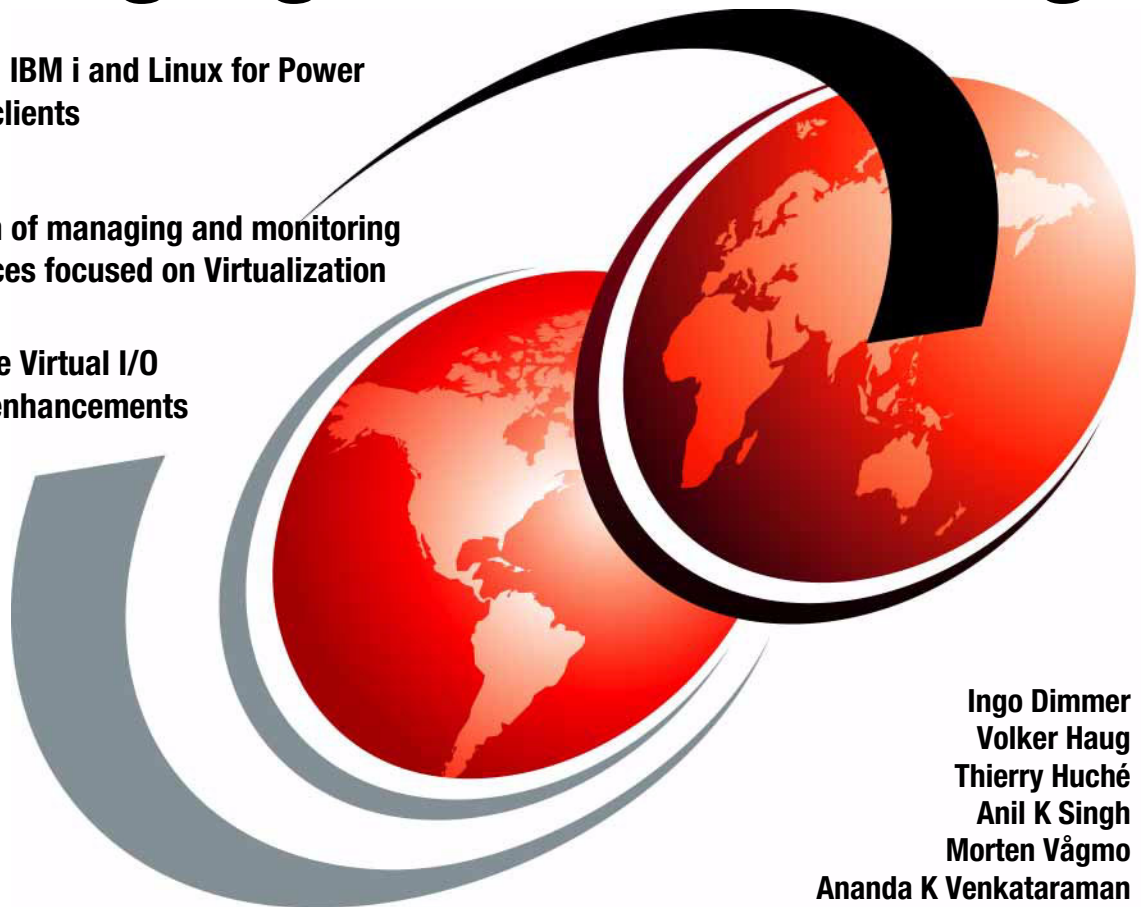


# PowerVM Virtualization on IBM Power Systems (Volume 2): Managing and Monitoring

Covers AIX, IBM i and Linux for Power  
virtual I/O clients

A collection of managing and monitoring  
best practices focused on Virtualization

Includes the Virtual I/O  
Server 2.1 enhancements



Ingo Dimmer  
Volker Haug  
Thierry Huché  
Anil K Singh  
Morten Vågmo  
Ananda K Venkataraman





International Technical Support Organization

**PowerVM Virtualization on Power Systems:  
Managing and Monitoring**

February 2009

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

## **Second Edition (February 2009)**

This edition applies to the following products:

IBM AIX 6.1 Technology Level 2 (5765-G62)

IBM AIX 5.3 Technology Level 9 (5765-G03)

IBM i 6.1 (5761-SS1)

Novell SUSE Linux Enterprise Server 10 for POWER (5639-S10)

Red Hat Enterprise Linux 5 for POWER (5639-RHL)

IBM PowerVM Express Edition (5765-PVX)

IBM PowerVM Standard Edition (5765-PVS)

IBM PowerVM Enterprise Edition (5765-PVE)

IBM Virtual I/O Server Version 2.1 with Fix Pack 20.1, or later

IBM Hardware Management Console Version 7.3.4, or later

IBM Tivoli Monitoring V6.1 for System p (5765-ITM)

IBM Tivoli Storage Manager (5608-ISM)

IBM Director 6.1

IBM Power Systems

© Copyright International Business Machines Corporation 2009. All rights reserved.

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

This document created or updated on November 7, 2008.



# Contents

<b>Figures</b> .....	xiii
<b>Tables</b> .....	xix
<b>Examples</b> .....	xxi
<b>Notices</b> .....	xxv
Trademarks .....	xxvi
<b>Preface</b> .....	xxvii
The team that wrote this book .....	xxviii
Become a published author .....	xxx
Comments welcome .....	xxx
<b>Part 1. PowerVM virtualization management</b> .....	1
<b>Chapter 1. Introduction</b> .....	1
1.1 PowerVM Editions .....	1
1.2 Maintenance strategy .....	6
1.3 New features for Virtual I/O Server Version 2.1 .....	7
1.4 Other PowerVM Enhancements .....	8
<b>Chapter 2. Virtual storage management</b> .....	11
2.1 Disk mapping options .....	12
2.1.1 Physical volumes .....	12
2.1.2 Logical volumes .....	13
2.1.3 File-backed devices .....	15
2.2 Using virtual optical devices .....	15
2.3 Using virtual tape devices .....	20
2.4 Using file-backed devices .....	25
2.5 Managing the mapping of LUNs over vSCSI to hdisks .....	29
2.5.1 Naming conventions .....	31
2.5.2 Virtual device slot numbers .....	32
2.5.3 Tracing a configuration .....	35
2.6 Replacing a disk on the Virtual I/O Server .....	41
2.6.1 Replacing a LV backed disk in the mirroring environment .....	41
2.6.2 Replacing a mirrored storage pool backed disk .....	47

2.7	Managing multiple storage security zones	51
2.8	Storage planning with migration in mind	52
2.8.1	Virtual adapter slot numbers	53
2.8.2	SAN considerations for LPAR migration	54
2.8.3	Backing devices and virtual target devices	55
2.9	N_Port ID virtualization	56
2.9.1	Introduction	56
2.9.2	Requirements	59
2.9.3	Managing virtual Fibre Channel adapters	60
2.9.4	Configuring NPIV on Power Systems with a new AIX LPAR	62
2.9.5	Configuring NPIV on Power Systems with existing AIX LPARs	74
2.9.6	Redundancy configurations for virtual Fibre Channel adapters	88
2.9.7	Replacing a Fibre Channel adapter configured with NPIV	92
2.9.8	Migration to virtual Fibre Channel adapter environments	93
2.9.9	Heterogeneous configuration with NPIV	107
<b>Chapter 3. Virtual network management</b>		<b>115</b>
3.1	Changing IP addresses or VLAN	116
3.1.1	Virtual I/O Server network address changes	116
3.1.2	Virtual I/O client network address changes	117
3.2	Managing the mapping of network devices	118
3.2.1	Virtual network adapters and VLANs	119
3.2.2	Virtual device slot numbers	120
3.2.3	Tracing a configuration	120
3.3	SEA threading on the Virtual I/O Server	126
3.4	Jumbo frame and path MTU discovery	127
3.4.1	Maximum transfer unit	127
3.4.2	Path MTU discovery	129
3.4.3	Using jumbo frames	132
3.4.4	Virtual Ethernet tuning with path MTU discovery	135
3.4.5	TCP checksum offload	135
3.4.6	Largesend option	135
3.5	SEA enhancements	137
3.5.1	Strict mode	138
3.5.2	Loose mode	138
3.5.3	Setting up QoS	138
3.5.4	Best practices in setting Mode for QoS	140
3.6	DoS Hardening	140
3.6.1	Solution	141
<b>Chapter 4. Virtual I/O Server security</b>		<b>143</b>
4.1	Network security	144
4.1.1	Stopping network services	144



4.1.2	Setting up the firewall . . . . .	144
4.1.3	Enabling ping through the firewall . . . . .	147
4.1.4	Security Hardening Rules . . . . .	148
4.1.5	DoS Hardening . . . . .	149
4.2	The Virtual I/O Server as an LDAP client . . . . .	149
4.2.1	Creating a key database file . . . . .	149
4.2.2	Configuring the LDAP server . . . . .	156
4.2.3	Configuring the Virtual I/O Server as an LDAP client . . . . .	161
4.3	Network Time Protocol configuration . . . . .	163
4.4	Setting up Kerberos on the Virtual I/O Server . . . . .	164
4.5	Managing users . . . . .	167
4.5.1	Creating a system administrator account . . . . .	167
4.5.2	Creating a service representative (SR) account . . . . .	168
4.5.3	Creating a read-only account . . . . .	169
4.5.4	Checking the global command log (gcl) . . . . .	169
	<b>Chapter 5. Virtual I/O Server maintenance . . . . .</b>	<b>171</b>
5.1	Installation or migration of Virtual I/O Server Version 2.1 . . . . .	172
5.1.1	Installation of Virtual I/O Server Version 2.1 . . . . .	173
5.1.2	Migration from an HMC . . . . .	173
5.1.3	Migration from DVD managed by an HMC . . . . .	174
5.1.4	Migration from DVD managed by IVM . . . . .	186
5.2	Virtual I/O server backup strategy . . . . .	188
5.2.1	Backup external device configuration . . . . .	189
5.2.2	Backup HMC resources . . . . .	189
5.2.3	Backup IVM resources . . . . .	190
5.2.4	Backup operating systems from the client logical partitions . . . . .	190
5.2.5	Backup the Virtual I/O Server operating system . . . . .	191
5.3	Scheduling backups of the Virtual I/O Server . . . . .	192
5.4	Backing up the Virtual I/O Server operating system . . . . .	193
5.4.1	Backing up to tape . . . . .	193
5.4.2	Backing up to a DVD-RAM . . . . .	194
5.4.3	Backing up to a remote file . . . . .	196
5.4.4	Backing up user-defined virtual devices . . . . .	199
5.4.5	Backing up using IBM Tivoli Storage Manager . . . . .	203
5.5	Restoring the Virtual I/O Server . . . . .	205
5.5.1	Restoring the HMC configuration . . . . .	206
5.5.2	Restoring other IT infrastructure devices . . . . .	206
5.5.3	Restoring the Virtual I/O Server operating system . . . . .	206
5.5.4	Recovering user-defined virtual devices and disk structure . . . . .	216
5.5.5	Restoring the Virtual I/O Server client operating system . . . . .	217
5.6	Rebuilding the Virtual I/O Server . . . . .	218
5.6.1	Rebuild the SCSI configuration . . . . .	221

5.6.2 Rebuild network configuration . . . . .	224
5.7 Updating the Virtual I/O Server . . . . .	226
5.7.1 Updating a single Virtual I/O Server environment . . . . .	226
5.7.2 Updating a dual Virtual I/O Server environment . . . . .	228
5.8 Error logging on the Virtual I/O Server . . . . .	239
5.8.1 Redirecting error logs to other servers . . . . .	240
5.8.2 Troubleshooting error logs . . . . .	241
<b>Chapter 6. Dynamic operations . . . . .</b>	<b>243</b>
6.1 Multiple Shared Processor Pools management . . . . .	244
6.2 Dynamic LPAR operations on AIX and IBM i . . . . .	248
6.2.1 Adding and removing processors dynamically . . . . .	248
6.2.2 Add memory dynamically . . . . .	251
6.2.3 Removing memory dynamically . . . . .	254
6.2.4 Add physical adapters dynamically . . . . .	256
6.2.5 Move physical adapters dynamically . . . . .	258
6.2.6 Removing physical adapters dynamically . . . . .	263
6.2.7 Add virtual adapters dynamically . . . . .	265
6.2.8 Removing virtual adapters dynamically . . . . .	269
6.2.9 Removing or replacing a PCI Hot Plug adapter . . . . .	271
6.3 Dynamic LPAR operations on Linux for Power . . . . .	272
6.3.1 Service and productivity tools for Linux for Power . . . . .	272
6.4 Dynamic LPAR operations on the Virtual I/O Server . . . . .	287
6.4.1 Ethernet adapter replacement on the Virtual I/O Server . . . . .	287
6.4.2 Replacing a Fibre Channel adapter on the Virtual I/O Server . . . . .	290
<b>Chapter 7. PowerVM Live Partition Mobility . . . . .</b>	<b>293</b>
7.1 What's new in PowerVM Live Partition Mobility . . . . .	294
7.2 PowerVM Live Partition Mobility requirements . . . . .	294
7.2.1 HMC requirements . . . . .	294
7.2.2 Common system requirements . . . . .	295
7.2.3 Source system requirements . . . . .	296
7.2.4 Destination system requirements . . . . .	296
7.2.5 Migrating partition requirements . . . . .	297
7.2.6 Active and inactive migrations . . . . .	297
7.3 Managing a live partition migration . . . . .	298
7.3.1 The migration validation . . . . .	298
7.3.2 Validation and migration . . . . .	298
7.3.3 How to fix missing requirements . . . . .	301
7.4 Differences with Live Application Mobility . . . . .	302
<b>Chapter 8. System Planning Tool . . . . .</b>	<b>305</b>
8.1 Sample scenario . . . . .	306
8.1.1 Preparation recommendation . . . . .	307

8.1.2 Planning the configuration with SPT . . . . .	308
8.1.3 Initial setup checklist . . . . .	317
<b>Chapter 9. Automated management . . . . .</b>	<b>321</b>
9.1 Automating remote operations . . . . .	322
9.1.1 Remotely powering a Power Systems server on and off . . . . .	324
9.1.2 Remotely starting and stopping logical partitions . . . . .	324
9.2 Scheduling jobs on the Virtual I/O Server . . . . .	325
<b>Chapter 10. High-level management . . . . .</b>	<b>327</b>
10.1 IBM Systems Director . . . . .	328
10.1.1 IBM Director installation on AIX . . . . .	331
10.1.2 Log on to IBM Systems Director . . . . .	333
10.1.3 Preparing managed systems . . . . .	334
10.1.4 Discover managed systems . . . . .	338
10.1.5 Collect inventory data . . . . .	341
10.1.6 View Managed resources . . . . .	344
10.1.7 Power Systems Management summary . . . . .	346
10.1.8 IBM Systems Director Virtualization Manager plug-in. . . . .	349
10.1.9 Manage Virtual I/O Server with IBM Systems Director . . . . .	350
10.1.10 Active Energy Manager. . . . .	359
10.2 Cluster Systems Management . . . . .	360
10.2.1 CSM architecture and components. . . . .	361
10.2.2 CSM and the Virtual I/O Server. . . . .	362
<b>Part 2. PowerVM virtualization monitoring . . . . .</b>	<b>365</b>
<b>Chapter 11. Virtual I/O Server monitoring agents . . . . .</b>	<b>369</b>
11.1 IBM Tivoli Monitoring. . . . .	370
11.1.1 What to monitor. . . . .	370
11.1.2 Agent configuration . . . . .	371
11.1.3 Using the Tivoli Enterprise Portal . . . . .	372
11.2 Configuring the IBM Tivoli Storage Manager client. . . . .	380
11.3 IBM Tivoli Usage and Accounting Manager agent . . . . .	381
11.4 IBM TotalStorage Productivity Center. . . . .	383
11.5 IBM Tivoli Application Dependency Discovery Manager. . . . .	386
<b>Chapter 12. Monitoring global system resource allocations . . . . .</b>	<b>389</b>
12.1 Hardware Management Console monitoring. . . . .	390
12.1.1 Partition properties monitoring . . . . .	391
12.1.2 HMC hardware information monitoring . . . . .	391
12.1.3 HMC virtual network monitoring . . . . .	393
12.1.4 HMC shell scripting . . . . .	394

12.2	Integrated Virtualization Manager monitoring . . . . .	395
12.3	Monitoring resources allocations from a partition . . . . .	397
12.3.1	Monitoring CPU and memory allocations from AIX . . . . .	397
12.3.2	Monitoring CPU and memory allocations from Linux . . . . .	398
	<b>Chapter 13. Monitoring commands on the Virtual I/O Server . . . . .</b>	<b>401</b>
13.1	Global system monitoring . . . . .	402
13.2	Device inspection . . . . .	402
13.3	Storage monitoring and listing . . . . .	403
13.4	Network monitoring . . . . .	404
13.5	User ID listing . . . . .	404
	<b>Chapter 14. CPU monitoring . . . . .</b>	<b>405</b>
14.1	CPU-related terminology and metrics . . . . .	406
14.1.1	Terminology and metrics common to POWER5 and POWER6 . . . . .	406
14.1.2	Terminology and metrics specific to POWER6 systems . . . . .	409
14.2	CPU metrics computation . . . . .	411
14.2.1	Processor Utilization of Resources Register (PURR) . . . . .	411
14.2.2	New PURR-based metrics . . . . .	412
14.2.3	System-wide tools modified for virtualization . . . . .	414
14.2.4	Scaled Processor Utilization of Resources Register (SPURR) . . . . .	414
14.3	Cross-partition CPU monitoring . . . . .	416
14.4	AIX and Virtual I/O Server CPU monitoring . . . . .	422
14.4.1	Monitoring using topas . . . . .	422
14.4.2	Monitoring using nmon . . . . .	425
14.4.3	Monitoring using vmstat . . . . .	428
14.4.4	Monitoring using lparstat . . . . .	429
14.4.5	Monitoring using sar . . . . .	431
14.4.6	Monitoring using mpstat . . . . .	434
14.4.7	Report generation for CPU utilization . . . . .	435
14.5	IBM i CPU monitoring . . . . .	442
14.6	Linux for Power CPU monitoring . . . . .	448
	<b>Chapter 15. Memory monitoring . . . . .</b>	<b>451</b>
15.1	Cross-partition memory monitoring . . . . .	452
15.2	IBM i memory monitoring . . . . .	453
15.3	Linux for Power memory monitoring . . . . .	458
	<b>Chapter 16. Virtual storage monitoring . . . . .</b>	<b>459</b>
16.1	Virtual I/O Server storage monitoring . . . . .	460
16.1.1	Checking storage health on the Virtual I/O Server . . . . .	460
16.1.2	Monitoring storage performance on the Virtual I/O Server . . . . .	460
16.2	AIX virtual I/O client storage monitoring . . . . .	461
16.2.1	Checking storage health on the AIX virtual I/O client . . . . .	462

16.2.2	Monitoring storage performance on the AIX virtual I/O client . . . .	465
16.3	IBM i virtual I/O client storage monitoring . . . . .	466
16.3.1	Checking storage health on the IBM i virtual I/O client . . . . .	466
16.3.2	Monitoring storage performance on the IBM i virtual I/O client . . .	467
16.4	Linux for Power virtual I/O client storage monitoring. . . . .	472
<b>Chapter 17. Virtual network monitoring . . . . .</b>		<b>475</b>
17.1	Monitoring the Virtual I/O Server. . . . .	476
17.1.1	Error logs. . . . .	476
17.1.2	IBM Tivoli Monitoring. . . . .	476
17.1.3	Testing your configuration. . . . .	477
17.2	Virtual I/O Server networking monitoring. . . . .	481
17.2.1	Describing the scenario. . . . .	481
17.2.2	Advanced SEA monitoring . . . . .	492
17.3	AIX client network monitoring . . . . .	497
17.4	IBM i client network monitoring . . . . .	497
17.4.1	Checking network health on the IBM i virtual I/O client. . . . .	497
17.4.2	Monitoring network performance on the IBM i virtual I/O client. . .	500
17.5	Linux for Power client network monitoring. . . . .	501
<b>Chapter 18. Third-party monitoring tools for AIX and Linux. . . . .</b>		<b>503</b>
18.1	nmon utility . . . . .	503
18.1.1	nmon on AIX 6.1 . . . . .	504
18.1.2	nmon on Linux. . . . .	506
18.1.3	Additional nmon statistics . . . . .	507
18.1.4	Recording with the nmon tool . . . . .	507
18.2	Sysstat utility . . . . .	507
18.3	Ganglia tool . . . . .	508
18.4	Other third party tools . . . . .	508
<b>Appendix A. Sample script for disk and NIB network checking and recovery on AIX virtual clients. . . . .</b>		<b>511</b>
<b>Abbreviations and acronyms . . . . .</b>		<b>519</b>
<b>Related publications . . . . .</b>		<b>523</b>
	IBM Redbooks . . . . .	523
	Other publications . . . . .	524
	Online resources . . . . .	524
	How to get Redbooks. . . . .	526
	Help from IBM . . . . .	526
<b>Index . . . . .</b>		<b>527</b>



# Figures

2-1	SCSI setup for shared optical device . . . . .	16
2-2	IBM i Work with Storage Resources screen . . . . .	17
2-3	IBM i Logical Hardware Resources screen I/O debug option . . . . .	18
2-4	IBM i Select IOP Debug Function screen IPL I/O processor option . . . . .	19
2-5	IBM i Select IOP Debug Function screen Reset I/O processor option . . . . .	20
2-6	SCSI setup for shared tape device . . . . .	21
2-7	Logical versus physical drive mapping . . . . .	30
2-8	Setting maximum number of virtual adapters in a partition profile . . . . .	34
2-9	IBM i SST Display Disk Configuration Status screen . . . . .	37
2-10	IBM i SST Display Disk Unit Details screen . . . . .	38
2-11	IBM i partition profile virtual adapters configuration . . . . .	39
2-12	AIX LVM mirroring environment with LV backed virtual disks . . . . .	42
2-13	AIX LVM mirroring environment with storage pool backed virtual disks . . . . .	47
2-14	Create virtual SCSI . . . . .	52
2-15	Slot numbers that are identical in the source and target system . . . . .	53
2-16	Server using redundant Virtual I/O Server partitions with NPIV . . . . .	58
2-17	Virtual adapter numbering. . . . .	63
2-18	Dynamically add virtual adapter. . . . .	65
2-19	Create Fibre Channel server adapter . . . . .	66
2-20	Set virtual adapter ID . . . . .	66
2-21	Save the Virtual I/O Server partition configuration. . . . .	67
2-22	Change profile to add virtual Fibre Channel client adapter. . . . .	68
2-23	Create Fibre Channel client adapter . . . . .	68
2-24	Define virtual adapter ID values . . . . .	69
2-25	Select virtual Fibre Channel client adapter properties . . . . .	71
2-26	Virtual Fibre Channel client adapter Properties . . . . .	71
2-27	LUN mapping on DS4800 . . . . .	73
2-28	NPIV configuration . . . . .	74
2-29	Dynamically add virtual adapter. . . . .	77
2-30	Create Fibre Channel server adapter . . . . .	78
2-31	Set virtual adapter ID . . . . .	78
2-32	Save the Virtual I/O Server partition configuration. . . . .	79
2-33	Dynamically add virtual adapter. . . . .	80
2-34	Create Fibre Channel client adapter . . . . .	81
2-35	Define virtual adapter ID values . . . . .	81
2-36	Save the virtual I/O client partition configuration. . . . .	82
2-37	Select virtual Fibre Channel client adapter properties . . . . .	84
2-38	Virtual Fibre Channel client adapter Properties . . . . .	85

2-39	LUN mapping within DS4800	87
2-40	Host bus adapter failover	89
2-41	Host bus adapter and Virtual I/O Server failover.	90
2-42	LUN mapped to a physical Fibre Channel adapter	94
2-43	Add Virtual Adapter to vios1 partition	95
2-44	Create virtual Fibre Channel server adapter in vios1 partition	96
2-45	Set Adapter IDs in vios1 partition	97
2-46	Add a virtual adapter to NPIV partition	98
2-47	Create virtual Fibre Channel client adapter in NPIV partition	99
2-48	Set Adapter IDs in NPIV partition	100
2-49	Add new host port	104
2-50	Remove a physical Fibre Channel adapter	105
2-51	Select adapter to be removed	106
2-52	Heterogeneous NPIV configuration	108
3-1	HMC Virtual Network Management	121
3-2	Virtual Ethernet adapter slot assignments	122
3-3	IBM i Work with Communication Resources screen	123
3-4	IBM i Display Resource Details screen	124
3-5	HMC IBMi61 partition properties screen	125
3-6	HMC virtual Ethernet adapter properties screen	125
3-7	IBM i Work with TCP/IP Interface Status screen	134
4-1	ikeyman program initial window	151
4-2	Create new key database screen	151
4-3	Creating the ldap_server key	152
4-4	Setting the key database password	152
4-5	Default certificate authorities available on the ikeyman program	153
4-6	Creating a self-signed certificate initial screen	154
4-7	Self-signed certificate information	155
4-8	Default directory information tree created by the mkseclap command	156
5-1	Define the System Console	176
5-2	Installation and Maintenance main menu	177
5-3	Virtual I/O Server Migration Installation and Settings	178
5-4	Change Disk Where You Want to Install	179
5-5	Virtual I/O Server Migration Installation and Settings - start migration	180
5-6	Migration Confirmation	181
5-7	Running migration	182
5-8	Set Terminal Type	183
5-9	Software License Agreements	184
5-10	Accept License Agreements	185
5-11	IBM Virtual I/O Server login menu	186
5-12	Example of a System Plan generated from a managed system	219
5-13	IBM i Work with TCP/IP Interface Status screen	230
5-14	Virtual I/O client running MPIO	232



5-15	Virtual I/O client partition software mirroring . . . . .	233
5-16	IBM i Display Disk Configuration Status screen . . . . .	234
6-1	Shared Processor Pool . . . . .	244
6-2	Modifying Shared Processor pool attributes . . . . .	245
6-3	Partitions assignment to Multiple Shared Processor Pools. . . . .	246
6-4	Assign a partition to a Shared Processor Pool . . . . .	246
6-5	Comparing partition weights from different Shared Processor Pools . . . . .	247
6-6	Add or remove processor operation . . . . .	249
6-7	Defining the amount of CPU processing units for a partition . . . . .	250
6-8	IBM i Work with System Activity screen . . . . .	251
6-9	Add or remove memory operation. . . . .	252
6-10	Changing the total amount of memory of the partition to 5 GB . . . . .	253
6-11	Dynamic LPAR operation in progress. . . . .	253
6-12	Add or remove memory operation. . . . .	254
6-13	Dynamically reducing 1 GB from a partition . . . . .	255
6-14	LPAR overview menu . . . . .	256
6-15	Add physical adapter operation. . . . .	257
6-16	Select physical adapter to be added . . . . .	258
6-17	I/O adapters properties for a managed system. . . . .	259
6-18	Move or remove physical adapter operation . . . . .	261
6-19	Selecting adapter in slot C2 to be moved to partition AIX_LPAR . . . . .	262
6-20	Save current configuration . . . . .	263
6-21	Remove physical adapter operation . . . . .	264
6-22	Select physical adapter to be removed. . . . .	265
6-23	Add virtual adapter operation . . . . .	266
6-24	Dynamically adding virtual SCSI adapter . . . . .	267
6-25	Virtual SCSI adapter properties . . . . .	268
6-26	Virtual adapters for an LPAR . . . . .	269
6-27	Remove virtual adapter operation . . . . .	270
6-28	Delete virtual adapter . . . . .	271
6-29	Add processor to a Linux partition. . . . .	280
6-30	Increasing the number of virtual processors . . . . .	281
6-31	DLPAR add or remove memory . . . . .	285
6-32	DLPAR adding 2 GB memory . . . . .	285
7-1	Partition Migration Validation . . . . .	299
7-2	Partition Migration . . . . .	300
7-3	Partition migration validation detailed information. . . . .	301
8-1	The partition and slot numbering plan of virtual storage adapters . . . . .	306
8-2	The partition and slot numbering plan for virtual Ethernet adapters . . . . .	307
8-3	The SPT Partition properties window . . . . .	308
8-4	The SPT SCSI connections window . . . . .	309
8-5	The SPT Edit Virtual Slots window . . . . .	310
8-6	The SPT System Plan ready to be deployed . . . . .	311

8-7 Deploy System Plan Wizard . . . . .	311
8-8 The System Plan validation screen . . . . .	312
8-9 The Partition Deployment menu . . . . .	313
8-10 The Operating Environment Install Deployment menu . . . . .	314
8-11 The Deployment Progress screen . . . . .	315
8-12 Partition profiles deployed on the HMC . . . . .	316
9-1 Creating a System Profile on the HMC . . . . .	322
9-2 The HMC Remote Command Execution menu . . . . .	323
10-1 IBM Systems Director Environment . . . . .	330
10-2 IBM Director login screen . . . . .	333
10-3 Welcome to IBM Systems Director screen . . . . .	334
10-4 HMC LAN Adapter Details . . . . .	335
10-5 IBM Director System Discovery . . . . .	338
10-6 IBM Director Discovered Systems table . . . . .	339
10-7 IBM Director Request Access . . . . .	340
10-8 IBM Director Access granted . . . . .	341
10-9 IBM Director View and Collect Inventory . . . . .	342
10-10 IBM Director Run Collect Inventory . . . . .	343
10-11 IBM Director Collection job . . . . .	344
10-12 IBM Director Navigate Resources . . . . .	345
10-13 IBM Director All Systems (View Members) . . . . .	345
10-14 IBM Director Power Systems Management summary . . . . .	346
10-15 IBM Director Power Systems Management Manage resources . . . . .	347
10-16 IBM Director Virtualization Manager . . . . .	350
10-17 IBM Director Create Group Summary . . . . .	351
10-18 IBM Director Create Virtual Server . . . . .	352
10-19 IBM Director Create Virtual Server . . . . .	353
10-20 IBM Director Create Virtual Server disk . . . . .	354
10-21 IBM Director Virtual LAN Adapters . . . . .	355
10-22 IBM Director Virtual SCSI Topology . . . . .	355
10-23 IBM Director Basic Topology map . . . . .	356
10-24 IBM Director Hardware Inventory . . . . .	357
10-25 IBM Director Virtualization Manager Monitor . . . . .	358
10-26 IBM Director CPU Utilization graph . . . . .	359
11-1 Tivoli Enterprise Portal login using web browser . . . . .	373
11-2 Tivoli Enterprise Portal login . . . . .	374
11-3 Storage Mappings Workspace selection . . . . .	374
11-4 ITM panel showing Storage Mappings . . . . .	375
11-5 ITM panel showing Network Mappings . . . . .	376
11-6 ITM panel showing Top Resources usage . . . . .	377
11-7 ITM panel showing CPU Utilization . . . . .	378
11-8 ITM panel showing System Storage Information . . . . .	379
11-9 ITM panel showing Network Adapter Utilization . . . . .	380

12-1	Available servers managed by the HMC . . . . .	390
12-2	Configuring the displayed columns on the HMC . . . . .	390
12-3	Virtual adapters configuration in the partition properties . . . . .	391
12-4	Virtual I/O Server hardware information context menu . . . . .	392
12-5	The Virtual I/O Server virtual SCSI topology window . . . . .	392
12-6	Virtual Network Management . . . . .	393
12-7	Virtual Network Management - detail information . . . . .	394
12-8	IVM partitions monitoring . . . . .	395
12-9	IVM virtual Ethernet configuration monitoring . . . . .	396
12-10	IVM virtual storage configuration monitoring . . . . .	396
14-1	16-core system with dedicated and shared CPUs . . . . .	407
14-2	A Multiple Shared-Processor Pool example on POWER6 . . . . .	410
14-3	Per-thread PURR . . . . .	412
14-4	Dedicated partition's Processor Sharing properties . . . . .	418
14-5	topas displaying monitoring information . . . . .	426
14-6	Initial screen of NMON application . . . . .	426
14-7	Display of command help for monitoring system resources . . . . .	427
14-8	Monitoring CPU activity with NMON . . . . .	427
14-9	NMON monitoring of CPU and network resources . . . . .	428
14-10	smitty topas for CPU utilization reporting . . . . .	436
14-11	Local CEC recording attributes screen . . . . .	437
14-12	Report generation . . . . .	438
14-13	Reporting Formats . . . . .	439
14-14	IBM i WRKSYSACT command output . . . . .	443
14-15	IBM i CPU Utilization and Waits Overview . . . . .	447
14-16	mpstat command output . . . . .	449
15-1	IBM i WRKSYSSTS command output . . . . .	454
15-2	IBM i System Director Navigator Page fault overview . . . . .	456
15-3	Linux monitoring memory statistics using meminfo . . . . .	458
16-1	AIX virtual I/O client using MPIO . . . . .	462
16-2	AIX virtual I/O client using LVM mirroring . . . . .	464
16-3	IBM i Display Disk Configuration Status . . . . .	467
16-4	IBM i WRKDSKSTS command output . . . . .	468
16-5	IBM i Navigator Disk Overview for System Disk Pool . . . . .	471
16-6	iostat command output showing the i/o output activity . . . . .	472
16-7	iostat output with -d flag and 5 sec interval as a parameter . . . . .	473
17-1	Network monitoring testing scenario . . . . .	482
18-1	nmon LPAR statistics on an AIX shared partition . . . . .	505
18-2	nmon LPAR statistics on an AIX dedicated partition . . . . .	506
18-3	nmon LPAR statistics report for a Linux partition . . . . .	506



# Tables

1-1	PowerVM Editions components, editions, and hardware support . . . . .	4
3-1	Typical maximum transmission units (MTUs) . . . . .	127
4-1	Default open ports on Virtual I/O Server . . . . .	144
4-2	Hosts on the network . . . . .	145
4-3	Task and associated command to manage Virtual I/O Server users . . .	167
5-1	Virtual I/O Server backup and restore methods . . . . .	192
5-2	Commands to save information about Virtual I/O Server . . . . .	202
5-3	Error log entry classes . . . . .	240
6-1	Service & Productivity tools description . . . . .	273
7-1	Missing requirements for PowerVM Live Partition Mobility . . . . .	301
7-2	PowerVM Live Partition Mobility versus Live Application Mobility . . . . .	303
10-1	Terms for IBM Systems Director . . . . .	348
10-2	Tools for monitoring resources in a virtualized environment . . . . .	366
11-1	TPC agent attributes, descriptions, and their values. . . . .	384
14-1	POWER5-based terminology and metrics . . . . .	408
14-2	POWER6-specific terminology and metrics . . . . .	410
14-3	IBM i CPU utilization guidelines . . . . .	445



# Examples

2-1	Making the virtual device for the tape drive. . . . .	22
2-2	Finding which LPAR is holding the tape drive using dsh . . . . .	23
2-3	Finding which LPAR is holding the optical drive using ssh . . . . .	24
2-4	Checking the version of the Virtual I/O Server . . . . .	25
2-5	Checking whether any virtual media repository is already defined . . . . .	25
2-6	List of available storage pools and defining a virtual media repository . . . . .	26
2-7	Creating a virtual optical media disk in the virtual media repository . . . . .	26
2-8	Creating an iso image from CD/DVD drive . . . . .	27
2-9	Creating an optical virtual target device . . . . .	27
2-10	Loading the virtual media on the virtual target device. . . . .	27
2-11	Checking the virtual optical device contents on a client . . . . .	28
2-12	Loading a new disk on the virtual media device . . . . .	29
2-13	The fget_config command for the DS4000 series. . . . .	31
2-14	SAN storage listing on the Virtual I/O Server version 2.1 . . . . .	32
2-15	Tracing virtual SCSI storage from Virtual I/O Server . . . . .	35
2-16	Tracing NPIV virtual storage from the Virtual I/O Server . . . . .	35
2-17	Displaying the Virtual I/O Server device mapping. . . . .	39
2-18	Virtual I/O Server hdisk to LUN tracing . . . . .	40
2-19	Find the disk to remove. . . . .	43
2-20	version command shows Fabric OS level. . . . .	63
2-21	List port configuration . . . . .	64
2-22	lsdev -dev vfchost* command on the Virtual I/O Server . . . . .	69
2-23	lsdev -dev fcs* command on the Virtual I/O Server . . . . .	69
2-24	lsnports command on the Virtual I/O Server . . . . .	70
2-25	vfcmmap command with vfchost2 and fcs3 . . . . .	70
2-26	lsmmap -npiv -vadapter vfchost0 command . . . . .	70
2-27	zonestow command before adding a new WWPN. . . . .	72
2-28	cfgsave and cfgenable commands . . . . .	72
2-29	zonestow command after adding a new WWPN . . . . .	72
2-30	version command shows Fabric OS level. . . . .	75
2-31	List port configuration . . . . .	75
2-32	lsdev -dev vfchost* command on the Virtual I/O Server . . . . .	82
2-33	lsdev -dev fcs* command on the Virtual I/O Server . . . . .	83
2-34	lsnports command on the Virtual I/O Server . . . . .	83
2-35	vfcmmap command with vfchost2 and fcs3 . . . . .	83
2-36	lsmmap -all -npiv command . . . . .	83
2-37	zonestow command before adding a WWPN . . . . .	85
2-38	cfgsave and cfgenable commands . . . . .	86

2-39	zonestow command after adding a new WWPN . . . . .	86
2-40	cfgmgr and lspv command from the AIX client partition . . . . .	87
2-41	mpio_get_config command from the AIX client partition. . . . .	88
2-42	Removing a NPIV Fibre Channel adapter in the Virtual I/O Server . . . . .	93
2-43	Show available Fibre Channel adapters . . . . .	94
2-44	WWPN of the virtual Fibre Channel client adapter in NPIV partition. . . . .	102
2-45	Zoning of WWPN for fcs2 . . . . .	102
2-46	Verifying MPIO in a heterogeneous configuration. . . . .	108
3-1	The default MSS value in AIX 6.1 . . . . .	128
3-2	Path MTU display . . . . .	131
3-3	Largesend option for SEA . . . . .	136
3-4	Configuring QoS for an SEA . . . . .	139
3-5	Configuring VLAN for existing vlan device . . . . .	139
3-6	Enabling network traffic regulation . . . . .	141
3-7	Using tcptr for Network traffic regulation for sendmail service . . . . .	141
4-1	Stopping network services . . . . .	144
4-2	Firewall view . . . . .	146
4-3	Firewall rules output . . . . .	146
4-4	High level Firewall settings . . . . .	148
4-5	Creating an ldap user on the Virtual I/O Server . . . . .	162
4-6	Log on to the Virtual I/O Server using an LDAP user . . . . .	162
4-7	Searching the LDAP server. . . . .	162
4-8	Content of the /home/padmin/config/ntp.conf file . . . . .	163
4-9	Too large time error. . . . .	163
4-10	Successful ntp synchronization. . . . .	163
4-11	Setting up a symbolic link for ntp.conf. . . . .	164
4-12	Creating a system administrator user and checking its attributes. . . . .	167
4-13	Creating a service representative account . . . . .	168
4-14	lsgcl command output. . . . .	169
5-1	Backing up the Virtual I/O Server to tape . . . . .	194
5-2	Backing up the Virtual I/O Server to DVD-RAM . . . . .	195
5-3	Backing up the Virtual I/O Server to the nim_resources.tar file . . . . .	198
5-4	Backing up the Virtual I/O Server to the mksysb image . . . . .	198
5-5	Sample output from the lsmmap command . . . . .	201
5-6	Restore of Virtual I/O Server to the same logical partition . . . . .	211
5-7	Devices recovered if restored to a different server . . . . .	213
5-8	Disks and volume groups to restore . . . . .	216
5-9	Creating an HMC system plan from the HMC command line . . . . .	218
5-10	lsmmap -all command . . . . .	222
5-11	The netstat -v comand on the virtual I/O client . . . . .	229
5-12	The netstat -cdlistats command on the primary Virtual I/O Server . . . . .	230
5-13	The netstat -cdlistats command on the secondary Virtual I/O Server. . . . .	231
5-14	The mdstat command showing a healthy environment. . . . .	234



5-15	AIX LVM Mirror Resync . . . . .	236
5-16	errlog short listing . . . . .	239
5-17	Detailed error listing . . . . .	239
5-18	Content of /tmp/syslog.add file . . . . .	241
5-19	Create new error log file . . . . .	241
5-20	Copy errlog and view it . . . . .	242
6-1	Removing the Fibre Channel adapter . . . . .	260
6-2	Installing Service and Productivity tools . . . . .	276
6-3	lscfg command on Linux . . . . .	277
6-4	lsvpd command . . . . .	278
6-5	Display virtual SCSI and network . . . . .	278
6-6	List the management server . . . . .	279
6-7	Linux finds new processors . . . . .	281
6-8	lparcfg command before adding CPU dynamically . . . . .	281
6-9	lparcfg command after addition of 0.1 CPU dynamically . . . . .	282
6-10	Ready to die message . . . . .	283
6-11	Display of total memory in the partition before adding memory . . . . .	284
6-12	Total memory in the partition after adding 1GB dynamically . . . . .	286
6-13	Rescanning a SCSI host adapter . . . . .	286
10-1	Installing IBM Director . . . . .	331
10-2	Installing IBM Director Common Agent . . . . .	337
10-3	Hardware definition example . . . . .	362
12-1	lparstat -i command output on AIX . . . . .	397
12-2	Listing partition resources on Linux . . . . .	398
14-1	topas -cecdisp command on Virtual I/O Server . . . . .	416
14-2	topas -C command on virtual I/O client . . . . .	417
14-3	topas -C global information with the g command . . . . .	420
14-4	Monitoring processor pools with topas -C . . . . .	421
14-5	Shared pool partitions listing in topas . . . . .	422
14-6	Basic topas monitoring . . . . .	423
14-7	Logical partition information report in topas (press L) . . . . .	424
14-8	Upper part of topas busiest CPU report . . . . .	425
14-9	Monitoring with the vmstat command . . . . .	428
14-10	Monitoring using the lparstat command . . . . .	430
14-11	Variable processor frequency view with lparstat . . . . .	431
14-12	Individual CPU Monitoring using the sar command . . . . .	432
14-13	sar command working a previously saved file . . . . .	433
14-14	Individual CPU Monitoring using the mpstat command . . . . .	434
14-15	IBM i Component Report for Component Interval Activity . . . . .	444
14-16	IBM i System Report for Resource Utilization Expansion . . . . .	446
14-17	Usage of iostat for CPU monitoring . . . . .	449
15-1	Cross-partition memory monitoring with topas -C . . . . .	452
15-2	IBM i Component Report for Storage Pool Activity . . . . .	455

16-1	Monitoring I/O performance with viostat . . . . .	460
16-2	AIX lspath command output . . . . .	462
16-3	AIX client lsattr command to show hdisk attributes. . . . .	463
16-4	Using the chdev command for setting hdisk recovery parameters . . . .	463
16-5	Check missing disk . . . . .	464
16-6	AIX command to recover from stale partitions . . . . .	464
16-7	Monitoring disk performance with iostat . . . . .	465
16-8	IBM i System Report for Disk Utilization . . . . .	469
16-9	IBM i Resource Report for Disk Utilization . . . . .	469
17-1	Verifying the active channel in an EtherChannel . . . . .	478
17-2	Errorlog message when the primary channel fails . . . . .	479
17-3	Verifying the active channel in an EtherChannel . . . . .	480
17-4	Manual switch to primary channel using entstat . . . . .	480
17-5	Checking for the Link Failure count. . . . .	481
17-6	Output of entstat on SEA . . . . .	483
17-7	entstat -all command on SEA . . . . .	484
17-8	entstat -all command after file transfer attempt 1 . . . . .	485
17-9	entstat -all command after file transfer attempt 2 . . . . .	487
17-10	entstat -all command after file transfer attempt 3 . . . . .	488
17-11	entstat -all command after reset of Ethernet adapters . . . . .	489
17-12	entstat -all command after opening one ftp session . . . . .	490
17-13	entstat -all command after opening two ftp session . . . . .	491
17-14	Enabling advanced SEA monitoring . . . . .	492
17-15	Sample seastat statistics. . . . .	493
17-16	seastat statistics using search criterion. . . . .	496
17-17	IBM i Work with TCP/IP Interface Status screen. . . . .	498
17-18	IBM i Work with Configuration Status screen . . . . .	498
17-19	IBM i Work with Communication Resources screen . . . . .	499
17-20	IBM i System Report for TCP/IP Summary . . . . .	500
17-21	IBM i Resource Report for Disk Utilization . . . . .	500
18-1	Using a script to update partitions. . . . .	512
18-2	Running the script and listing output. . . . .	513

# 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

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

Active Memory™	Geographically Dispersed	POWER5™
AIX 5L™	Parallel Sysplex™	POWER5+™
AIX®	GPFS™	POWER6™
BladeCenter®	HACMP™	PowerHA™
DB2®	i5/OS®	PowerVM™
DS4000™	IBM®	PTX®
DS6000™	iSeries®	Redbooks®
DS8000™	Micro-Partitioning™	Redbooks (logo)  ®
EnergyScale™	OS/400®	System i®
Enterprise Storage Server®	Parallel Sysplex®	System p5®
eServer™	POWER™	System p®
GDPS®	POWER Hypervisor™	System Storage™
General Parallel File System™	Power Systems™	Tivoli®
	POWER4™	TotalStorage®

The following terms are trademarks of other companies:

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

VMware, the VMware "boxes" logo and design are registered trademarks or trademarks of VMware, Inc. in the United States and/or other jurisdictions.

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

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

UNIX is a registered trademark of The Open Group 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

PowerVM™ virtualization technology is a combination of hardware and software that supports and manages the virtual environments on POWER5™, POWER5+™ and POWER6™-based systems. It is a major tool to help simplify and optimize your IT infrastructure.

Available on most IBM® System p®, IBM Power Systems™, and the IBM BladeCenter® JS12 and JS22 servers as optional Editions and supported by the AIX®, IBM i, and Linux® for POWER operating systems, this set of comprehensive systems technologies and services is designed to enable you to aggregate and manage resources using a consolidated, logical view. The key benefits of deploying PowerVM virtualization and IBM Power Systems are as follows:

- ▶ Cut energy costs through server consolidation
- ▶ Reduce the cost of existing infrastructure
- ▶ Manage growth, complexity, and risk on your infrastructure

To achieve this goal, PowerVM virtualization provides the following technologies:

- ▶ Virtual Ethernet
- ▶ Shared Ethernet Adapter
- ▶ Virtual SCSI
- ▶ Micro-Partitioning™ technology

Additionally, these new technologies are available on POWER6 systems:

- ▶ Multiple Shared-Processor Pools
- ▶ N\_Port Identifier Virtualization
- ▶ PowerVM Live Partition Mobility (optional available with PowerVM Enterprise Edition)

To take complete advantage of these technologies and master your infrastructure needs as they evolve, you need to be able to correctly monitor and optimally manage your resources.

This publication is an extension of *PowerVM on System p: Introduction and Configuration*, SG24-7940. It provides an organized view of best practices for managing and monitoring your PowerVM environment with respect to virtualized resources managed by the Virtual I/O Server.

This publication is divided into two parts:

- ▶ The first part focuses on system management and describes some best practices to optimize the resources with some practical examples. It also details how to secure and maintain your virtual environments. Finally new features of the Virtual I/O Server Version 2.1, such as N\_Port ID Virtualization are covered and explained.
- ▶ The second part describes how to monitor a PowerVM virtualization infrastructure. Rather than presenting a list of tools, it addresses practical situations to help you select and use the monitoring tool that best shows the resources you are interested in. Some reminders on the key PowerVM features are also provided.

Although this publication can be read as a whole, you can also jump directly to sections in the managing or monitoring parts you are interested in.

## The team that wrote this book

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

**Ingo Dimmer** is an IBM Consulting IT Specialist for System i® and a PMI® Project Management Professional working in the IBM STG Europe storage support organization in Mainz, Germany. He has nine years of experience in enterprise storage support from working in IBM post-sales and pre-sales support. He holds a degree in Electrical Engineering from the Gerhard-Mercator University Duisburg. His areas of expertise include System i external disk and tape storage solutions, PowerVM virtualization, I/O performance and tape encryption for which he has been an author of several IBM Redbooks® and White Paper publications.

**Volker Haug** is a certified Consulting IT Specialist within IBM Systems and Technology Group located in Stuttgart, Germany. He holds a Bachelor's degree in Business Management from the University of Applied Studies in Stuttgart. His career has included more than 21 years working in IBM's PLM and Power Systems divisions as a RISC and AIX Systems Engineer. Volker is an expert in workstations and server hardware, AIX, and PowerVM virtualization. He is a member of the EMEA Power Champions team and also a member of the German Technical Expert Council, a affiliate to the IBM Academy of Technology. He wrote several books and whitepapers about AIX, workstations, servers, and PowerVM virtualization.

**Thierry Huché** is an IT Specialist working at the Products and Solutions Support Center in Montpellier, France, as an IBM Power Systems Benchmark Manager.

He has worked at IBM France for 19 years, and he has more than 17 years of AIX System Administration and Power Systems experience working in the pSeries Benchmark Center and AIX support center in Paris. He is an IBM Certified pSeries AIX System Administrator. His areas of expertise include benchmarking, virtualization, performance tuning, networking, high availability. He coauthored the IBM Redbook Communications Server for AIX Explored, SG24-2591 and IBM eServer Certification Study Guide: eServer p5 and pSeries Enterprise Technical Support AIX 5L V5.3SG24-7197.

**Anil K Singh** is a Senior software developer in IBM India. He has 7 years of experience in testing and development, including 4 years in AIX. He holds a Bachelor of Engineering degree in Computer Science. His area of expertise include programming and testing in Kernel and User mode for TCP/IP, malloc subsystem, ksh, vi, WPAR and WLM. He also has experience in planning and configuring hardware and software environment for WPAR and TCP/IP stress testing. He has good insight of Storage key and ProbeVue programming. He has co-authored one IBM Redbook and published two abstracts for the IBM Academy of Technology.

**Morten Vågmo** is a certified Consulting IT Specialist in IBM Norway with 20 years of AIX and Power Systems experience. Morten is Nordic AIX Competence Leader and member of the EMEA Power Champions team. He is working in technical pre-sales support and is now focusing on PowerVM implementations. He co-authored the update of the Advanced POWER Virtualization on IBM System p5: Introduction and Configuration redbook, and the PowerVM Virtualization on IBM System p: Introduction and Configuration redbook. He is speaker at technical conferences on the topic of virtualization. Morten holds a degree in Marine Engineering from the Technical University of Norway.

**Ananda K Venkataraman** is a software engineer in the Linux Technology Center at IBM Austin. He has four years of experience in Linux. His current focus is on Linux for Power, SAN storage technology, PowerVM Virtualization, and serial port device drivers. Ananda holds a Master's degree in Computer Science from the Texas State University.

**Scott Vetter (PMP)** is a Certified Executive Project Manager at the International Technical Support Organization, Austin Center. He has enjoyed 23 years of rich and diverse experience working for IBM in a variety of challenging roles. His latest efforts are directed at providing world-class Power Systems Redbooks, whitepapers, and workshop collateral.

The authors of the first edition are:

<b>Tomas Baublys</b>	IBM Germany
<b>Damien Faure</b>	Bull France

**Jackson Alfonso Krainer** IBM Brazil

**Michael Reed** IBM US

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

Francisco J. Alanis, Ray Anderson, Rich Avery, Paul S. Bostrom, Shamsundar Ashok, Jim Czenkusch, Carol Dziuba, Tom Edgerton, Jim Fall, Kevin W. Juhl, Bob Kovacs, Derek Matocha, Dawn May, Tommy Mclane, Dave Murray, Nguyen Nguyen, Niraj Patel, Vani Ramagiri, Bhargavi B. Reddy, Jacob Rosales, Simeon Rotich, Kate Tinklenberg, Scott Tran, Scott Urness, Vasu Vallabhaneni, Jonathan Van Niewaal, James Y. Wang, Kristopher Whitney, Linette Williams  
**IBM, US**

Nigel Griffiths  
**IBM, UK**

## Become a published author

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

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

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

[ibm.com/redbooks/residencies.html](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 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 e-mail 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  
2455 South Road  
Poughkeepsie, NY 12601-5400





# Part 1

# PowerVM virtualization management

Part 1 describes best practices to manage your Power Systems and PowerVM environment.

This chapter gives you a summary of the different PowerVM Editions and the maintenance strategies. It also includes the new functions of the Virtual I/O Server Version 2.1. In the next chapters the following topics are covered:

- ▶ Virtual storage management
- ▶ Virtual I/O Server security
- ▶ Virtual I/O Server maintenance
- ▶ Dynamic operations
- ▶ PowerVM Live Partition Mobility
- ▶ System Planning Tool

|

- ▶ Automated management.



# Introduction

This chapter describes the available PowerVM Editions, and gives an overview of the new Virtual I/O Server Version 2.1 features and PowerVM enhancements.

## 1.1 PowerVM Editions

Virtualization technology is offered in a three editions on Power Systems. All Power Systems servers can utilize standard virtualization functions or logical partitioning (LPAR) technology by using either the Hardware Management Console (HMC) or the Integrated Virtualization Manager (IVM). Logical partitions enable clients to run separate workloads in different partitions on a same physical server. This helps lowering costs and improving energy efficiency. LPARs are designed to be shielded from each other to provide a high level of data security and increased application availability which has been certified by the German Federal Office for Security Information. The complete report and others can visited at:

<http://www.bsi.de/zertifiz/zert/reporte/0461a.pdf>

Dynamic LPAR operations allows clients to dynamically allocate many system resources to application partitions without rebooting, simplifying overall systems administration and workload balancing and enhancing availability.

PowerVM Editions extends the base system functions to include IBM Micro-Partitioning and Virtual I/O Server capabilities, which are designed to allow businesses to increase system utilization, while helping to ensure that applications continue to get the resources they need. Micro-Partitioning technology can help lower costs by allowing the system to be finely tuned to consolidate multiple independent workloads. Micro partitions can be defined as small as 1/10th of a processor and be changed in increments as small as 1/100th of a processor. Up to 10 micro-partitions may be created per core on a server.

The Virtual I/O Server allows for the sharing of expensive disk and optical devices and communications and Fibre Channel adapters to help drive down complexity and systems and administrative expenses. Also included is support for Multiple Shared Processor Pools, which allows for automatic non disruptive balancing of processing power between partitions assigned to the shared pools, resulting in increased throughput and the potential to reduce processor-based software licensing costs and Shared Dedicated Capacity, which helps optimize use of processor cycles.

An uncapped partition enables the processing capacity of that partition to exceed its entitled capacity when the shared processing pool has available resources. This means that idle processor resource within a server can be used by any uncapped partition, resulting in an overall increase of the physical processor resource utilization. However, in an uncapped partition, the total number of virtual processors configured limits the total amount of physical processor resource that the partition can potentially consume.

Using an example, a server has eight physical processors. The uncapped partition is configured with two processing units (equivalent of two physical processors) as its entitled capacity and four virtual processors. In this example, the partition is only ever able to use a maximum of four physical processors. This is because a single virtual processor can only ever consume a maximum equivalent of one physical processor. A dynamic LPAR operation to add more virtual processors would be required to enable the partition to potentially use more physical processor resource.

In the following sections the three different PowerVM Editions are described in detail.

### **PowerVM Express Edition**

PowerVM Express Edition is offered only on the IBM Power 520 and Power 550 servers. It is designed for clients who want to have an introduction to advanced virtualization features at a highly affordable price. With PowerVM Express Edition clients can create up to three partitions on a server (two client partitions and one for the Virtual I/O Server and Integrated Virtualization Manager), leveraging virtualized disk and optical devices and even try out the shared processor pool.

All virtualization features, such as Micro-Partitioning, Shared Processor Pool, Virtual I/O Server, PowerVM LX86, Shared Dedicated Capacity, N Port ID Virtualization, and Virtual Tape can be managed by using the Integrated Virtualization Manager.

### **PowerVM Standard Edition**

For clients ready to get the full value out of their server, IBM offers PowerVM Standard Edition, which provides the most complete virtualization functionality for UNIX® and Linux in the industry. This option is available for all IBM Power Systems servers. With PowerVM Standard Edition clients can create up to 254 partitions on a server, leveraging virtualized disk and optical devices and even try out the shared processor pool. All virtualization features, such as Micro-Partitioning, Shared Processor Pool, Virtual I/O Server, PowerVM LX86, Shared Dedicated Capacity, N Port ID Virtualization, and Virtual Tape can be managed by using an Hardware Management Console or the Integrated Virtualization Manager.

### **PowerVM Enterprise Edition**

PowerVM Enterprise Edition is offered exclusively on POWER6 servers and includes all the features of PowerVM Standard Edition plus a capability named PowerVM Live Partition Mobility. PowerVM Live Partition Mobility allows for the movement of a running partition from one POWER6 technology-based server to another with no application downtime, resulting in better system utilization, improved application availability, and energy savings. With PowerVM Live Partition Mobility, planned application downtime due to regular server maintenance can be a thing of the past.

For more information on PowerVM Live Partition Mobility refer to “PowerVM Live Partition Mobility” on page 293.

Table 1-1 gives describes each component of the PowerVM Editions feature, the editions in which each component is included, and the processor-based hardware on which each component is available.

Table 1-1 PowerVM Editions components, editions, and hardware support

Component	Description	PowerVM Edition	Hardware
Micro-Partitioning technology	The ability to allocate processors to logical partitions in increments of 0.01 allowing multiple logical partitions to share the system's processing power.	<ul style="list-style-type: none"> <li>▶ Express Edition</li> <li>▶ Standard Edition</li> <li>▶ Enterprise Edition</li> </ul>	<ul style="list-style-type: none"> <li>▶ POWER6</li> <li>▶ POWER5</li> </ul>
Virtual I/O Server	Software that facilitates the sharing of physical I/O resources between client logical partitions within the server.	<ul style="list-style-type: none"> <li>▶ Express Edition</li> <li>▶ Standard Edition</li> <li>▶ Enterprise Edition</li> </ul>	<ul style="list-style-type: none"> <li>▶ POWER6</li> <li>▶ POWER5</li> </ul>
Integrated Virtualization Manager	The graphical interface of the Virtual I/O Server management partition on some servers that are not managed by an Hardware Management Console.	<ul style="list-style-type: none"> <li>▶ Express Edition</li> <li>▶ Standard Edition</li> <li>▶ Enterprise Edition</li> </ul>	<ul style="list-style-type: none"> <li>▶ POWER6</li> <li>▶ POWER5</li> </ul>
Live Partition Mobility	The ability to migrate an active or inactive AIX or Linux logical partition from one system to another.	<ul style="list-style-type: none"> <li>▶ Enterprise Edition</li> </ul>	<ul style="list-style-type: none"> <li>▶ POWER6</li> </ul>



Component	Description	PowerVM Edition	Hardware
Partition Load Manager	Software that provides processor and memory resource management and monitoring across AIX logical partitions within a single central processor complex.	▶ Standard Edition	▶ POWER5
Lx86	A product that makes a Power system compatible with x86 applications. This extends the application support for Linux on Power systems, allowing applications that are available on x86 but not on Power systems to be run on the Power system.	▶ Express Edition ▶ Standard Edition ▶ Enterprise Edition	▶ POWER6 running SUSE® or Red Hat Linux

### How to find out which PowerVM Editions feature was ordered

As PowerVM Editions comes in three options: Express, Standard, Enterprise, you can determine the appropriate Edition when reviewing the VET code on the POD web site at:

<http://www-912.ibm.com/pod/pod>

Use bits 25-28 from the VET code listed on the web site. Here is an example of different VET codes:

```
450F28E3D581AF72732400000000041FA
B905E3D284DF097DCA1F00002c0000418F
0F0DA0E9B40C5449CA1F00002c20004102
```

where

0000 = PowerVM Express Edition

2c00 = PowerVM Standard Edition

2c20 = PowerVM Enterprise Edition

## Software licensing

From a software licensing perspective, vendors have different pricing structures on which they license their applications running in an uncapped partition. Because an application has the potential of using more processor resource than the partition's entitled capacity, many software vendors that charge on a per-processor basis require additional processor licenses to be purchased simply based on the possibility that the application might consume more processor resource than it is entitled. When deciding to implement an uncapped partition, check with your software vendor for more information about their licensing terms.

## 1.2 Maintenance strategy

Having a maintenance strategy is very important in any computing environment. It is often recommended to consider guidelines for updates and changes before going into production. Hence when managing a complex virtualization configuration on a Power Systems server running some dozens of partitions managed by different departments or customers, you have to plan maintenance window requests.

There is no ideal maintenance strategy for every enterprise and situation. Each has to be individually developed based on the business availability goals.

PowerVM also offers various techniques to avoid the need for service window requests. PowerVM Live Partition Mobility can be used to move a workload from one server to another without any interruption. Dual Virtual I/O Server configuration allows Virtual I/O Server maintenance without any disruption for clients. Combining Power Systems virtualization and SAN technologies allows you to create flexible and responsive implementation in which any hardware or software can be exchanged and upgraded.

PowerVM demonstrates its ability to be manageable and enterprise-ready for years. Cross-platform tools such as IBM Systems Director, Tivoli® and Cluster Systems Management offer single management interfaces for multiple physical and virtual systems. The Hardware Management Console allows managing multiple virtual systems on Power Systems. The Integrated Virtualization Manager manages virtual systems on a single server.

The advantages provided by virtualization—infrastructure simplification, energy savings, flexibility and responsiveness—also include manageability. Even if the managed virtual environment looks advanced, keep in mind that virtualized environment replaces not a single server but dozens and sometimes hundreds of hard-to-manage, minimally utilized standalone servers.

## 1.3 New features for Virtual I/O Server Version 2.1

IBM PowerVM technology has been enhanced to boost the flexibility of Power Systems servers with support for the following features:

▶ N\_Port ID Virtualization (NPIV)

N\_Port ID Virtualization provides direct access to Fiber Channel adapters from multiple client partitions. It simplifies the management of SAN environments. NPIV support is included with PowerVM Express, Standard, and Enterprise Edition and supports LPARs running AIX 5.3 or AIX 6.1. At the time of writing the Power 520, Power 550, Power 560, and Power 570 servers are supported.

**Note:** IBM intends to support N\_Port ID Virtualization (NPIV) on the POWER6 processor-based Power 595, BladeCenter JS12, and BladeCenter JS22 in 2009. IBM intends to support NPIV with IBM i and Linux environments in 2009.

For more information about NPIV refer to chapter 2.9, “N\_Port ID virtualization” on page 56.

▶ Virtual tape

Two methods for using SAS tape devices are supported:

- Access to SAN tape libraries using shared physical HBA resources through NPIV.
- Virtual tape support allows serial sharing of selected SAS tape devices.

For more information about virtual tape support refer to chapter 2.3, “Using virtual tape devices” on page 20.

**Note:** IBM intends to support Virtual I/O Server virtual tape capabilities on IBM i and Linux environments in 2009.

▶ Enhancements to PowerVM Live Partition Mobility

Two major functions have been added to PowerVM Live Partition Mobility:

- Multiple path support enhances the flexibility and redundancy of Live Partition Mobility.
- PowerVM Live Partition Mobility is now supported in environments with two Hardware Management Consoles (HMC). This enables larger and flexible configurations.

For more information about these new functions refer to “What’s new in PowerVM Live Partition Mobility” on page 294.

▶ Enhancements to PowerVM Lx86

PowerVM Lx86 offers the following additional enhancements:

- Enhanced PowerVM Lx86 installer supports archiving the previously installed environment for backup or migration to other systems.
- Automate installation for non-interactive installation.
- Automate installation from an archive.
- Support installation using the IBM Installation Toolkit for Linux.
- SUSE Linux is supported by PowerVM Lx86 when running on RHEL.

For more information on PowerVM Lx86, visit:

<http://www.ibm.com/systems/power/software/linux>

▶ Enhancements to PowerVM monitoring

New functions have been added to monitor logical volumes statistics, volume group statistics, network and Shared Ethernet Adapter (SEA) statistics and disk service time metrics. A new **topasrec** tool will help you to record monitoring statistics. Also a screen panel has been added to view the Virtual I/O Server and Client throughput.

## 1.4 Other PowerVM Enhancements

There are three additional enhancements for PowerVM which are described as follows:

▶ Active Memory™ Sharing

Active Memory Sharing will intelligently flow memory from one partition to another for increased utilization and flexibility of memory.

**Note:** IBM intends to enhance PowerVM with Active Memory Sharing, an advanced memory virtualization technology, in 2009.

▶ IP Version 6 Support

Beginning with Virtual I/O Server Version 1.5.2 also IP Version 6 is now supported in a Virtual I/O Server partition.

▶ Dual HMC support for PowerVM Live Partition Mobility

PowerVM Live Partition Mobility is now supported in environments with two Hardware Management Consoles supporting larger and more flexible configurations. Partitions may be migrated between systems that are managed by two different HMC. This support is provided for AIX V5.3, AIX V6.1, and Linux partitions on POWER6 processor-based servers.





# Virtual storage management

The Virtual I/O Server maps physical storage to virtual I/O clients. This chapter outlines the best practices for managing disk, tape and optical storage in the virtual environment, keeping track of physical storage and allocating it to virtual I/O clients.

We describe maintenance scenarios such as replacing a physical disk on the Virtual I/O Server which is used as a backing device.

We also include migration scenarios, including moving a partition with virtual storage from one server to another, and moving existing storage (for example, physical or dedicated) into the virtual environment where possible.

## 2.1 Disk mapping options

The Virtual I/O Server presents disk storage to virtual I/O clients as virtual SCSI disks. These virtual disks must be mapped to physical storage by the Virtual I/O Server. There are three ways to perform this mapping, each with its own advantages:

- ▶ Physical volumes
- ▶ Logical volumes
- ▶ File backing devices

The general rule for choosing between these options for dual Virtual I/O Server configurations is that disk devices being accessed through a SAN should be exported as physical volumes, with storage allocation managed in the SAN. Internal and SCSI-attached disk devices should be exported with either logical volumes or storage pools so that storage can be allocated in the server.

### Notes:

For IBM i client partitions we recommend for performance reasons to map dedicated *physical volumes* (hdisks) on the Virtual I/O Server to virtual SCSI client disks.

Up to 16 virtual disk LUNs *and* up to 16 virtual optical LUNs are supported per IBM i virtual SCSI client adapter.

This chapter covers mapping of physical storage to file backed devices. The mapping of physical storage to physical volumes and logical volumes is covered in *PowerVM Virtualization on IBM System p: Introduction and Configuration*, SG24-7490.

### 2.1.1 Physical volumes

The Virtual I/O Server can export physical volumes intact to virtual I/O clients. This method of exporting storage has several advantages over logical volumes:

- ▶ Physical disk devices can be exported from two or more Virtual I/O Servers concurrently for multipath redundancy.
- ▶ The code path for exporting physical volumes is shorter, which might lead to better performance.
- ▶ Physical disk devices can be moved from one Virtual I/O Server to another with relative ease.



- ▶ In some cases, existing LUNs from physical servers can be migrated into the virtual environment with the data intact.
- ▶ One consideration for exporting physical volumes is that the size of the device is not managed by the Virtual I/O Server, and the Virtual I/O Server does not allow partitioning of a single device among multiple clients. This is generally only a concern for internal and SCSI-attached disks.

There is no general requirement to subdivide SAN-attached disks, because storage allocation can be managed at the storage server. In the SAN environment, provision and allocate LUNs for each LPAR on the storage servers and export them from the Virtual I/O Server as physical volumes.

When a SAN disk is available, all storage associated with a virtual I/O client should be stored in the SAN, including rootvg and paging space. This makes management simpler because partitions will not be dependent on both internal logical volumes and external LUNs. It also makes it easier to move LPARs from one Virtual I/O Server to another. For more information, see Chapter 7, “PowerVM Live Partition Mobility” on page 293.

## 2.1.2 Logical volumes

The Virtual I/O Server can export logical volumes to virtual I/O clients. This method does have some advantages over physical volumes:

- ▶ Logical volumes can subdivide physical disk devices between different clients.
- ▶ The logical volume interface is familiar to those with AIX experience.

**Note:** Using the rootvg on the Virtual I/O Server to host exported logical volumes is not recommended. Certain types of software upgrades and system restores might alter the logical volume to target device mapping for logical volumes within rootvg, requiring manual intervention.

When an internal or SCSI-attached disk is used, the logical volume manager (LVM) enables disk devices to be subdivided between different virtual I/O clients. For small servers, this enables several LPARs to share internal disks or RAID arrays.

### Best practices for exporting logical volumes

The Integrated Virtualization Manager (IVM) and HMC-managed environments present two different interfaces for storage management under different names. The storage pool interface under the IVM is essentially the same as the logical volume manager interface under the HMC, and in some cases, the documentation uses the terms interchangeably. The remainder of this chapter

uses the term *volume group* to refer to both volume groups and storage pools, and the term *logical volume* to refer to both logical volumes and storage pool backing devices.

Logical volumes enable the Virtual I/O Server to subdivide a physical volume between multiple virtual I/O clients. In many cases, the physical volumes used will be internal disks, or RAID arrays built of internal disks.

A single volume group should not contain logical volumes used by virtual I/O clients and logical volumes used by the Virtual I/O Server operating system. Keep Virtual I/O Server file systems within the rootvg, and use other volume groups to host logical volumes for virtual I/O clients.

A single volume group or logical volume cannot be accessed by two Virtual I/O Servers concurrently. Do not attempt to configure MPIO on virtual I/O clients for VSCSI devices that reside on logical volumes. If redundancy is required in logical volume configurations, use LVM mirroring on the virtual I/O client to mirror across different logical volumes on different Virtual I/O Servers.

Although logical volumes that span multiple physical volumes are supported, a logical volume should reside wholly on a single physical volume for optimum performance. To guarantee this, volume groups can be composed of single physical volumes.

**Note:** Keeping an exported storage pool backing device or logical volume on a single hdisk results in optimized performance.

When exporting logical volumes to clients, the mapping of individual logical volumes to virtual I/O clients is maintained in the Virtual I/O Server. The additional level of abstraction provided by the logical volume manager makes it important to track the relationship between physical disk devices and virtual I/O clients. For more information, see 2.5, “Managing the mapping of LUNs over vSCSI to hdisks” on page 29.

### Storage pools

When managed by the Integrated Virtualization Manager (IVM), the Virtual I/O Server can export storage pool backing devices to virtual I/O clients. This method is similar to logical volumes, and it does have some advantages over physical volumes:

- ▶ Storage pool backing devices can subdivide physical disk devices between different clients.
- ▶ The storage pool interface is easy to use through IVM.

**Important:** The default storage pool in IVM is the root volume group of the Virtual I/O Server. Be careful not to allocate backing devices within the root volume group because certain types of software upgrades and system restores might alter the logical volume to target device mapping for logical volumes in rootvg, requiring manual intervention.

Systems in a single server environment under the management of IVM are often not attached to a SAN, and these systems typically use internal and SCSI-attached disk storage. The IVM interface allows storage pools to be created on physical storage devices so that a single physical disk device can be divided among several virtual I/O clients.

As with logical volumes, storage pool backing devices cannot be accessed by multiple Virtual I/O Servers concurrently, so they cannot be used with MPIO on the virtual I/O client.

We recommend that the virtual I/O client use LVM mirroring if redundancy is required.

### 2.1.3 File-backed devices

On Version 1.5 of Virtual I/O Server there is a new feature called file-backed virtual SCSI devices. This feature provides additional flexibility for provisioning and managing virtual SCSI devices. In addition to backing a virtual SCSI device (disk or optical) by physical storage, a virtual SCSI device can now be backed to a file. File-backed virtual SCSI devices continue to be accessed as standard SCSI-compliant storage.

**Note:** If LVM mirroring is used in the client, make sure that each mirror copy is placed on a separate disk. It is recommended to mirror across storage pools.

## 2.2 Using virtual optical devices

Virtual optical devices can be accessed by AIX, IBM i, and Linux client partitions.

### Using a virtual optical device on AIX

Chapter 3, Basic Virtual I/O Scenario of *PowerVM on system p Introduction and Configuration*, SG24-7940 describes the setup and management of a virtual optical device such as CD or DVD for AIX client partitions.

**Note:** Both virtual optical devices and virtual tape devices are assigned dedicated server-client pairs. Since the server adapter is configured with the Any client partition can connect option, these pairs are not suited for client disks.

### Using a virtual optical device on IBM i

The following sections show how to dynamically allocate or deallocate an optical device on IBM i virtualized by the Virtual I/O Server and shared between multiple Virtual I/O Server client partitions eliminating the need to use dynamic LPAR to move around any physical adapter resources.

**Important:** An active IBM i partition will by default automatically configure an accessible optical device making it unavailable for usage by other partitions unless the IBM i virtual IOP is disabled using an IOP reset or removed using dynamic LPAR operation. For this reason the IOP should remain disabled when not using the DVD.

Figure 2-1 shows the Virtual I/O Server and client partition virtual SCSI setup for the shared optical device with the Virtual I/O Server owning the physical optical device and virtualizing it via its virtual SCSI server adapter in slot 50 configured for *Any client partition can connect*.

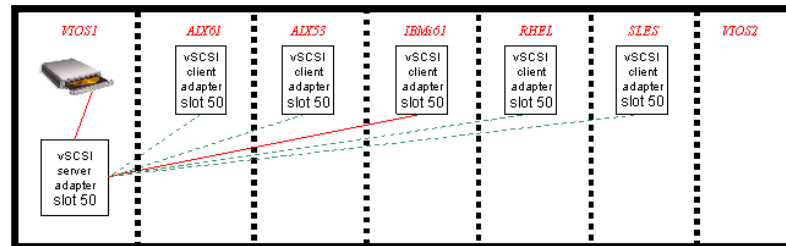


Figure 2-1 SCSI setup for shared optical device

### Allocating a shared optical device on IBM i

1. Use the `WRKHDWRSCT *STG` command to verify the IBM i virtual IOP (type 290A) for the optical device is *operational*.

If it is inoperational as shown in Figure 2-2, locate the logical resource for the virtual IOP in SST Hardware Service Manager and re-IPL the virtual IOP using the **I/O debug** and **IPL I/O processor** option as shown in Figure 2-3 and Figure 2-4.

```
Work with Storage Resources                                     System:E101F170
Type options, press Enter.
  7=Display resource detail  9=Work with resource

Opt Resource          Type-model Status          Text
CMB01          290A-001 Operational      Storage Controller
  DC01          290A-001 Operational      Storage Controller
CMB02          290A-001 Operational      Storage Controller
  DC02          290A-001 Operational      Storage Controller
CMB03          290A-001 Inoperative      Storage Controller
  DC03          290A-001 Inoperative      Storage Controller
CMB05          268C-001 Operational      Storage Controller
  DC05          6B02-001 Operational      Storage Controller

Bottom
F3=Exit  F5=Refresh  F6=Print  F12=Cancel
```

Figure 2-2 IBM i Work with Storage Resources screen

```

Logical Hardware Resources

Type options, press Enter.
  2=Change detail   4=Remove   5=Display detail   6=I/O debug
  7=Verify          8=Associated packaging resource(s)

Opt Description                Type-Model  Status      Resource
6 Virtual IOP                  290A-001   Disabled    CMB03

F3=Exit      F5=Refresh   F6=Print     F9=Failed resources
F10=Non-reporting resources F11=Display serial/part numbers F12=Cancel
CMB03       located successfully.

```

Figure 2-3 IBM i Logical Hardware Resources screen I/O debug option

```

Select IOP Debug Function

Resource name . . . . . : CMB03
Dump type . . . . . : Normal

Select one of the following:

    1. Read/Write I/O processor data
    2. Dump I/O processor data
    3. Reset I/O processor
    4. IPL I/O processor
    5. Enable I/O processor trace
    6. Disable I/O processor trace

Selection
-

F3=Exit      F12=Cancel
F8=Disable I/O processor reset      F9=Disable I/O processor IPL
Re-IPL of IOP was successful.

```

Figure 2-4 IBM i Select IOP Debug Function screen IPL I/O processor option

2. After the IOP is *operational* vary-on the optical drive using the command:

```
VRYCFG CFGOBJ(OPT01) CFGTYPE(*DEV) STATUS(*ON)
```

**Note:** Alternatively to the **VRYCFG** command the corresponding make available/unavailable (vary on/off) options from the Work with Configuration Status screen accessible via the **WRKCFGSTS \*DEV** command can be used.

### ***Deallocating a shared virtual optical device on IBM i***

1. Use the following VRYCFG command to vary-off the optical device from IBM i:

```
VRYCFG CFGOBJ(OPT01) CFGTYPE(*DEV) STATUS(*OFF)
```

2. To release the optical device for usage by other Virtual I/O Server client partitions *disable* its virtual IOP from the SST Hardware Service Manager by locating the logical resource for the virtual IOP first and selecting the **I/O debug** and **Reset I/O processor** option as shown in Figure 2-5

```
Select IOP Debug Function

Resource name . . . . . : CMB03
Dump type . . . . . : Normal

Select one of the following:

    1. Read/Write I/O processor data
    2. Dump I/O processor data
    3. Reset I/O processor
    4. IPL I/O processor
    5. Enable I/O processor trace
    6. Disable I/O processor trace

Selection
    -

F3=Exit      F12=Cancel
F8=Disable I/O processor reset      F9=Disable I/O processor IPL
Reset of IOP was successful.
```

Figure 2-5 IBM i Select IOP Debug Function screen Reset I/O processor option

## Using a virtual optical device on Linux

A virtual optical device can be assigned to a Red Hat or Novell SuSE Linux partition. However, the partition needs to be rebooted to be able to assign the free drive.

Likewise, the partition needs to be shutted down to release the drive.

## 2.3 Using virtual tape devices

Virtual tape devices are assigned and operated similarly to virtual optical devices.

Only one virtual I/O client can have access at a time. The advantage of a virtual tape device is that you do not have to move the parent SCSI adapter between virtual I/O clients.



**Note:** The virtual tape drive cannot be moved to another Virtual I/O Server since client SCSI adapters cannot be created in a Virtual I/O Server. If you want the tape drive in another Virtual I/O Server, the virtual device must be unconfigured and the parent SAS adapter must be unconfigured and moved using dynamic LPAR. The physical tape drive is a SAS device, but mapped to virtual clients as a virtual SCSI device.

If the tape drive is to be used locally in the parent Virtual I/O Server, the virtual device must be unconfigured first.

Figure 2-6 shows the virtual slot setup for virtual tape.

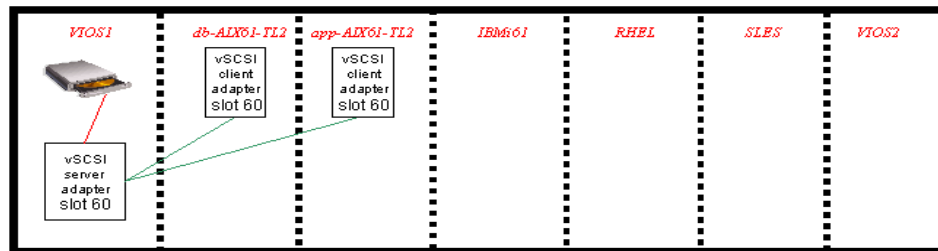


Figure 2-6 SCSI setup for shared tape device

Supported tape drives at the time of writing are:

- ▶ Feature Code 5907: 36/72GB 4mm DAT72 SAS Tape Drive
- ▶ Feature Code 5619: DAT160: 80/160GB DAT160 SAS Tape Drive
- ▶ Feature Code 5746: Half High 800GB/1.6TB LTO4 SAS Tape Drive

#### Notes:

At the time of writing virtual tape is supported in AIX client partitions only. IBM intends to support Virtual I/O Server virtual tape capabilities with IBM i and Linux environments in 2009.

AIX client partitions need to be running AIX Version 5.3 TL9, AIX Version 6.1 TL2 or higher on a POWER6 System for virtual tape support.

SAN Fibre Channel tape drives are supported through N-port ID virtualization (NPIV).

## How to setup a virtual tape drive

The setup steps are as follows:

1. Let the tape drive be assigned to a Virtual I/O Server.
2. Create a virtual SCSI server adapter using the HMC where any partition can connect to.

**Note:** This should not be an adapter shared with disks since it will be removed or unconfigured when not holding the tape drive.

3. Run the `cfgdev` command to configure the new vhost adapter. You can find the new adapter number with the `lsdev -virtual` command.
4. In the Virtual I/O Server, VIOS1, you create the virtual target device with the following command:

```
mkvdev -vdev tape_drive -vadapter vhostn -dev device_name
```

where *n* is the number of the vhost adapter and *device\_name* is the name for the virtual target device. See Example 2-1.

*Example 2-1 Making the virtual device for the tape drive*

---

```
$ mkvdev -vdev rmt0 -vadapter vhost3 -dev vtape
```

---

5. Create a virtual SCSI client adapter in each LPAR using the HMC. The client adapter should point to the server adapter created in the previous step. In the scenario slot 60 is used for the server adapter and slot 60 for all client adapters.

**Tip:** It is useful to use the same slot number for all the clients.

6. In the client, run the `cfgmgr` command that will assign the drive to the LPAR. If the drive is already assigned to another LPAR you will get an error message and you will have to release the drive from the LPAR holding it.

## How to move the virtual tape drive

If your documentation does not provide the vscsi adapter number, you can find it with the `lscfg|grep Cn` command, where *n* is the slot number of the virtual client adapter from the HMC.

1. Use the `rmdev -Rl vscsin` command to change the vscsi adapter and the tape drive to a defined state in the AIX client partition that holds the drive.

**Note:** Adding the `-d` option also removes the adapter from the ODM.

- The `cfgmgr` command in the target LPAR will make the drive available.

**Note:** Provided that the tape drive is not assigned to another LPAR, the drive will show up as an install device in the SMS menu.

## How to find the partition that holds the virtual tape drive

You can use the `dsh` command to find the LPAR currently holding the drive, as shown in Example 2-2. `dsh` is installed by default in AIX. You can use `dsh` with `rsh`, `ssh` or Kerberos authentication as long as `dsh` can run commands without being prompted for a password. When using SSH, a key exchange is required to be able to run commands without being prompted for password. The `dshbak` command sorts the output by target system.

**Note:** Set the `DSH_REMOTE_CMD=/usr/bin/ssh` variable if you use SSH for authentication:

```
# export DSH_REMOTE_CMD=/usr/bin/ssh
# export DSH_LIST=<file listing lpars>
# dsh lsdev -Cc tape | dshbak
```

*Example 2-2 Finding which LPAR is holding the tape drive using dsh*

---

```
# dsh lsdev -Cc tape | dshbak
HOST: app-aix61-TL2
-----
rmt0 Defined Virtual Tape Drive

HOST: db-aix61-TL2
-----
rmt0 Available Virtual Tape Drive
```

---

**Tip:** Put the `DSH_LIST` and `DSH_REMOTE_CMD` definitions in `.profile` on your admin server. You can change the file containing names of target LPARs without redefining `DSH_LIST`.

**Note:** If some partitions do not appear in the list, it is usually because the drive has never been assigned to the partition or completely removed with the `-d` option.

Or use the `ssh` command. See Example 2-3 on page 24.

*Example 2-3 Finding which LPAR is holding the optical drive using ssh*

---

```
# for i in db-aix61-TL2 app-aix61-TL2
> do
> echo $i; ssh $i lsdev -Cc tape
> done
db-aix61-TL2
rmt0 Available Virtual Tape Drive
app-aix61-TL2
rmt0 Defined Virtual Tape Drive
```

---

**Tip:** You can also find the Partition ID of the partition holding the drive from the `lsmap -a11` command on the Virtual I/O Server.

**Note:** AIX6 offers a graphical interface to system management called IBM Systems Console for AIX. This has a menu setup for `dsh`.

## How to unconfigure a virtual tape drive for local use in the Virtual I/O Server

The following are the steps to unconfigure the virtual tape drive when it is going to be used in the Virtual I/O Server for local backups:

1. Release the drive from the partition holding it.
2. Unconfigure the virtual device in the Virtual I/O Server with the `rmdev -dev name -ucfg` command.
3. When finished using the drive locally, use the `cfgdev` command in the Virtual I/O Server to restore the drive as a virtual drive.

## How to unconfigure a virtual tape drive to be moved for local use in another partition

The following are the steps to unconfigure the virtual tape drive in one Virtual I/O Server when it is going to be moved *physically* to another partition and to move it back.

1. Release the drive from the partition holding it.
2. Unconfigure the virtual device in the Virtual I/O Server.
3. Unconfigure the SAS adapter recursively with the `rmdev -dev adapter -recursive -ucfg` command. The correct adapter can be identified with the `lsdev -slots` command.

**Note:** The `lsdev -slots` command will show all adapters that could be subject to dynamic LPAR operations.

4. Use the HMC to move the adapter to the target partition.
5. Run the `cfgmgr` command on an AIX partition or the `cfgdev` command for a Virtual I/O Server partition to configure the drive.
6. When finished, remove the SAS adapter recursively.
7. Use the HMC to move the adapter back to the Virtual I/O Server.
8. Run the `cfgdev` command to configure the drive and the virtual SCSI adapter in the Virtual I/O Server partition. This will make the virtual tape drive available again for client partitions.

## 2.4 Using file-backed devices

With file-backed devices it is possible, for example, to use an ISO image as a virtual device and share it among all the partitions on your system, such as a virtualized optical drive.

The first thing to do is to make sure that the version of Virtual I/O Server is greater than 1.5. The virtual media repository is used to store virtual optical media which can be conceptually inserted into file-backed virtual optical devices. To check this, log into the Virtual I/O Server using the `padmin` user and run the `ioslevel` command. The output of the command should be similar to the one in Example 2-4.

*Example 2-4 Checking the version of the Virtual I/O Server*

---

```
$ ioslevel
2.1.0.1-FP-20.0
```

---

Once you are sure that you are running the right version of the Virtual I/O Server, you can check whether a virtual media repository has already been created. If it has, you can use the `lsrep` command to list and display information about it. If you have output similar to Example 2-5 on page 25, it means that you do not have a virtual media repository set up yet.

*Example 2-5 Checking whether any virtual media repository is already defined*

---

```
$ lsrep
```

---

The DVD repository has not been created yet.

---

Use the **mkrep** command to define it. The command creates the virtual media repository in the specific storage pool. To list the storage pools defined on the Virtual I/O Server, use the **lssp** command. As shown in Example 2-6, there is only one storage pool defined (rootvg). This storage pool can be used to create a virtual media repository with 14 GB (enough space to fit 3 DVD images of 4.7 GB each). After that, recheck the virtual media repository definition with the **lsrep** command.

*Example 2-6 List of available storage pools and defining a virtual media repository*

---

```
$ lssp
Pool                Size(mb)  Free(mb)  Alloc Size(mb)  BDs Type
rootvg              69888    46848    128             0 LVP00L
$ mkrep -sp rootvg -size 14G
Virtual Media Repository Created
Repository created within "VMLibrary_LV" logical volume
$ lsrep
Size(mb) Free(mb) Parent Pool      Parent Size      Parent Free
14277    14277 rootvg             69888            32512
```

---

The next step is to copy the ISO image to the Virtual I/O Server. Secure Copy (SCP) can be used to accomplish this. Once the file is uploaded to the Virtual I/O Server, a virtual optical media disk can be created in the virtual media repository using the **mkvopt** command.

**Note:** By default, the virtual optical disk is created as DVD-RAM media. If the **-ro** flag is specified with the **mkvopt** command, the disk is created as DVD-ROM media.

In Example 2-7, a virtual optical media disk named `ibm_directory_cd1` is created from the `tds61-aix-ppc64-cd1.iso` ISO image located on the `/home/padmin` directory. Using the **-f** flag with the **mkvopt** command will copy the ISO file from its original location into the repository. So after executing this command, the file from the `/home/padmin` directory can be removed since it will be stored on the virtual media repository. The repository configuration can be checked with the **lsrep** command.

*Example 2-7 Creating a virtual optical media disk in the virtual media repository*

---

```
$ mkvopt -name ibm_directory_cd1 -file
/home/padmin/tds61-aix-ppc64-cd1.iso
$ rm tds61-aix-ppc64-cd1.iso
rm: Remove tds61-aix-ppc64-cd1.iso? y
```

```

$ lsrep
Size(mb) Free(mb) Parent Pool          Parent Size      Parent Free
    14278    13913 rootvg                      69888            32512

Name                               File Size Optical
Access
ibm_directory_cd1                   365 None          rw

```

---

Alternatively you can create an ISO image directly from CD/DVD drive as shown in Example 2-8. The default path for created ISO image will be `/var/vio/VMLibrary`. So `file.iso` will be stored as `/var/vio/VMLibrary/file.iso`.

*Example 2-8 Creating an iso image from CD/DVD drive*

---

```

$ mkvopt -name file.iso -dev cd0 -ro
$ lsrep
Size(mb) Free(mb) Parent Pool          Parent Size      Parent Free
    10198    5715 clientvg                 355328           340992

Name                               File Size Optical
Access
file.iso                           4483 None          ro
$ ls /var/vio/VMLibrary/
file.iso  lost+found

```

---

Now that a file-backed virtual optical device has been created, it is necessary to map it to the virtual server adapter. Because it is not possible to use the optical virtual device created in Example 2-7 on page 26 as an input to the `mkvdev` command, a special virtual device is required. This device is a special type of virtual target device called *virtual optical device* and can be created using the `mkvdev` command with the `-fbo` flag. The creation of this new virtual adapter is shown in Example 2-9.

*Example 2-9 Creating an optical virtual target device*

---

```

$ mkvdev -fbo -vadapter vhost1
vtopt0 Available

```

---

The virtual optical device cannot be used until the virtual media is loaded into the device. To load the media, the `loadopt` command is used, as shown in Example 2-10.

*Example 2-10 Loading the virtual media on the virtual target device*

---

```

$ loadopt -disk ibm_directory_cd1 -vtd vtopt0

```

```

$ lsmapi -vadapter vhost1
SVSA          Physloc          Client Partition
ID
-----
vhost1        U9117.MMA.100F6A0-V1-C20  0x00000002

VTD          vnim_rvg
Status       Available
LUN          0x8100000000000000
Backing device hdisk12
Physloc
U789D.001.DQDWWHY-P1-C2-T1-W200400A0B8110D0F-L12000000000000

VTD          vtopt0
Status       Available
LUN          0x8300000000000000
Backing device /var/vio/VMLibrary/ibm_directory_cd1
Physloc
U789D.001.DQDWWHY-P1-C2-T1-W200400A0B8110D0F-L130000000000000

```

Once this is done, you can just go to the client partition and a new CD-ROM unit will appear – on AIX after running the `cfgmgr` command first. If you select this unit as an installation media location, you will be able to see the contents of the ISO file on the Virtual I/O Server and use it as a DVD-RAM unit.

In Example 2-11, the new `cd1` device created in our AIX client is shown. A list of its contents is also provided.

*Example 2-11 Checking the virtual optical device contents on a client*

```

# cfgmgr
# lsdev -C |grep cd
cd0          Defined          Virtual SCSI Optical Served by VIO Server
cd1          Available       Virtual SCSI Optical Served by VIO Server
# lscfg -vl cd1
cd1          U9117.MMA.100F6A0-V2-C20-T1-L830000000000 Virtual SCSI Optical Served by VIO
Server
# installp -L -d /dev/cd1
gskjs:gskjs.rte:7.0.3.30::I:C::::N:AIX Certificate and SSL Java only Base Runtime::::0::
gksa:gksa.rte:7.0.3.30::I:C::::N:AIX Certificate and SSL Base Runtime ACME Toolkit::::0::
gskta:gskta.rte:7.0.3.30::I:C::::N:AIX Certificate and SSL Base Runtime ACME Toolkit::::0::

```

Once you are done with one disk, you might want to insert another disk. In order to do that you need to create a new virtual optical media as described in Example 2-7 on page 26. In this example, three more virtual disk media were created. In Example 2-12 we show how to check which virtual media device is loaded and how to unload a virtual media device (`ibm_directory_cd1`) from the virtual target device (`vtopt0`) and load a new virtual media device (`ibm_directory_cd13`) into it. To unload the media, use the `unloadopt` command



with the **-vtd** flag and the virtual target device. To load a new virtual media, just reuse the **loadopt** command as already described.

*Example 2-12 Loading a new disk on the virtual media device*

---

```

$ lsrep
Size(mb) Free(mb) Parent Pool      Parent Size      Parent Free
   14279   13013 rootvg              69888             32384

Name                File Size Optical      Access
ibm_directory_cd1   365 vtopt0             rw
ibm_directory_cd2   308 None                rw
ibm_directory_cd3   351 None                rw
ibm_directory_cd4   242 None                rw
$ unloadopt -vtd vtopt0
$ loadopt -disk ibm_directory_cd3 -vtd vtopt0
$ lsrep
Size(mb) Free(mb) Parent Pool      Parent Size      Parent Free
   14279   13013 rootvg              69888             32384

Name                File Size Optical      Access
ibm_directory_cd1   365 None                rw
ibm_directory_cd2   308 None                rw
ibm_directory_cd3   351 vtopt0             rw
ibm_directory_cd4   242 None                rw

```

---

It is not required that you have to unload virtual media each time you want to load a new virtual media. You can instead create more than one virtual target devices as shown in Example 2-9 on page 27 and load individual virtual media on new virtual optical devices as shown in Example 2-10 on page 27.

## 2.5 Managing the mapping of LUNs over vSCSI to hdisks

One of the keys to managing a virtual environment is keeping track of what virtual objects correspond to what physical objects. This is particularly challenging in the storage arena where individual LPARs can have hundreds of virtual disks. This mapping is critical to manage performance and to understand what systems will be affected by hardware maintenance.

Virtual disks can be mapped to physical disks in one of two ways (Figure 2-7 on page 30):

- ▶ Physical volumes
- ▶ Logical volumes

Logical volumes can be mapped from volume groups or storage pools.

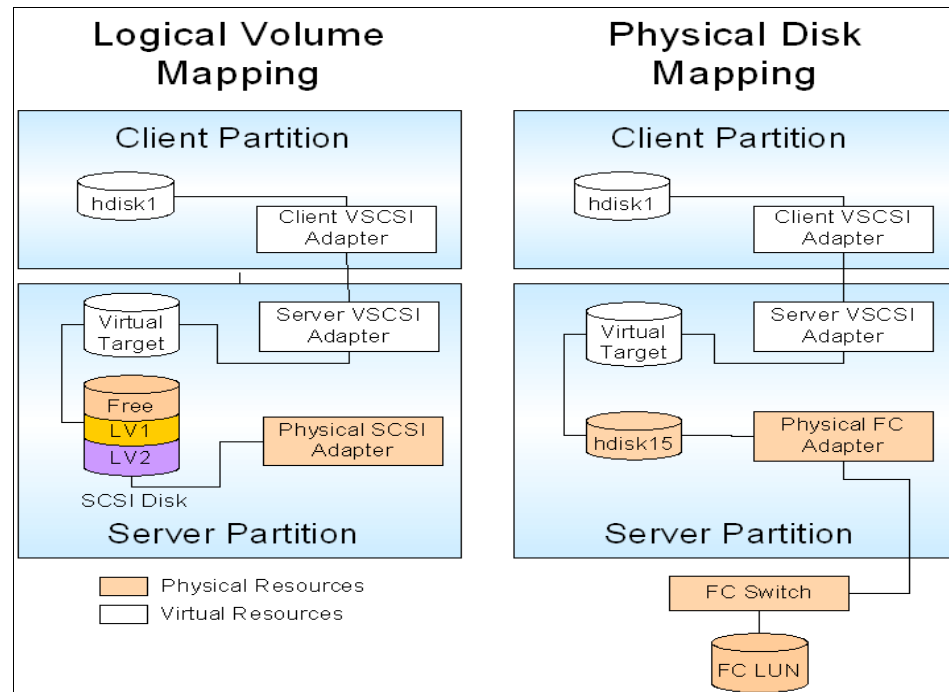


Figure 2-7 Logical versus physical drive mapping

Depending on which method you choose, you might need to track the following information:

- ▶ Virtual I/O Server
  - Server host name
  - Physical disk location
  - Physical adapter device name
  - Physical hdisk device name
  - Volume group or storage pool name<sup>1</sup>
  - Logical volume or storage pool backing device name<sup>1</sup>
  - Virtual SCSI adapter slot
  - Virtual SCSI adapter device name
  - Virtual target device
- ▶ Virtual I/O client
  - Client host name
  - Virtual SCSI adapter slot
  - Virtual SCSI adapter device name

<sup>1</sup> For logical volume or storage pool export only

- Virtual disk device name

The System Planning Tool (SPT) is recommended for planning and documenting your configuration. The SPT system plan can be deployed through an HMC or the Integrated Virtualization Manager (IVM) and ensures correct naming and numbering. For more information about deploying a SPT system plan refer to Chapter 8, “System Planning Tool” on page 305.

## 2.5.1 Naming conventions

A good naming convention is key to managing information. One strategy for reducing the amount of data that must be tracked is to make settings match on the virtual I/O client and server wherever possible.

This can include corresponding volume group, logical volume, and virtual target device names. Integrating the virtual I/O client host name into the virtual target device name can simplify tracking on the server.

When using Fibre Channel disks on a storage server that supports LUN naming, this feature can be used to make it easier to identify LUNs. Commands such as **lssdd** for the IBM System Storage™ DS8000™ and DS6000™ series storage servers, and the **fget\_config** or **mpio\_get\_config** command for the DS4000™ series can be used to match hdisk devices with LUN names.

Prior to Virtual I/O Server version 2.1, you can use the **fget\_config** command as shown in Example 2-13 on page 31. Because the **fget\_config** command is part of a storage device driver, you must use the **oem\_setup\_env** command.

*Example 2-13 The fget\_config command for the DS4000 series*

---

```
$ oem_setup_env
# fget_config -Av

---dar0---
```

User array name = 'FAST200'  
dac0 ACTIVE dac1 ACTIVE

Disk	DAC	LUN	Logical Drive
utm		31	
hdisk4	dac1	0	Server1_LUN1
hdisk5	dac1	1	Server1_LUN2
hdisk6	dac1	2	Server-520-2-LUN1
hdisk7	dac1	3	Server-520-2-LUN2

---

In many cases, using LUN names can be simpler than tracing devices using Fibre Channel world wide port names and numeric LUN identifiers.

The Virtual I/O Server version 2.1 uses MPIO as default device driver. Example 2-14 shows the listing of a DS4800 disk subsystem. Proper User Label naming in the SAN makes it much easier to track the LUN to hdisk relation.

*Example 2-14 SAN storage listing on the Virtual I/O Server version 2.1*

---

```

$ oem_setup_env
# mpio_get_config -Av
Frame id 0:
  Storage Subsystem worldwide name: 60ab800114632000048ed17e
  Controller count: 2
  Partition count: 1
  Partition 0:
    Storage Subsystem Name = 'ITS0_DS4800'
    hdisk      LUN #  Ownership      User Label
    hdisk6     0    A (preferred)  VIOS1
    hdisk7     1    A (preferred)  AIX61
    hdisk8     2    B (preferred)  AIX53
    hdisk9     3    A (preferred)  SLES10
    hdisk10    4    B (preferred)  RHEL52
    hdisk11    5    A (preferred)  IBMi61_0
    hdisk12    6    B (preferred)  IBMi61_1
    hdisk13    7    A (preferred)  IBMi61_0m
    hdisk14    8    B (preferred)  IBMi61_1m

```

---

## 2.5.2 Virtual device slot numbers

After establishing the naming conventions, also establish slot numbering conventions for the virtual I/O adapters.

All Virtual SCSI and Virtual Ethernet devices have slot numbers. In complex systems, there will tend to be far more storage devices than network devices because each virtual SCSI device can only communicate with one server or client. For this reason it is recommended to reserve lower slot numbers for Ethernet adapters. It is also recommended to allow for growth both for Ethernet and SCSI to avoid mixing slot number ranges.

Virtual I/O Servers typically have a lot more virtual adapters than client partitions.

**Note:** Several disks can be mapped to the same server-client SCSI adapter pair.

Management can be simplified by keeping slot numbers consistent between the virtual I/O client and server. However, when partitions are moved from one server to another, this might not be possible. In environments with only one Virtual I/O Server, add storage adapters incrementally starting with slot 21 and higher. When clients are attached to two Virtual I/O Servers, the adapter slot numbers should be alternated from one Virtual I/O Server to the other. The first Virtual I/O Server should use odd numbered slots starting at 21, and the second should use even numbered slots starting at 22. In a two-server scenario, allocate slots in pairs, with each client using two adjacent slots such as 21 and 22, or 33 and 34.

Set the maximum virtual adapters number to 100 as shown in Figure 2-8, the default value is 10 when you create an LPAR. The appropriate number for your environment depends on the number of LPARs and adapters expected on each system. Each unused virtual adapter slot consumes a small amount of memory, so the allocation should be balanced. Use the System Planning Tool available from the following URL to plan memory requirements for your system configuration:

<http://www.ibm.com/servers/eserver/series/lpar/systemdesign.html>

**Important:**

When planning for the number of virtual I/O slots on your LPAR, the maximum number of virtual adapter slots available on a partition is set by the partition's profile. To increase the maximum number of virtual adapters you have to change the profile, stop the partition (not just a reboot) and start the partition. It is recommended to leave plenty of room for expansion when setting the maximum number of slots so that new virtual I/O clients can be added without shutting down the LPAR or Virtual I/O Server partition.

The maximum number of virtual adapters should not be set higher than 1024.

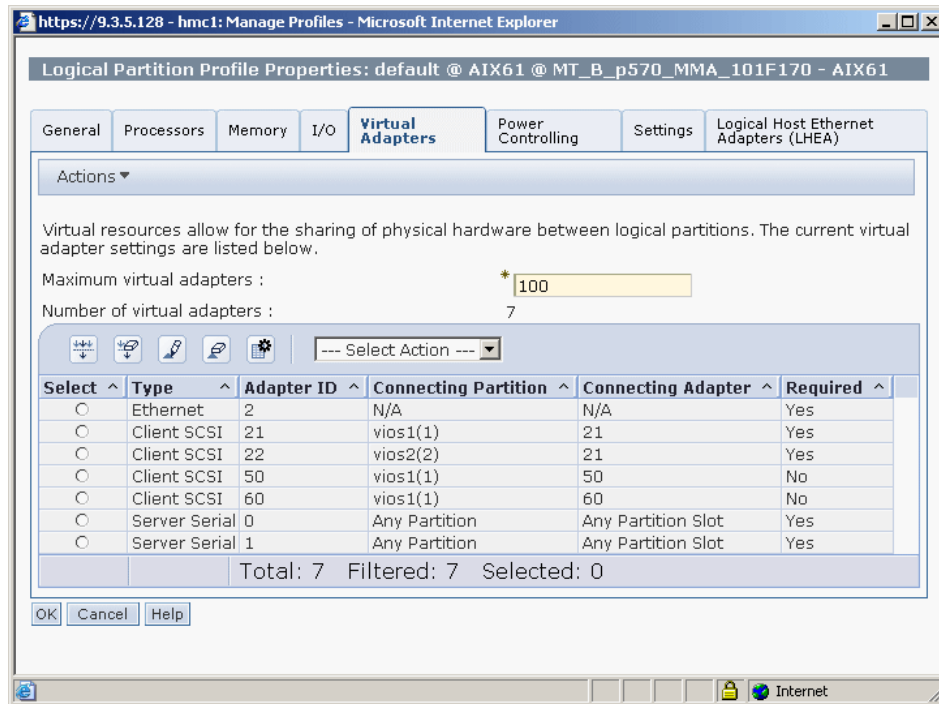


Figure 2-8 Setting maximum number of virtual adapters in a partition profile

Because virtual SCSI connections operate at memory speed, there is generally no performance gain from adding multiple adapters between a Virtual I/O Server and client. For AIX virtual I/O client partitions each adapter pair can handle up to 85 virtual devices with the default queue depth of three, for IBM i clients up to 16 virtual disk and 16 optical devices are supported. In situations where virtual devices per partition are expected to exceed that number, or where the queue depth on some devices might be increased above the default, reserve additional adapter slots – for the Virtual I/O Server and the virtual I/O client partition. When tuning queue depths, the VSCSI adapters have a fixed queue depth. There are 512 command elements of which 2 are used by the adapter, 3 are reserved for each VSCSI LUN for error recovery and the rest are used for I/O requests. Thus, with the default queue depth of 3 for VSCSI LUNs, that allows for up to 85 LUNs to use an adapter:  $(512 - 2) / (3 + 3) = 85$  rounding down. So if we need higher queue depths for the devices, then the number of LUNs per adapter is reduced. For Example, if we want to use a queue depth of 25, that allows  $510/28 = 18$  LUNs per adapter for an AIX client partition.

## 2.5.3 Tracing a configuration

Despite the best intentions in record keeping, it sometimes becomes necessary to manually trace a client virtual disk back to the physical hardware.

### AIX virtual storage configuration tracing

AIX virtual storage (including NPIV) can be traced from Virtual I/O Server using the **lsmmap** command. In Example 2-15 tracing of virtual SCSI storage is shown from the Virtual I/O Server.

*Example 2-15 Tracing virtual SCSI storage from Virtual I/O Server*

```

$ lsmmap -all
SVSA          Physloc          Client Partition
ID
-----
vhost0       U9117.MMA.101F170-V1-C21  0x00000003

VTD          aix61_rvg
Status       Available
LUN          0x8100000000000000
Backing device hdisk7
Physloc
U789D.001.DQDYKYW-P1-C2-T2-W201300A0B811A662-L1000000000000

SVSA          Physloc          Client Partition
ID
-----
vhost1       U9117.MMA.101F170-V1-C22  0x00000004

VTD          NO VIRTUAL TARGET DEVICE FOUND
:
:

```

Example 2-16 shows how to trace NPIV storage devices from the Virtual I/O Server. **ClntID** shows the LPAR ID as seen from the HMC and **ClntName** depicts the hostname. For more information on NPIV refer to section 2.9, “N\_Port ID virtualization” on page 56.

*Example 2-16 Tracing NPIV virtual storage from the Virtual I/O Server*

```

$ lsmmap -npiv -all
Name          Physloc          ClntID ClntName      ClntOS
=====
vfchost0     U9117.MMA.101F170-V1-C31  3 AIX61        AIX

```

```

Status:LOGGED_IN
FC name:fcs3                      FC loc code:U789D.001.DQDYKYW-P1-C6-T2
Ports logged in:2
Flags:a<LOGGED_IN,STRIP_MERGE>
VFC client name:fcs2              VFC client DRC:U9117.MMA.101F170-V3-C31-T1

Name          Physloc                      CIntID CIntName      CIntOS
=====
vfchost1     U9117.MMA.101F170-V1-C32          4
=====

Status:NOT_LOGGED_IN
FC name:fcs3                      FC loc code:U789D.001.DQDYKYW-P1-C6-T2
Ports logged in:0
Flags:4<NOT_LOGGED>
VFC client name:                  VFC client DRC:
:
:

```

---

The IBM Systems Hardware Information Center contains a guide to tracing virtual disks, available at:

[http://publib.boulder.ibm.com/infocenter/systems/scope/hw/index.jsp?topic=/iphb1/iphb1\\_vios\\_managing\\_mapping.htm](http://publib.boulder.ibm.com/infocenter/systems/scope/hw/index.jsp?topic=/iphb1/iphb1_vios_managing_mapping.htm)

### IBM i Virtual SCSI disk configuration tracing

Virtual LUNs always show up as device type 6B22 model 050 on the IBM i client regardless of which storage subsystem ultimately provides the backing physical storage. Figure 2-9 shows an example of the disk configuration from a new IBM i client after SLIC install set up for mirroring its disk units across two Virtual I/O Servers.



```
Display Disk Configuration Status

      Serial
ASP Unit  Number      Type Model Name      Status
  1
    1 Y3WUTVVQMM4G    6B22 050 DD001      Active
    1 YYUUH3U9UELD    6B22 050 DD004      Resume Pending
    2 YD598QUY5XR8    6B22 050 DD003      Active
    2 YTM3C79KY4XF    6B22 050 DD002      Resume Pending

Press Enter to continue.

F3=Exit      F5=Refresh      F9=Display disk unit details
F11=Disk configuration capacity  F12=Cancel
```

Figure 2-9 IBM i SST Display Disk Configuration Status screen

In order to trace down the IBM i disk units to the corresponding SCSI devices on the Virtual I/O Server we look at the disk unit details information as shown in Figure 2-10.

Display Disk Unit Details

Type option, press Enter.  
5=Display hardware resource information details

OPT	ASP	Unit	Serial Number	Sys Bus	Sys Card	I/O Adapter	I/O Bus	Ctl	Dev	Compressed
1	1	1	Y3WUTVVQMM4G	255	21		0	1	0	No
1	1	1	YYUUH3U9UELD	255	22		0	2	0	No
1	2	2	YD598QUY5XR8	255	21		0	2	0	No
1	2	2	YTM3C79KY4XF	255	22		0	1	0	No

F3=Exit                      F9=Display disk units                      F12=Cancel

Figure 2-10 IBM i SST Display Disk Unit Details screen

**Note:** To trace down an IBM i virtual disk unit to the corresponding virtual target device (VTD) and backing hdisk on the Virtual I/O Server use the provided system card Sys Card and controller Ctl information from the IBM i client as follows:

- ▶ Sys Card shows the IBM i virtual SCSI client adapter slot as configured in the IBM i partition profile
- ▶ Ctl XOR 0x80 corresponds to the virtual target device LUN information on the Virtual I/O Server

In the following example we illustrate tracing the IBM i mirrored load source (disk unit 1) reported on IBM i at Sys Card 21 Ctl 1 and Sys Card 22 Ctl 2 down to the devices on the two Virtual I/O Servers and the SAN storage system:

1. We first look at the virtual adapters mapping in the IBM i partition properties on the HMC as shown in Figure 2-11. Since the Sys Card 21 and 22 information from the IBM i client corresponds to the virtual SCSI adapter slot numbers, the partition properties information shows us that the IBM i client

virtual SCSI client adapters 21 and 22 connect to the virtual SCSI server adapters 23 and 23 of the Virtual I/O Server partitions vios1 and vios2.

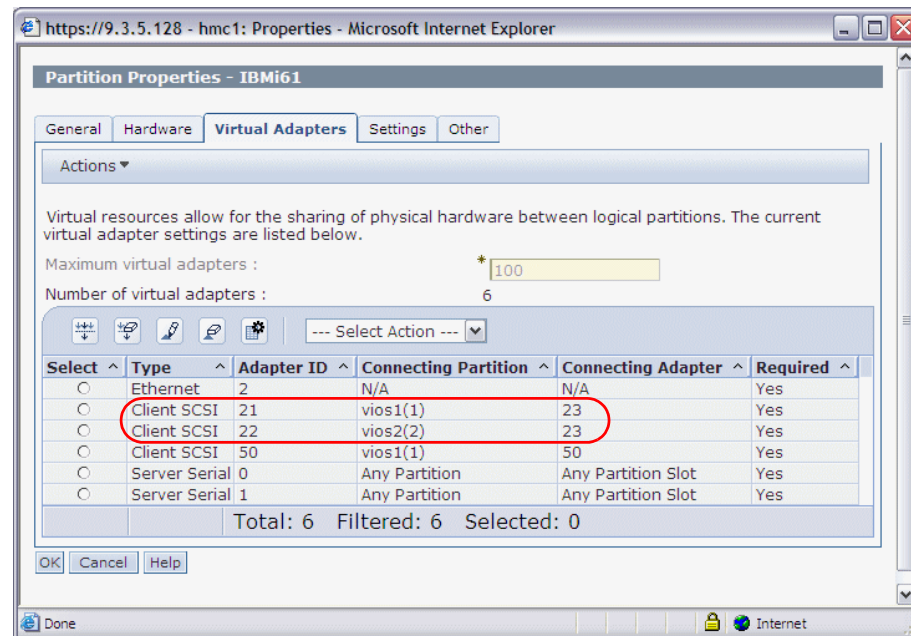


Figure 2-11 IBM i partition profile virtual adapters configuration

- Knowing the corresponding virtual SCSI server adapter slots 23 and 23 we can look at the device mapping on our two Virtual I/O Servers. Using the **lsmmap** command on our Virtual I/O Server vios1 we see the device mapping between physical and virtual devices as shown in Example 2-17.

*Example 2-17 Displaying the Virtual I/O Server device mapping*

```
$ lsdev -slots
# Slot                Description          Device(s)
U789D.001.DQDYKYW-P1-T1 Logical I/O Slot    pci4 usbh0c usbh1
U789D.001.DQDYKYW-P1-T3 Logical I/O Slot    pci3 sissas0
U9117.MMA.101F170-V1-C0 Virtual I/O Slot    vsa0
U9117.MMA.101F170-V1-C2 Virtual I/O Slot    vasi0
U9117.MMA.101F170-V1-C11 Virtual I/O Slot    ent2
U9117.MMA.101F170-V1-C12 Virtual I/O Slot    ent3
U9117.MMA.101F170-V1-C13 Virtual I/O Slot    ent4
U9117.MMA.101F170-V1-C21 Virtual I/O Slot    vhost0
U9117.MMA.101F170-V1-C22 Virtual I/O Slot    vhost1
U9117.MMA.101F170-V1-C23 Virtual I/O Slot    vhost2
U9117.MMA.101F170-V1-C24 Virtual I/O Slot    vhost3
U9117.MMA.101F170-V1-C25 Virtual I/O Slot    vhost4
```

```

U9117.MMA.101F170-V1-C50 Virtual I/O Slot vhost5
U9117.MMA.101F170-V1-C60 Virtual I/O Slot vhost6

$ lsmmap -vadapter vhost2
SVSA          Physloc          Client Partition
ID
-----
vhost2        U9117.MMA.101F170-V1-C23    0x00000005

VTD          IBMi61_0
Status       Available
LUN          0x8100000000000000
Backing device hdisk11
Physloc
U789D.001.DQDYKYW-P1-C2-T2-W201300A0B811A662-L5000000000000

VTD          IBMi61_1
Status       Available
LUN          0x8200000000000000
Backing device hdisk12
Physloc
U789D.001.DQDYKYW-P1-C2-T2-W201300A0B811A662-L6000000000000

```

3. Remembering our IBM i client disk unit 1 connected to Sys Card 21 connected to vios1 showed Ctl 1 we find the corresponding virtual target device LUN on the Virtual I/O Server as follows: *Ctl 1 XOR 0x80 = 0x81*, i.e. LUN 0x81 which is backed by hdisk11 corresponds to the disk unit 1 of our IBM i client whose mirror side is connected to vios1.
4. To further trace down hdisk11 on the Virtual I/O Server vios1 to its physical device respectively LUN on the SAN storage system we use the **lsdev** command to see its kind of multipath device, then knowing it is a MPIO device we use the **mpio\_get\_config** command as shown in Example 2-18 to finally figure out that our IBM i disk unit 1 corresponds to LUN 5 on our DS4800 storage subsystem.

---

*Example 2-18 Virtual I/O Server hdisk to LUN tracing*

---

```

$ lsdev -dev hdisk11
name          status      description
hdisk11       Available  MPIO Other DS4K Array Disk

$ oem_setup_env
# mpio_get_config -Av
Frame id 0:
  Storage Subsystem worldwide name: 60ab800114632000048ed17e
  Controller count: 2
  Partition count: 1
  Partition 0:

```

```
Storage Subsystem Name = 'ITSO_DS4800'
  hdisk      LUN #  Ownership      User Label
  hdisk6     0    A (preferred)  VIOS1
  hdisk7     1    A (preferred)  AIX61
  hdisk8     2    B (preferred)  AIX53
  hdisk9     3    A (preferred)  SLES10
  hdisk10    4    B (preferred)  RHEL52
  hdisk11    5    A (preferred) IBMi61_0
  hdisk12    6    B (preferred)  IBMi61_1
  hdisk13    7    A (preferred)  IBMi61_0m
  hdisk14    8    B (preferred)  IBMi61_1m
```

---

## 2.6 Replacing a disk on the Virtual I/O Server

If it becomes necessary to replace a disk on the Virtual I/O Server, you must first identify the virtual I/O clients affected and the target disk drive.

**Note:** Before replacing a disk device using this procedure, check that the disk can be hotswapped.

If you run in a single Virtual I/O Server environment without disk mirroring on the virtual I/O clients, replacement of a non-RAID protected physical disk requires data to be restored. This also applies if you have the same disk exported through two Virtual I/O Servers using MPIO. MPIO by itself does not protect against outages due to disk replacement. You should evaluate protecting the data on a disk or LUN using either mirroring or RAID technology.

This section covers the following disk replacement procedures in a dual Virtual I/O Server environment using software mirroring on the client:

- ▶ 2.6.1, “Replacing a LV backed disk in the mirroring environment” on page 41
- ▶ 2.6.2, “Replacing a mirrored storage pool backed disk” on page 47

### 2.6.1 Replacing a LV backed disk in the mirroring environment

In the logical volume (LV) backed mirroring scenario we want to replace hdisk2 on the Virtual I/O Server which is part of its volume group `vioc_rootvg_1` and which contains the LV `vioc_1_rootvg` mapped to the virtual target device `vtscsi0`. It has the following attributes:

- ▶ The size is 32 GB.
- ▶ The virtual disk is software mirrored on the virtual I/O client.

- ▶ The failing disk on the AIX virtual I/O client is hdisk1.
- ▶ The virtual SCSI adapter on the virtual I/O client is vscsi1.
- ▶ The volume group on the AIX virtual I/O client is rootvg.

Figure 2-12 shows the setup using an AIX virtual I/O client as an example however the following replacement procedure covers an AIX client as well as an IBM i client.

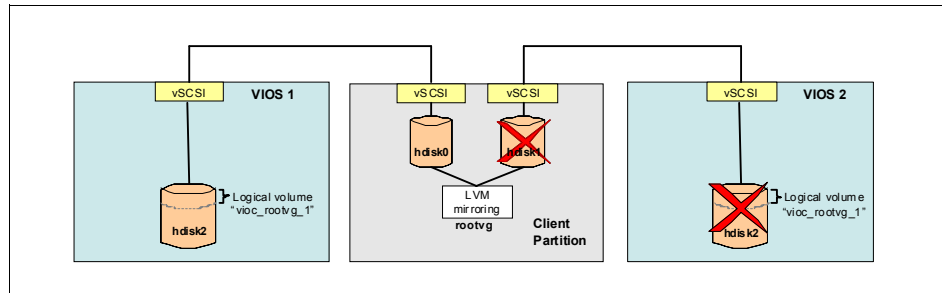


Figure 2-12 AIX LVM mirroring environment with LV backed virtual disks

Check the state of the client disk first for an indication that a disk replacement might be required:

- ▶ On the AIX client use the `lsvg -pv volume group` command to check if the PV STATE information shows *missing*.
- ▶ On the IBM i client enter the **STRSST** command and login to System Service Tools (SST) selecting the options **3. Work with disk units** → **1. Display disk configuration** → **1. Display disk configuration status** to check if a mirrored disk unit shows *suspended*.

**Note:** Before replacing the disk, document the virtual I/O client, logical volume (LV), backing devices, vhost and vtscsi associated, and the size of the LV mapped to the vtscsi device. See 2.5, “Managing the mapping of LUNs over vSCSI to hdisks” on page 29 for more information about managing this.

### ***Procedure to replace a physical disk on the Virtual I/O Server***

1. Identify the physical disk drive with the `diagmenu` command.
2. Then, select **Task Selection (Diagnostics, Advanced Diagnostics, Service Aids, etc.)**. In the next list, select **Hot Plug Task**.
3. In this list, select **SCSI and SCSI RAID Hot Plug Manager** and select **Identify a Device Attached to a SCSI Hot Swap Enclosure Device**.

4. In the next list, find the hdisk and press Enter. A window similar to the one in Example 2-19 opens. Note that this is an example from a p5-570 with internal disks.

*Example 2-19 Find the disk to remove*

---

```
IDENTIFY DEVICE ATTACHED TO SCSI HOT SWAP ENCLOSURE DEVICE
802483
```

The following is a list of devices attached to SCSI Hot Swap Enclosure devices.

Selecting a slot will set the LED indicator to Identify.

Make selection, use Enter to continue.

```
[MORE...4]
  slot 3+-----+
  slot 4|
  slot 5|
  slot 6| The LED should be in the Identify state for the
        | selected device.
ses1   | Use 'Enter' to put the device LED in the
  slot 1| Normal state and return to the previous menu.
  slot 2|
  slot 3|
  slot 4|
  slot 5|
  slot 6|
[BOTTOM] | F3=Cancel      F10=Exit      Enter
F1=Help  +-----+
```

---

5. For an AIX virtual I/O client:

- a. Unmirror the rootvg, as follows:

```
# unmirrorvg -c 1 rootvg hdisk1
0516-1246 rmlvcopy: If hd5 is the boot logical volume, please run
'chpv -c <diskname>' as root user to clear the boot record and
avoid a potential boot off an old boot image that may reside on
the disk from which this logical volume is moved/removed.
```

```
0301-108 mkboot: Unable to read file blocks. Return code: -1
0516-1132 unmirrorvg: Quorum requirement turned on, reboot system
for this to take effect for rootvg.
```

0516-1144 unmirrorvg: rootvg successfully unmirrored, user should perform bosboot of system to reinitialize boot records. Then, user must modify bootlist to just include: hdisk0.

- b. Reduce the AIX client rootvg:

```
# reducevg rootvg hdisk1
```

- c. Remove hdisk1 device from the AIX client configuration:

```
# rmdev -l hdisk1 -d
hdisk1 deleted
```

6. On the Virtual I/O Server, remove the vtscsi/vhost association:

```
$ rmdev -dev vtscsi0
vtscsi0 deleted
```

7. On the Virtual I/O Server, reduce the volume group. If you get an error, as in the following example, and you are sure that you have only one hdisk per volume group, you can use the **deactivatevg** and **exportvg** commands.

**Note:** If you use the **exportvg** command, it will delete all the logical volumes inside the volume group's ODM definition, and if your volume group contains more than one hdisk, the logical volumes on this hdisk are also affected. Use the **lspv** command to check. In this case, it is safe to use the **exportvg vioc\_rootvg\_1** command:

```
$ lspv
NAME          PVID          VG          STATUS
hdisk0        00c478de00655246  rootvg     active
hdisk1        00c478de008a399b  rootvg     active
hdisk2        00c478de008a3ba1  vioc_rootvg_1  active
hdisk3        00c478deb4b0d4b0  None
```

```
$ reducevg -rmlv -f vioc_rootvg_1 hdisk2
```

Some error messages may contain invalid information for the Virtual I/O Server environment.

```
0516-062 lqueryvg: Unable to read or write logical volume manager
record. PV may be permanently corrupted. Run diagnostics
0516-882 reducevg: Unable to reduce volume group.
$ deactivatevg vioc_rootvg_1
```

Some error messages may contain invalid information for the Virtual I/O Server environment.



```
0516-062 lqueryvg: Unable to read or write logical volume manager
record. PV may be permanently corrupted. Run diagnostics
$ exportvg vioc_rootvg_1
```

8. On the Virtual I/O Server, remove the hdisk device:

```
$ rmdev -dev hdisk2
hdisk2 deleted
```

9. Replace the physical disk drive.

10. On the Virtual I/O Server, configure the new hdisk device with the **cfgdev** command and check the configuration using the **lspv** command to determine that the new disk is configured:

```
$ cfgdev
$ lspv
NAME                PVID                VG                STATUS
hdisk2             none               None
hdisk0              00c478de00655246   rootvg           active
hdisk1              00c478de008a399b   rootvg           active
hdisk3              00c478deb4b0d4b0   None
```

11. On the Virtual I/O Server, extend the volume group with the new hdisk using the **mkvg** command if you only have one disk per volume group. Use the **extendvg** command if you have more disks per volume group. In this case, we have only one volume group per disk, which is recommended. If the disk has a PVID, use the **-f** flag on the **mkvg** command.

```
$ mkvg -vg vioc_rootvg_1 hdisk2
vioc_rootvg_1
0516-1254 mkvg: Changing the PVID in the ODM.
```

**Note:** Alternatively to the manual disk replacement steps 7 to 11 on the Virtual I/O Server the **replphyvol** command may be used.

12. On the Virtual I/O Server, recreate the logical volume of exactly the original size for the vtscsi device:

```
$ mklv -lv vioc_1_rootvg vioc_rootvg_1 32G
vioc_1_rootvg
```

**Note:** For an IBM i client partition do not attempt to determine the size for the virtual target device from the IBM i client partition because due to the 8-to-9 sector conversion (520 byte sectors on IBM i vs. 512 byte sectors on the Virtual I/O Server) the IBM i client shows less capacity for the disk unit than what actually needs to be configured on the Virtual I/O Server.

If you haven't noted down the size for the LV before the Virtual I/O Server disk failure determine the size from the output of the `lslv logicalvolume` command on the Virtual I/O Server of the active mirror side by multiplying the logical partitions (LPs) with the physical partition size (PP SIZE) information.

13. On the Virtual I/O Server, check that the LV does not span disks:

```
$ lslv -pv vioc_1_rootvg
vioc_1_rootvg:N/A
PV          COPIES          IN BAND          DISTRIBUTION
hdisk2      512:000:000    42%              000:218:218:076:000
```

14. On the Virtual I/O Server recreate the virtual device:

```
$ mkvdev -vdev vioc_1_rootvg -vadapter vhost0
vtscsi0 Available
```

15. For an AIX virtual I/O client:

- a. Reconfigure the new hdisk1 – if the parent device is unknown then the `cfgmgr` command can be executed without any parameters:

```
# cfgmgr -l vscsi1
```

- b. Extend the rootvg:

```
# extendvg rootvg hdisk1
0516-1254 extendvg: Changing the PVID in the ODM.
```

- c. Mirror the rootvg again:

```
# mirrorvg -c 2 rootvg hdisk1
0516-1124 mirrorvg: Quorum requirement turned off, reboot system
for this to take effect for rootvg.
0516-1126 mirrorvg: rootvg successfully mirrored, user should
perform bosboot of system to initialize boot records. Then, user
must modify bootlist to include: hdisk0 hdisk1.
```

- d. Initialize boot records and set the bootlist:

```
# bosboot -a
bosboot: Boot image is 18036 512 byte blocks.
# bootlist -m normal hdisk0 hdisk1
```

16. For an IBM i client verify in SST the previously suspended disk unit automatically changed its mirrored state to *resuming* and finally *active* when the mirror re-synchronization completed.

## 2.6.2 Replacing a mirrored storage pool backed disk

In the storage pool backed mirroring scenario we want to replace hdisk2 on the Virtual I/O Server in the storage pool vioc\_rootvg\_1 which contains the backing device vioc\_1\_rootvg associated to vhost0. It has the following attributes:

- ▶ The size is 32 GB.
- ▶ The virtual disk is software mirrored on the virtual I/O client.
- ▶ The volume group on the AIX virtual I/O client is rootvg.

**Note:** The *storage pool* and *backing device* concept on the Virtual I/O Server is similar to the volume group and logical volume concept known from AIX but hides some of its complexity.

Figure 2-13 shows the setup using an AIX virtual I/O client as an example however the following replacement procedure covers an AIX client as well as an IBM i client.

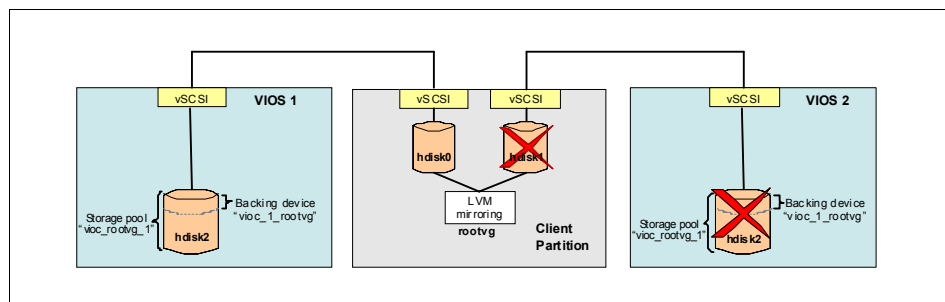


Figure 2-13 AIX LVM mirroring environment with storage pool backed virtual disks

Check the state of the client disk first for an indication that a disk replacement might be required:

- ▶ On the AIX client use the `lsvg -pv volumegroup` command to check if the PV STATE information shows *missing*.
- ▶ On the IBM i client enter the **STRSST** command and login to System Service Tools (SST) selecting the options **3. Work with disk units** → **1. Display disk configuration** → **1. Display disk configuration status** to check if a mirrored disk unit shows *suspended*.

**Note:** Before replacing the disk, document the virtual I/O client, logical volume (LV), backing devices, vhost and vtscsi associated, and the size of the backing device. See 2.5, “Managing the mapping of LUNs over vSCSI to hdisks” on page 29 for more information about managing this.

### ***Procedure to replace a physical disk on the Virtual I/O Server***

1. Identify the physical disk drive; see step 1 on page 42.
2. For an AIX virtual I/O client:
  - a. Unmirror the rootvg as follows:

```
# unmirrorvg -c 1 rootvg hdisk1
0516-1246 rmlvcopy: If hd5 is the boot logical volume, please run 'chpv -c
<diskname>'
    as root user to clear the boot record and avoid a potential boot
    off an old boot image that may reside on the disk from which this
    logical volume is moved/removed.
0301-108 mkboot: Unable to read file blocks. Return code: -1
0516-1132 unmirrorvg: Quorum requirement turned on, reboot system for this
    to take effect for rootvg.
0516-1144 unmirrorvg: rootvg successfully unmirrored, user should perform
    bosboot of system to reinitialize boot records. Then, user must
modify
    bootlist to just include: hdisk0.
```
  - b. Reduce the AIX client rootvg:

```
# reducevg rootvg hdisk1
```
  - c. Remove hdisk1 device from the AIX client configuration:

```
# rmdev -l hdisk1 -d
hdisk1 deleted
```
3. On the Virtual I/O Server remove the backing device:

```
$ rmbdsp -bd vioc_1_rootvg
vtscsi0 deleted
```
4. Remove the disk from the disk pool. If you receive an error message, such as in the following example, and you are sure that you have only one hdisk per storage pool, you can use the **deactivatevg** and **exportvg** commands.

**Note:** If you use the `exportvg` command, it will delete all the logical volumes inside the volume group, and if your volume group contains more than one `hdisk`, the logical volumes on this `hdisk` are also affected. Use the `lspv` command to check. In this case, it is safe to use the **`exportvg vioc_rootvg_1`** command:

```
$ lspv
NAME          PVID          VG          STATUS
hdisk0       00c478de00655246  rootvg     active
hdisk1       00c478de008a399b  rootvg     active
hdisk2       00c478de008a3ba1  vioc_rootvg_1  active
hdisk3       00c478deb4b0d4b0  None
```

```
$ chsp -rm -f -sp vioc_rootvg_1 hdisk2
```

Some error messages may contain invalid information for the Virtual I/O Server environment.

```
0516-062 lqueryvg: Unable to read or write logical volume manager
record. PV may be permanently corrupted. Run diagnostics
0516-882 reducevg: Unable to reduce volume group.
```

```
$ deactivatevg vioc_rootvg_1
```

Some error messages may contain invalid information for the Virtual I/O Server environment.

```
0516-062 lqueryvg: Unable to read or write logical volume manager
record. PV may be permanently corrupted. Run diagnostics
```

```
$ exportvg vioc_rootvg_1
```

5. On the Virtual I/O Server, remove the `hdisk` device:

```
$ rmdev -dev hdisk2
hdisk2 deleted
```

6. Replace the physical disk drive.

7. On the Virtual I/O Server, configure the new `hdisk` device using the `cfgdev` command and check the configuration using the `lspv` command to determine that the new disk is configured:

```
$ cfgdev
```

```
$ lspv
```

```
NAME          PVID          VG          STATUS
hdisk2       none         None
hdisk0       00c478de00655246  rootvg     active
hdisk1       00c478de008a399b  rootvg     active
hdisk3       00c478deb4b0d4b0  None
```

8. On the Virtual I/O Server, add the hdisk to the storage pool using the **chsp** command when you have more than one disk per storage pool. If you only have one storage pool per hdisk, use the **mksp** command:

```
$ mksp vioc_rootvg_1 hdisk2
vioc_rootvg_1
0516-1254 mkvg: Changing the PVID in the ODM.
```

**Note:** Alternatively to the manual disk replacement steps 4 to 7 on the Virtual I/O Server the **replphyvol** command may be used.

9. On the Virtual I/O Server, recreate the backing device of exactly the original size and attach it to the virtual device:

```
$ mkbdsp -sp vioc_rootvg_1 32G -bd vioc_1_rootvg -vadapter vhost0
Creating logical volume "vioc_1_rootvg" in storage pool "vioc_rootvg_1".
vtscsi0 Available
vioc_1_rootvg
```

**Note:** For an IBM i client partition do not attempt to determine the size for the virtual target backing device from the IBM i client partition because due to the 8-to-9 sector conversion (520 byte sectors on IBM i vs. 512 byte sectors on the Virtual I/O Server) the IBM i client shows less capacity for the disk unit than what actually needs to be configured on the Virtual I/O Server.

If you haven't noted down the size for the backing device before the Virtual I/O Server disk failure determine the size from the output of the **lslv backingdevice** command on the Virtual I/O Server of the active mirror side by multiplying the logical partitions (LPs) with the physical partition size (PP SIZE) information.

10. On the Virtual I/O Server, check that the backing device does not span a disk in the storage pool. In this case, we have only have one hdisk per storage pool.
11. For an AIX virtual I/O client:
- Reconfigure the new hdisk1 – if the parent device is unknown then the **cfgmgr** command can be executed without any parameters:
 

```
# cfgmgr -l vscsi1
```
  - Extend the rootvg:
 

```
# extendvg rootvg hdisk1
0516-1254 extendvg: Changing the PVID in the ODM.
```
  - Re-establish mirroring of the AIX client rootvg:

```
# mirrorvg -c 2 rootvg hdisk1
0516-1124 mirrorvg: Quorum requirement turned off, reboot system
for this
    to take effect for rootvg.
0516-1126 mirrorvg: rootvg successfully mirrored, user should
perform bosboot of system to initialize boot records. Then, user
must modify bootlist to include: hdisk0 hdisk1.
```

- d. Initialize the AIX boot record and set the bootlist:

```
# bosboot -a
bosboot: Boot image is 18036 512 byte blocks.
# bootlist -m normal hdisk0 hdisk1
```

12. For an IBM i client verify in SST the previously suspended disk unit automatically changed its mirrored state to *resuming* and finally *active* when the mirror re-synchronization completed.

## 2.7 Managing multiple storage security zones

**Note:** Security in a virtual environment depends on the integrity of the Hardware Management Console and the Virtual I/O Server. Access to the HMC and Virtual I/O Server must be closely monitored because they are able to modify existing storage assignments and establish new storage assignments on LPARs within the managed systems.

When planning for multiple storage security zones in a SAN environment, study the enterprise security policy for the SAN environment and the current SAN configuration.

If different security zones or disk subsystems share SAN switches, the virtual SCSI devices can share the HBAs, because the hypervisor firmware acts in a manner similar to a SAN switch. If a LUN is assigned to a partition by the Virtual I/O Server, it cannot be used or seen by any other partitions. The hypervisor is designed in a way that no operation within a client partition can gain control of or use a shared resource that is not assigned to the client partition.

When you assign a LUN in the SAN environment for the different partitions, remember that the zoning is done by the Virtual I/O Server. Therefore, in the SAN environment, assign all the LUNs to the HBAs used by the Virtual I/O Server. The Virtual I/O Server assigns the LUNs (hdisk) to the virtual SCSI server adapters (vhost) that are associated to the virtual SCSI client adapters (vscsi) used by the partitions. See Figure 2-14.

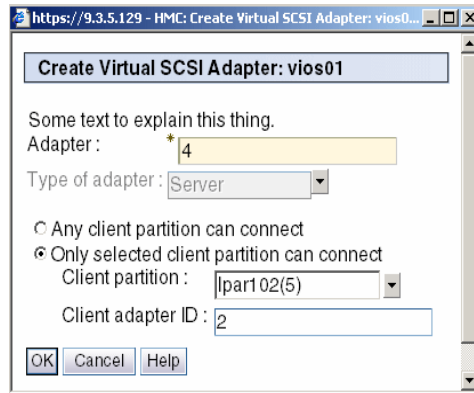


Figure 2-14 Create virtual SCSI

If accounting or security audits are made from the LUN assignment list, you will not see the true owners of LUNs, because all the LUNs are assigned to the same HBA. This might cause audit remarks.

You can produce the same kind of list from the Virtual I/O Server by using the **lsmap** command, but if it is a business requirement that the LUN mapping be at the storage list, you must use different HBA pairs for each account and security zone. You can still use the same Virtual I/O Server, because this will not affect the security policy.

If it is a security requirement, and not a hardware issue, that security zones or disk subsystems do not share SAN switches, you cannot share an HBA. In this case, you cannot use multiple Virtual I/O Servers to virtualize the LUN in one managed system, because the hypervisor firmware will act as one SAN switch.

## 2.8 Storage planning with migration in mind

Managing storage resources during an LPAR move can be more complex than managing network resources. Careful planning is required to ensure that the storage resources belonging to an LPAR are in place on the target system. This section assumes that you fairly good knowledge of PowerVM Live Partition Mobility. But if you don't know much about that please refer to *IBM System p Live Partition Mobility*, SG24-7460.



## 2.8.1 Virtual adapter slot numbers

Virtual SCSI and virtual Fibre Channel adapters are tracked by slot number and partition ID on both the Virtual I/O Server and client. The number of virtual adapters in a Virtual I/O Server must equal the sum of the virtual adapters in a client partition which it serves. The Virtual I/O Server *vhost* or *vfhost* adapter slot numbers are not required to match the client partition *vscsi* or *fscsi* slot numbers as shown for virtual SCSI in Figure 2-15).

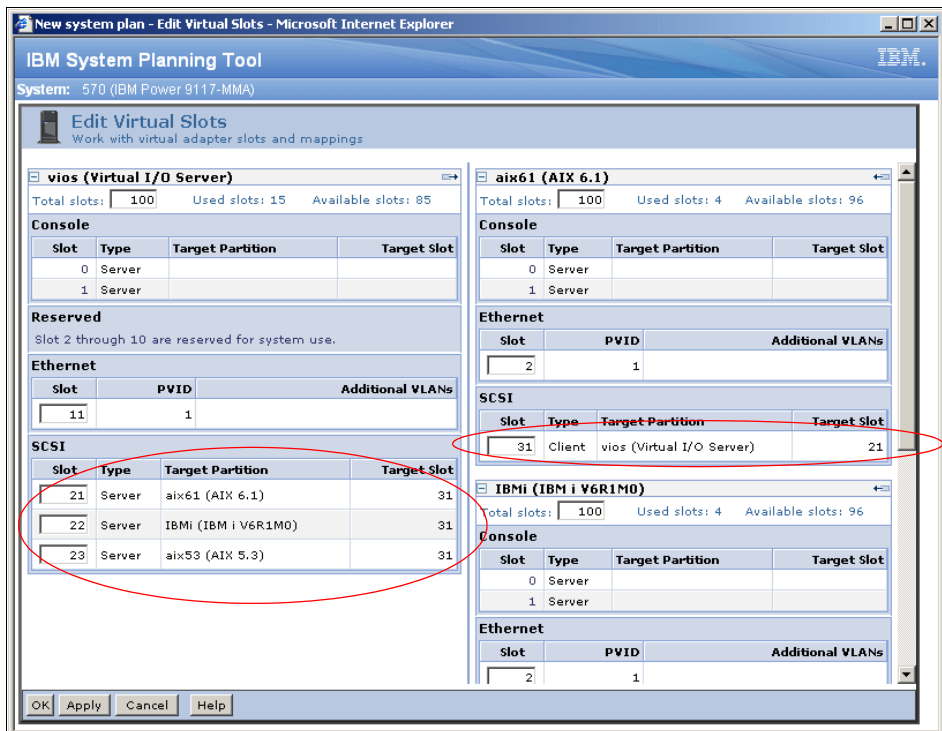


Figure 2-15 Slot numbers that are identical in the source and target system

You can apply any numbering scheme as long as server-client adapter pairs match. It is recommended to reserve a range of slot numbers for each type of virtual adapters to avoid interchanging types and slot numbers. This is also important when partitions are moved between systems.

**Note:** Do not increase maximum number of adapters for a partition beyond 1024.

## 2.8.2 SAN considerations for LPAR migration

All storage associated with a partition to be moved should be hosted on LUNs in a Fibre Channel SAN and exported as physical volumes in the Virtual I/O Server. The LUNs used by the LPAR to be moved must be visible to both the source and target Virtual I/O Servers. This can involve zoning, LUN masking, or other configuration of the SAN fabric and storage devices, which is beyond the scope of this document.

If multipath software is employed in the Virtual I/O Servers on the source system, the same multipath software must be in place on the target Virtual I/O Servers. For example, if SDDPCM (SDD or Powerpath) is in use on the source server, the same level must also be in use on the target server. Storage should not be migrated between different multipath environments during an LPAR move because this might affect the visibility of the unique tag on the disk devices.

Another important consideration is whether to allow concurrent access to the LUNs used by the LPAR. By default, the Virtual I/O Server acquires a SCSI reserve on a LUN when its hdisk is configured as a virtual SCSI target device. This means that only one Virtual I/O Server can export the LUN to an LPAR at a time. The SCSI reserve prevents the LPAR from being started on multiple systems at once, which could lead to data corruption.

The SCSI reserve does make the move process more complicated, because configuration is required on both the source and target Virtual I/O Servers between the time that the LPAR is shut down on the source and activated on the target server.

Turning off the SCSI reserve on the hdisk devices associated with the LUNs makes it possible to move the LPAR with no configuration changes on the Virtual I/O Servers during the move. However, it raises the possibility of data corruption if the LPAR is accidentally activated on both servers concurrently.

**Note:** We recommend leaving the SCSI reserve active on the hdisks in rootvg to eliminate the possibility of booting the operating system on two servers concurrently, which could result in data corruption.

Query the SCSI reserve setting with the `lsdev` command and modify it with the `chdev` command on the Virtual I/O Server. The exact setting can differ for different types of storage. The setting for LUNs on IBM storage servers should not be `no_reserve`.

```
$ lsdev -dev hdisk7 -attr reserve_policy
value
no_reserve
```

```
$ chdev -dev hdisk7 -attr reserve_policy=single_path
hdisk7 changed
$ lsdev -dev hdisk7 -attr reserve_policy
value

single_path
```

Consult the documentation from your storage vendor for the reserve setting on other types of storage.

**Note:** In a dual Virtual I/O Server configuration, both servers need to have access to the same LUNs. In this case, the reserve policy must be set to `no_reserve` on the LUNs on both Virtual I/O Servers.

In situations where the LPAR normally participates in concurrent data access, such as a GPFS™ cluster, the SCSI reserve should remain deactivated on hdisks that are concurrently accessed. These hdisks should be in separate volume groups, and the reserve should be active on all hdisks in rootvg to prevent concurrent booting of the partition.

### 2.8.3 Backing devices and virtual target devices

The source and destination partitions must have access to the same backing devices from the Virtual I/O Servers on the source and destination system. Each backing device must have a corresponding virtual target device. The virtual target device refers to a SCSI target for the backing disk or LUN, while the destination server is the system to which the partition is moving.

**Note:** Fibre Channel LUNs might have different hdisk device numbers on the source and destination Virtual I/O Server. The hdisk device numbers increment as new devices are discovered, so the order of attachment and number of other devices can influence the hdisk numbers assigned. Use the WWPN and LUN number in the device physical location to map corresponding hdisk numbers on the source and destination partitions.

Use the `lsmap` command on the source Virtual I/O Server to list the virtual target devices that must be created on the destination Virtual I/O Server and corresponding backing devices. If the vhost adapter numbers for the source Virtual I/O Server are known, run the `lsmap` command with the `-vadapter` flag for the adapter or adapters. Otherwise, run the `lsmap` command with the `-all` flag, and any virtual adapters attached to the source partition should be noted. The following listing is for an IBM System Storage DS4000 series device:

```
$ lsmap -all
```

SVSA ID	Physloc	Client Partition
-----		
vhost0	U9117.570.107CD9E-V1-C4	0x00000007
VTD	lpar07_rootvg	
LUN	0x8100000000000000	
Backing device	hdisk5	
Physloc	U7879.001.DQD186N-P1-C3-T1-W200400A0B8110D0F-L0	

The Physloc identifier for each backing device on the source Virtual I/O Server can be used to identify the appropriate hdisk device on the destination Virtual I/O Server from the output of the **lsdev -vpd** command. In some cases, with multipath I/O and multicontroller storage servers, the Physloc string can vary by a few characters, depending on which path or controller is in use on the source and destination Virtual I/O Server.

```
$ lsdev -vpd -dev hdisk4
  hdisk4          U787A.001.DNZ00XY-P1-C5-T1-W200500A0B8110D0F-L0  3542
(20
0) Disk Array Device
```

PLATFORM SPECIFIC

```
Name: disk
Node: disk
Device Type: block
```

Make a note of the hdisk device on the destination Virtual I/O Server that corresponds to each backing device on the source Virtual I/O Server.

## 2.9 N\_Port ID virtualization

N\_Port ID Virtualization (NPIV) is a new virtualization feature which was announced on October 7, 2008. The next sections describe the requirements for NPIV and how to configure it. Also an overview of supported infrastructure scenarios is given.

### 2.9.1 Introduction

NPIV is an industry standard technology that provides the capability to assign a physical Fibre Channel adapter to multiple unique world wide port names (WWPN). To access physical storage from a SAN, the physical storage is mapped to logical units (LUNs) and the LUNs are mapped to the ports of physical

Fibre Channel adapters. Then the Virtual I/O Server maps the LUNs to the virtual Fibre Channel adapter of the virtual I/O client.

**Note:** Both disk and tape SAN storage devices are supported with NPIV.

To enable NPIV on the managed system, you must create a Virtual I/O Server partition at Version 2.1, or later that provides virtual resources to virtual I/O client partitions. You assign at least one 8 Gigabit PCI Express Dual Port Fibre Channel Adapter to the Virtual I/O Server logical partition. Then, you create virtual client and server Fibre Channel adapter pair in each partition profile through the HMC or IVM. Refer to “Virtual Fibre Channel for HMC-managed systems” on page 60 and “Virtual Fibre Channel for IVM-managed systems” on page 61 for more information.

**Note:** A virtual Fibre Channel client adapter is a virtual device that provides virtual I/O client partitions with a Fibre Channel connection to a storage area network through the Virtual I/O Server partition.

The Virtual I/O Server partition provides the connection between the virtual Fibre Channel server adapters and the physical Fibre Channel adapters assigned to the Virtual I/O Server partition on the managed system.

Figure 2-16 shows a managed system configured to use NPIV, running two Virtual I/O Server partitions each with one physical Fibre Channel card. Each Virtual I/O Server partition provides virtual Fibre Channel adapters to the virtual I/O client. For increased serviceability you can use MPIO in the AIX virtual I/O client.

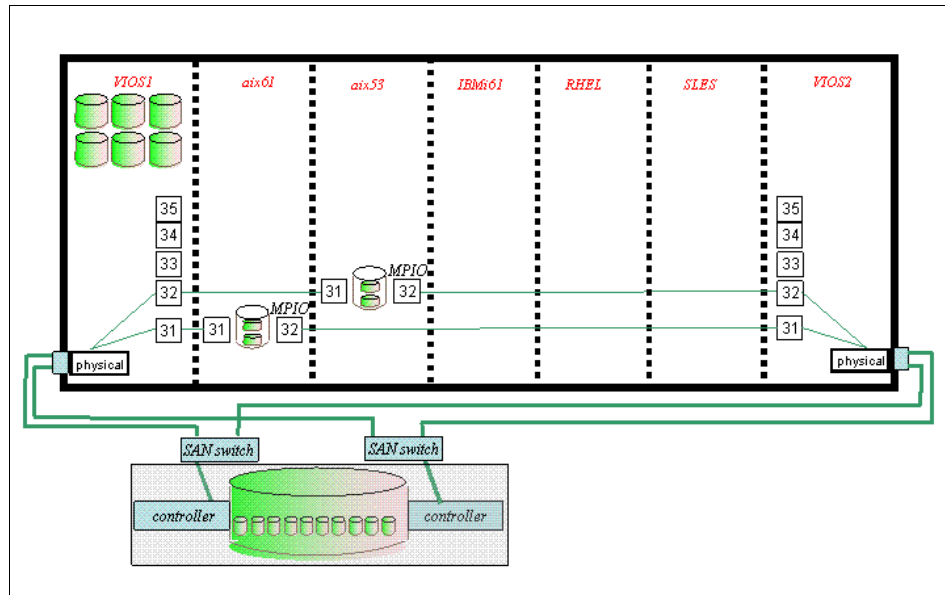


Figure 2-16 Server using redundant Virtual I/O Server partitions with NPIV

Figure 2-16 shows the following connections:

- ▶ A SAN connects several LUNs from an external physical storage system to a physical Fibre Channel adapter that is located on the managed system. Each LUN is connected through both Virtual I/O Servers for redundancy. The physical Fibre Channel adapter is assigned to the Virtual I/O Server and supports NPIV.
- ▶ There are five virtual Fibre Channel adapters available in the Virtual I/O Server. Two of them are mapped with the physical Fibre Channel adapter (adapter slot 31 and 32). All two virtual Fibre Channel server adapters are mapped to the same physical port on the physical Fibre Channel adapter.
- ▶ Each virtual Fibre Channel server adapter on the Virtual I/O Server partition connects to one virtual Fibre Channel client adapter on a virtual I/O client partition. Each virtual Fibre Channel client adapter receives a pair of unique WWPNs. The virtual I/O client partition uses one WWPN to log into the SAN at any given time. The other WWPN is used by the system when you move the virtual I/O client partition to another managed system with PowerVM Live Partition Mobility.

Using their unique WWPNs and the virtual Fibre Channel connections to the physical Fibre Channel adapter, the AIX operating system that runs in the virtual I/O client partitions discovers, instantiates, and manages their physical storage located on the SAN. The Virtual I/O Server provides the virtual I/O client

partitions with a connection to the physical Fibre Channel adapters on the managed system.

There is always a one-to-one relationship between virtual Fibre Channel client adapter and the virtual Fibre Channel server adapter.

Using the SAN tools of the SAN switch vendor, you zone your NPIV-enabled switch to include WWPNs that are created by the HMC for any virtual Fibre Channel client adapter on virtual I/O client partitions with the WWPNs from your storage device in a zone – same like for a physical environment. The SAN uses zones to provide access to the targets based on WWPNs.

Redundancy configurations help to increase the serviceability of your Virtual I/O Server environment. With NPIV, you can configure the managed system so that multiple virtual I/O client partitions can independently access physical storage through the same physical Fibre Channel adapter. Each virtual Fibre Channel client adapter is identified by a unique WWPN, which means that you can connect each virtual I/O partition to independent physical storage on a SAN.

Similar to virtual SCSI redundancy, virtual Fibre Channel redundancy can be achieved using Multi-path I/O (MPIO) and mirroring at the virtual I/O client partition. The difference between traditional redundancy with SCSI adapters and the NPIV technology using virtual Fibre Channel adapters, is that the redundancy occurs on the client, because only the client recognizes the disk. The Virtual I/O Server is essentially just a passthru managing the data transfer through the POWER hypervisor.

## 2.9.2 Requirements

You need to meet the following requirements to set up and use NPIV:

1. Hardware
  - Any POWER6-based system

**Note:** IBM intends to support N\_Port ID Virtualization (NPIV) on the POWER6 processor-based Power 595, BladeCenter JS12, and BladeCenter JS22 in 2009.

- Install a minimum System Firmware level of EL340\_036 for the IBM Power 520 and Power 550, and EM340\_036 for the IBM Power 560 and IBM Power 570.
- Minimum of one 8 Gigabit PCI Express Dual Port Fibre Channel Adapter (Feature Code 5735)

Check the latest available firmware for the adapter at:

<http://www.ibm.com/support/us/en>

Select **Power** at the support type, then go to **Firmware updates**.

**Note:** At the time of writing only the 8 Gigabit PCI Express Dual Port Fibre Channel Adapter (Feature Code 5735) was announced.

- NPIV enabled SAN switch

Only the first SAN switch which is attached to the Fibre Channel adapter in the Virtual I/O Server needs to be NPIV capable. Other switches in your SAN environment do not need to be NPIV capable.

**Note:** Check with the storage vendor, if your SAN switch is NPIV enabled.

For information on IBM SAN switches you can also refer to *Implementing an IBM/Brocade SAN with 8 Gbps Directors and Switches*, SG24-6116 and search for NPIV.

Use the latest available firmware level for your SAN switch.

## 2. Software

- HMC V7.3.4, or later
- Virtual I/O Server Version 2.1 with Fix Pack 20.1, or later
- AIX 5.3 TL9, or later
- AIX 6.1 TL2, or later

**Statement of Direction:** IBM intends to support NPIV with IBM i and Linux environments in 2009.

### 2.9.3 Managing virtual Fibre Channel adapters

Whether your server is managed by an HMC or by IVM, during the next sections you will get information about the management.

#### Virtual Fibre Channel for HMC-managed systems

On HMC-managed systems, you can dynamically add and remove virtual Fibre Channel adapters to and from the Virtual I/O Server partition and each virtual I/O client partition. You can also view information about the virtual and physical Fibre



Channel adapters and the WWPNs by using Virtual I/O Server commands. To enable NPIV on the managed system, you create the required virtual Fibre Channel adapters and connections as follows:

- ▶ You use the HMC to create virtual Fibre Channel server adapters on the Virtual I/O Server partition and associate them with virtual Fibre Channel client adapters on the virtual I/O client partitions.
- ▶ On the HMC you create virtual Fibre Channel client adapters on each virtual I/O client partition and associate them with virtual Fibre Channel server adapters on the Virtual I/O Server partition. When you create a virtual Fibre Channel client adapter on a client logical partition, the HMC generates a pair of unique WWPNs for the virtual Fibre Channel client adapter.
- ▶ Then you map the virtual Fibre Channel server adapters on the Virtual I/O Server to the physical port of the physical Fibre Channel adapter by running the `vfcmap` command on the Virtual I/O Server. The POWER hypervisor generates WWPNs based on the range of names available for use with the prefix in the vital product data on the managed system. This 6–digit prefix comes with the purchase of the managed system and includes 32,000 pairs of WWPNs. When you delete a virtual Fibre Channel client adapter from a virtual I/O client partition, the hypervisor does not reuse the WWPNs that are assigned to the virtual Fibre Channel client adapter on the client logical partition.

**Note:** The POWER hypervisor does not reuse the deleted WWPNs when generating WWPNs for virtual Fibre Channel adapters in the future.

If you run out of WWPNs, you must obtain an activation code that includes another prefix with another 32,000 pairs of WWPNs.

**Note:** For more information how to obtain the activation code, contact your IBM sales representative or your IBM Business Partner representative.

### Virtual Fibre Channel for IVM-managed systems

On systems that are managed by the Integrated Virtualization Manager (IVM), you can dynamically change the physical ports that are assigned to a logical partition and you can dynamically change the logical partitions that are assigned to a physical port. You can also view information about the virtual and physical Fibre Channel adapters and the WWPNs. To use NPIV on the managed system, you assign logical partitions directly to the physical ports of the physical Fibre Channel adapters. You can assign multiple logical partitions to one physical port. When you assign a logical partition to a physical port, the IVM automatically creates the following connections:

- ▶ The IVM creates a virtual Fibre Channel server adapter on the management partition and associates it with the virtual Fibre Channel adapter on the logical partition.
- ▶ The IVM generates a pair of unique WWPNs and creates a virtual Fibre Channel client adapter on the logical partition. The IVM assigns the WWPNs to the virtual Fibre Channel client adapter on the logical partition, and associates the virtual Fibre Channel client adapter on the logical partition with the virtual Fibre Channel server adapter on the management partition.
- ▶ The IVM connects the virtual Fibre Channel server adapter on the management partition to the physical port on the physical Fibre Channel adapter.

The IVM generates WWPNs based on the range of names available for use with the prefix in the vital product data on the managed system. This 6–digit prefix comes with the purchase of the managed system and includes 32,000 pairs of WWPNs. When you remove the connection between a logical partition and a physical port, the hypervisor deletes the WWPNs that are assigned to the virtual fibre channel client adapter on the logical partition.

**Note:** The IVM does not reuse the deleted WWPNs when generating WWPNs for virtual fibre channel client adapters in the future.

If you run out of WWPNs, you must obtain an activation code that includes another prefix with 32,000 pairs of WWPNs.

**Note:** For more information how to obtain the activation code, contact your IBM sales representative or your IBM Business Partner representative.

## 2.9.4 Configuring NPIV on Power Systems with a new AIX LPAR

This section describes the installation of an AIX operating system on a virtual I/O client on an external disk mapped through a Virtual Fibre Channel adapter. A IBM 2109-F32 SAN switch, a IBM Power 570 server, and a IBM System Storage DS4800 storage system has been used in our lab environment to describe the setup of a NPIV environment. The example and HMC screen outputs are based on this environment, but might look different in your environment, depending on your systems.

At the time of writing, NPIV is supported for AIX 5.3 and AIX 6.1. IBM intends to support NPIV for IBM i and Linux in 2009.

The following configuration is used to describe the set up of NPIV in this section, as shown in Figure 2-17.

- ▶ Virtual Fibre Channel server adapter slot 31 is used in the Virtual I/O Server partition `vios1`
- ▶ Virtual Fibre Channel client adapter slot 31 is used in the virtual I/O client partition `aix61`.
- ▶ `aix61` partition access physical storage through virtual Fibre Channel adapter.
- ▶ AIX is installed on the physical storage (boot from SAN).

Follow the next steps to set up the NPIV environment:

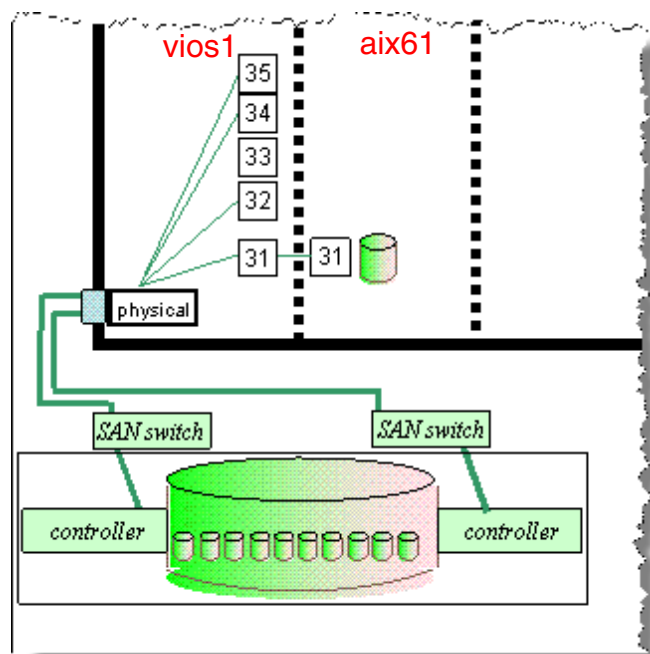


Figure 2-17 Virtual adapter numbering

1. On the SAN switch two things need to be done before it can be used for NPIV.
  - a. Update the firmware to a minimum level of Fabric OS (FOS) 5.3.0. To check the level of Fabric OS on the switch, log on to the switch and run the `version` command, as shown in Example 2-20:

Example 2-20 `version` command shows Fabric OS level

```
itsosan02:admin> version
Kernel:      2.6.14
```

```
Fabric OS: v5.3.0
Made on: Thu Jun 14 19:04:02 2007
Flash: Mon Oct 20 12:14:10 2008
BootProm: 4.5.3
```

**Note:** You can find the firmware for IBM SAN switches at:

<http://www-03.ibm.com/systems/storage/san/index.html>

Click **Support** and select **Storage are network (SAN)** in the Product family and then select your SAN product.

- b. After a successful firmware update, you have to enable the NPIV capability on each port of the SAN switch. Run the `portCfgNPIVPort` command to enable NPIV on port 16:

```
itsosan02:admin> portCfgNPIVPort 16, 1
```

The `portCfgshow` command lists information for all ports, as shown in Example 2-21:

*Example 2-21 List port configuration*

```
itsosan02:admin> portCfgshow
Ports of Slot 0  0  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
Speed           AN AN AN AN AN AN AN AN AN AN AN AN AN AN AN AN AN
Trunk Port      ON ON ON ON ON ON ON ON ON ON ON ON ON ON ON ON ON
Long Distance   .. .. .. .. .. .. .. .. .. .. .. .. .. .. .. .. ..
VC Link Init    .. .. .. .. .. .. .. .. .. .. .. .. .. .. .. .. ..
Locked L_Port   .. .. .. .. .. .. .. .. .. .. .. .. .. .. .. .. ..
Locked G_Port   .. .. .. .. .. .. .. .. .. .. .. .. .. .. .. .. ..
Disabled E_Port .. .. .. .. .. .. .. .. .. .. .. .. .. .. .. .. ..
ISL R_RDY Mode  .. .. .. .. .. .. .. .. .. .. .. .. .. .. .. .. ..
RSCN Suppressed .. .. .. .. .. .. .. .. .. .. .. .. .. .. .. .. ..
Persistent Disable.. .. .. .. .. .. .. .. .. .. .. .. .. .. .. .. ..
NPIV capability .. ON ON ON ON ON ON ON ON ON .. .. .. ON ON ON

Ports of Slot 0 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
Speed           AN AN AN AN AN AN AN AN AN AN AN AN AN AN AN AN AN
Trunk Port      ON ON ON ON ON ON ON ON ON ON ON ON ON ON ON ON ON
Long Distance   .. .. .. .. .. .. .. .. .. .. .. .. .. .. .. .. ..
VC Link Init    .. .. .. .. .. .. .. .. .. .. .. .. .. .. .. .. ..
Locked L_Port   .. .. .. .. .. .. .. .. .. .. .. .. .. .. .. .. ..
Locked G_Port   .. .. .. .. .. .. .. .. .. .. .. .. .. .. .. .. ..
Disabled E_Port .. .. .. .. .. .. .. .. .. .. .. .. .. .. .. .. ..
ISL R_RDY Mode  .. .. .. .. .. .. .. .. .. .. .. .. .. .. .. .. ..
RSCN Suppressed .. .. .. .. .. .. .. .. .. .. .. .. .. .. .. .. ..
```



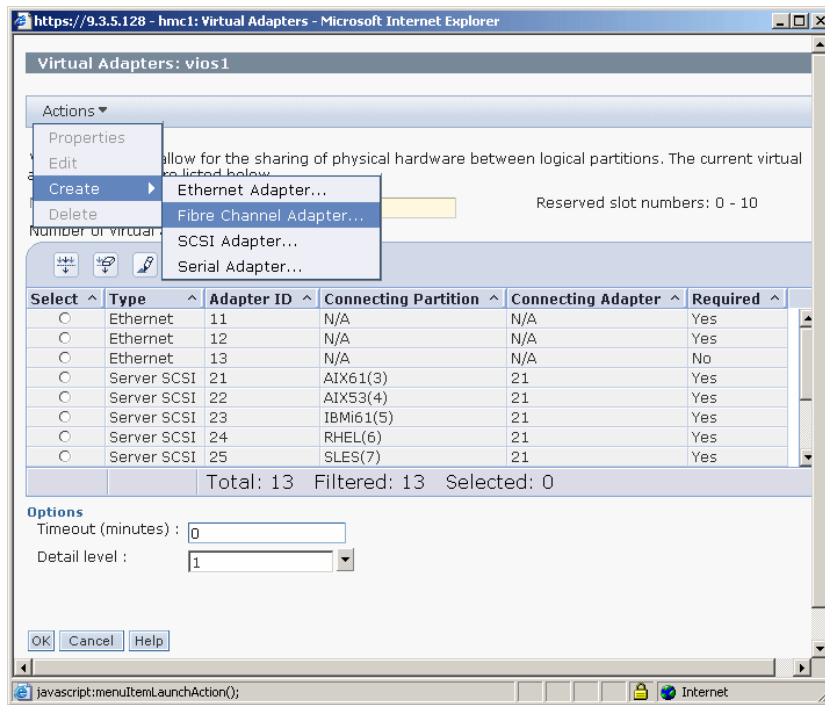


Figure 2-19 Create Fibre Channel server adapter

- d. Enter the virtual slot number for the Virtual Fibre Channel server adapter, select Client Partition to which the adapter may be assigned to, and enter the Client adapter ID as shown in Figure 2-20. and click **OK**.

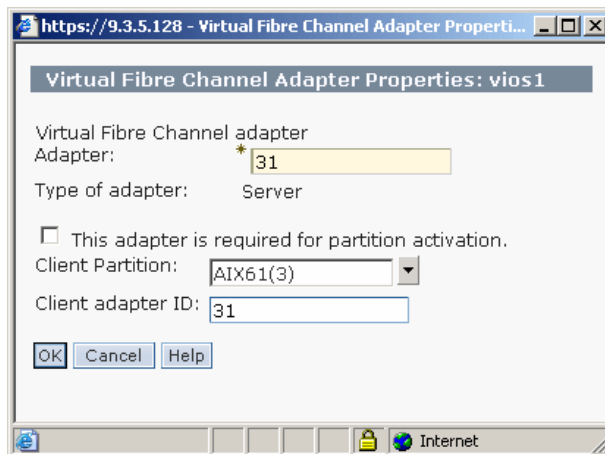


Figure 2-20 Set virtual adapter ID

- e. Click **OK**.
- f. Remember to update the profile of the Virtual I/O Server partition for the change to be reflected across restarts of the partitions. Alternatively, use the **Configuration** → **Save Current Configuration** option to save the changes to the new profile. See Figure 2-21.

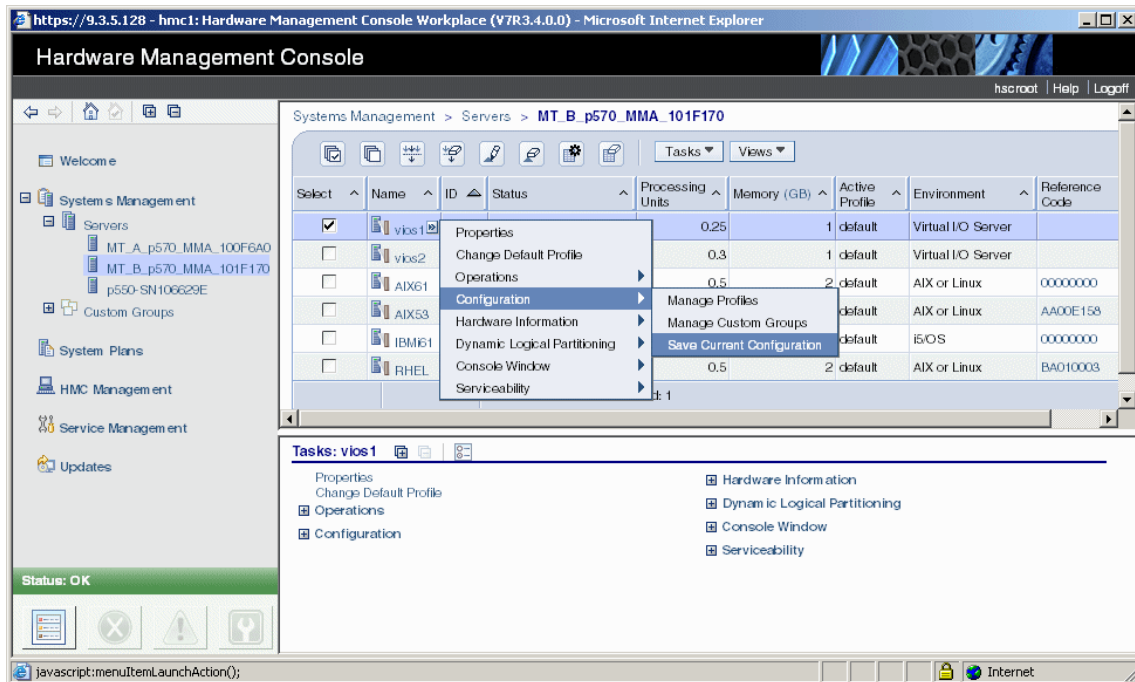


Figure 2-21 Save the Virtual I/O Server partition configuration.

- g. Change the name of the profile if required and click **OK**.
3. Use the following steps to create virtual Fibre Channel client adapter in the virtual I/O client partition.
  - a. Select the virtual I/O client partition on which the virtual Fibre Channel client adapter is to be configured. Then select **Tasks** → **Configuration** → **Manage Profiles** as shown in Figure 2-22:

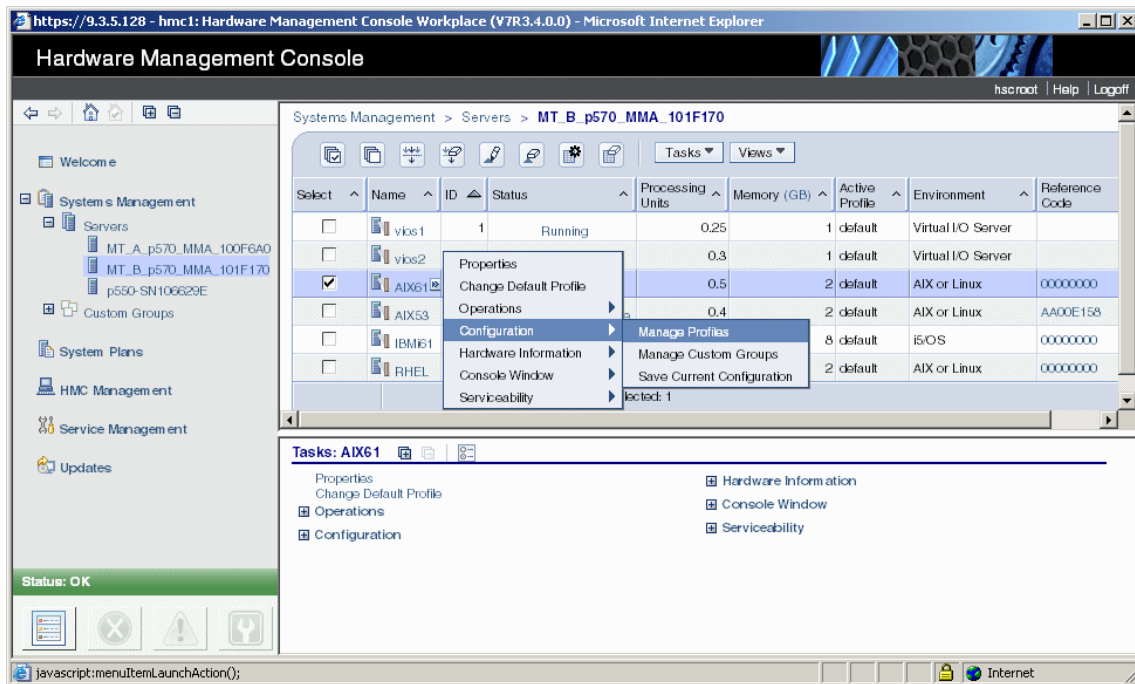


Figure 2-22 Change profile to add virtual Fibre Channel client adapter

- b. To create a virtual Fibre Channel client adapter select the profile, click on **Actions** → **Edit**, expand Virtual Adapters tab, click on **Actions** → **Create** → **Fibre Channel Adapter** as shown in Figure 2-23.

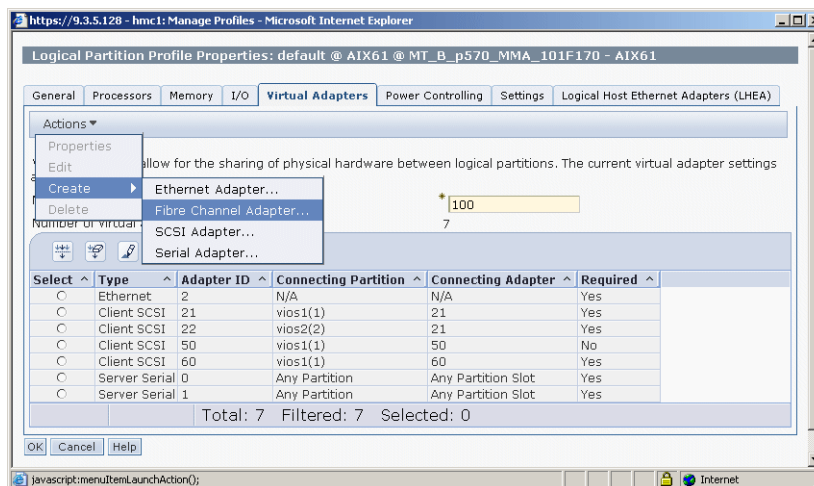


Figure 2-23 Create Fibre Channel client adapter



- c. Enter the virtual slot number for the Virtual Fibre Channel client adapter, select Virtual I/O Server partition to whom the adapter may be assigned to, and enter the Server adapter ID as shown in Figure 2-24. and click **OK**.

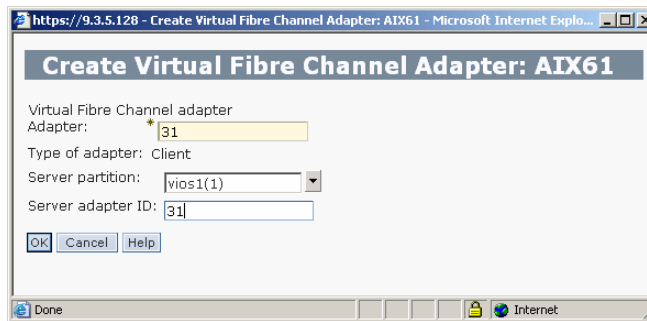


Figure 2-24 Define virtual adapter ID values

- d. Click **OK** → **OK** → **Close**.
4. Logon to the Virtual I/O Server partition as user padmin.
5. Run the `cfgdev` command to get the virtual Fibre Channel server adapter(s) configured.
6. The command `lsdev -dev vfchost*` lists all available virtual Fibre Channel server adapters in the Virtual I/O Server partition before mapping to a physical adapter as in Example 2-22.

Example 2-22 `lsdev -dev vfchost*` command on the Virtual I/O Server

---

```
$ lsdev -dev vfchost*
name          status      description
vfchost0      Available   Virtual FC Server Adapter
```

---

7. The `lsdev -dev fcs*` command lists all available physical Fibre Channel server adapters in the Virtual I/O Server partition, as shown in Example 2-23.

Example 2-23 `lsdev -dev fcs*` command on the Virtual I/O Server

---

```
$ lsdev -dev fcs*
name          status      description
fcs0          Available   4Gb FC PCI Express Adapter (df1000fe)
fcs1          Available   4Gb FC PCI Express Adapter (df1000fe)
fcs2          Available   8Gb PCI Express Dual Port FC Adapter (df1000f114108a03)
fcs3          Available   8Gb PCI Express Dual Port FC Adapter (df1000f114108a03)
```

---

8. Run the `lsnports` command to check the Fibre Channel adapter NPIV readiness of the adapter and the SAN switch. Example 2-24 shows that the

*fabric* attribute for the physical Fibre Channel adapter in slot C6 is set to 1. This means the adapter and the SAN switch is NPIV ready. If the value is equal 0, then the adapter or a SAN switch is not NPIV ready and you should check the SAN switch configuration.

*Example 2-24 lsports command on the Virtual I/O Server*

---

```
$ lsports
name          physloc          fabric tports aports swwpns awwpns
fcs3          U789D.001.DQDYKYW-P1-C6-T2      1     64     63   2048   2046
```

---

9. Before mapping the virtual FC adapter to a physical adapter, get the `vfchost` name of the virtual adapter you created, and the `fcs` name for the FC adapter from the previous `lsdev` commands output.
10. To map the virtual adapters `vfchost0` to the physical Fibre Channel adapter `fcs3`, use the `vfcmmap` command as shown in Example 2-25.

*Example 2-25 vfcmmap command with vfchost2 and fcs3*

---

```
$ vfcmmap -vadapter vfchost0 -fcp fcs3
vfchost0 changed
```

---

11. To list the mappings use the `lsmmap -npiv -vadapter vfchost0` command, as shown in Example 2-26.

*Example 2-26 lsmmap -npiv -vadapter vfchost0 command*

---

```
$ lsmmap -npiv -vadapter vfchost0
Name          Physloc          CIntID CIntName      CIntOS
=====
vfchost0      U9117.MMA.101F170-V1-C31      3
```

```
Status:NOT_LOGGED_IN
FC name:          FC loc code:
Ports logged in:0
Flags:1<NOT_MAPPED,NOT_CONNECTED>
VFC client name:          VFC client DRC:
```

---

12. After you have created the virtual Fibre Channel server adapters in the Virtual I/O server partition and in the virtual I/O client partition, you need to do the correct zoning in the SAN switch. To do so, follow the next steps:
  - a. Get the information about the WWPN of the virtual Fibre Channel client adapter created in the virtual I/O client partition.
    - i. Select the appropriate virtual I/O client partition, then click **Task** → **Properties**, expand Virtual Adapters tab, select the Client Fibre Channel client adapter, click on **Actions** → **Properties** to list the

properties of the virtual Fibre Channel client adapter, as shown in Figure 2-25.

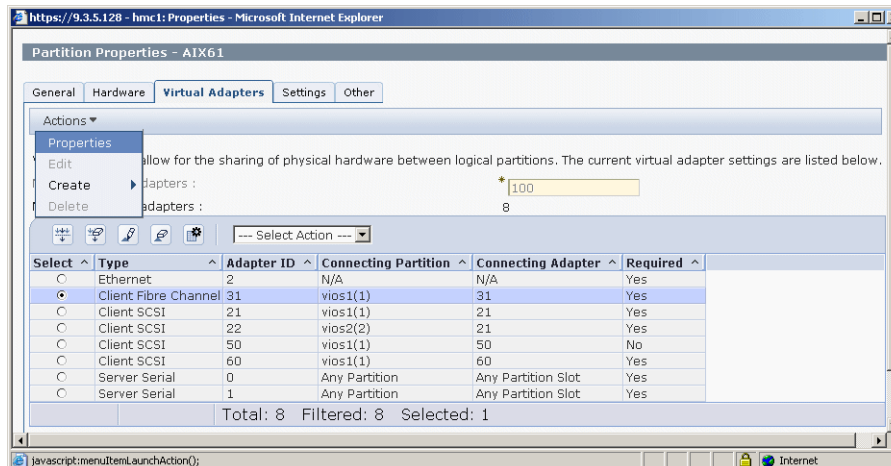


Figure 2-25 Select virtual Fibre Channel client adapter properties

- ii. Figure 2-26 shows the properties of the virtual Fibre Channel client adapter. Here you can get the WWPN which is required for the zoning.

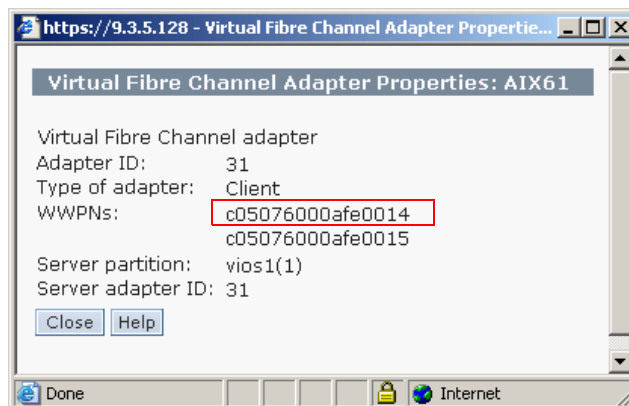


Figure 2-26 Virtual Fibre Channel client adapter Properties

- b. Logon to your SAN switch and create a new zoning or customize an existing one.

The command **zoneshow**, available on the IBM 2109-F32 switch lists the existing zones, as shown in Example 2-27.

*Example 2-27 zonestow command before adding a new WWPN*


---

```

itsosan02:admin> zonestow
Defined configuration:
  cfg:  npiv  vios1; vios2
  zone:  vios1  20:32:00:a0:b8:11:a6:62; c0:50:76:00:0a:fe:00:18
  zone:  vios2  C0:50:76:00:0A:FE:00:12; 20:43:00:a0:b8:11:a6:62

Effective configuration:
  cfg:  npiv
  zone:  vios1  20:32:00:a0:b8:11:a6:62
           c0:50:76:00:0a:fe:00:18
  zone:  vios2  c0:50:76:00:0a:fe:00:12
           20:43:00:a0:b8:11:a6:62

```

---

To add the WWPN c0:50:76:00:0a:fe:00:14 to the zone named vios1 run the command:

```
itsosan02:admin> zoneadd "vios1", "c0:50:76:00:0a:fe:00:14"
```

To save and enable the new zoning, run the **cfgsave** and **cfgenable npiv** commands, as shown in Example 2-28.

*Example 2-28 cfgsave and cfgenable commands*


---

```

itsosan02:admin> cfgsave
You are about to save the Defined zoning configuration. This
action will only save the changes on Defined configuration.
Any changes made on the Effective configuration will not
take effect until it is re-enabled.
Do you want to save Defined zoning configuration only? (yes, y, no, n): [no]
y
Updating flash ...
itsosan02:admin> cfgenable npiv
You are about to enable a new zoning configuration.
This action will replace the old zoning configuration with the
current configuration selected.
Do you want to enable 'npiv' configuration (yes, y, no, n): [no] y
zone config "npiv" is in effect
Updating flash ...

```

---

With the **zonestow** command, you can check if the added WWPN is active, as shown in Example 2-29:

*Example 2-29 zonestow command after adding a new WWPN*


---

```

itsosan02:admin> zonestow
Defined configuration:
  cfg:  npiv  vios1; vios2

```

```

zone:  vios1    20:32:00:a0:b8:11:a6:62; c0:50:76:00:0a:fe:00:18;
           c0:50:76:00:0a:fe:00:14
zone:  vios2    C0:50:76:00:0A:FE:00:12; 20:43:00:a0:b8:11:a6:62

```

Effective configuration:

```

cfg:  npiv
zone:  vios1    20:32:00:a0:b8:11:a6:62
           c0:50:76:00:0a:fe:00:18
           c0:50:76:00:0a:fe:00:14
zone:  vios2    c0:50:76:00:0a:fe:00:12
           20:43:00:a0:b8:11:a6:62

```

- c. After you have finished with the zoning, you need to map the LUN device(s) to the WWPN. In our example the LUNs named NPIV\_AIX61 is mapped to the Host Group named VIOS1\_NPIV, as shown in Figure 2-27.

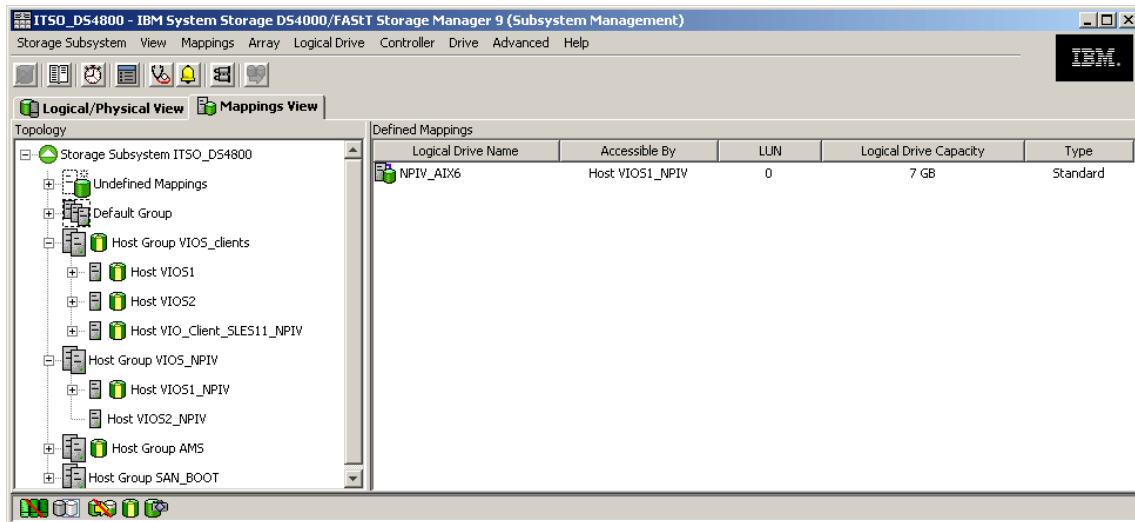


Figure 2-27 LUN mapping on DS4800

13. Activate your AIX client partition and boot it into SMS.
14. Select the correct boot devices within SMS, such as a DVD or a NIM Server.
15. Continue to boot the LPAR into the AIX Installation Main menu.
16. Select the disk where you want to install the operating system and continue to install AIX.

## 2.9.5 Configuring NPIV on Power Systems with existing AIX LPARs

This section describes how to add an external disk mapped through a Virtual Fibre Channel adapter to an AIX virtual I/O client partition. A IBM 2109-F32 SAN switch, a IBM Power 570 server, and a IBM System Storage DS4800 storage system has been used in our lab environment to describe the setup of a NPIV environment. The example and HMC screen outputs are based on this environment, but might look different in your environment, depending on your systems.

At the time of writing, NPIV is supported for AIX 5.3 and AIX 6.1. IBM intends to support NPIV for IBM i and Linux in 2009. Figure 2-28 shows a general overview of possible configurations.

To describe the configuration of NPIV in this section the following set up is used.

- ▶ Virtual Fibre Channel server adapter slot 31 is used in the Virtual I/O Server partition *vios1*.
- ▶ Virtual Fibre Channel adapter client slot 31 is used in the virtual I/O client partition *aix61*.
- ▶ *aix61* partition access physical storage through virtual Fibre Channel adapter.

Follow the next steps to set up the environment:

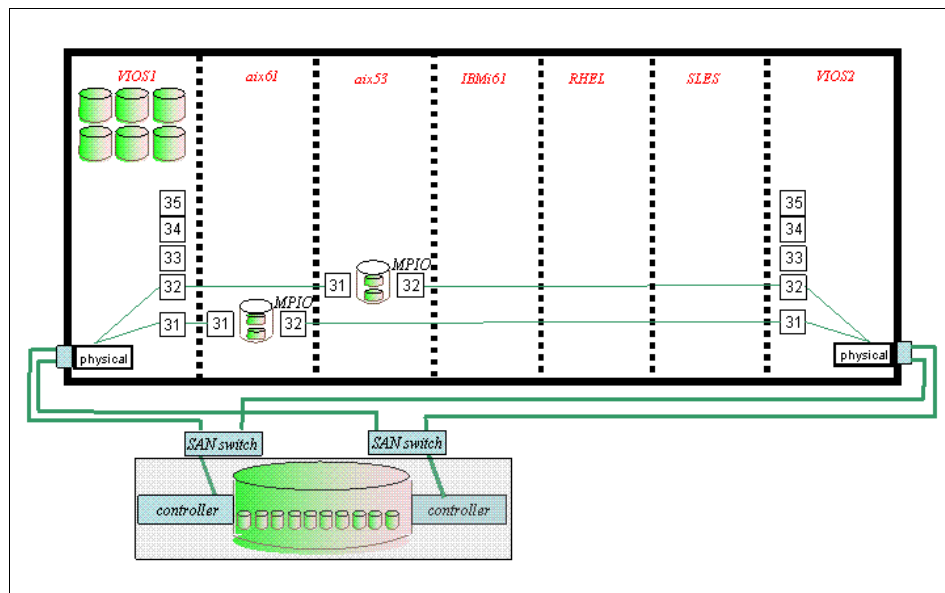


Figure 2-28 NPIV configuration

1. On the SAN switch two things need to be done before it could be used for NPIV.
  - a. Update the firmware to a minimum level of Fabric OS (FOS) 5.3.0. To check the level of Fabric OS on the switch, log on to the switch and run the **version** command, as shown in Example 2-30:

*Example 2-30 version command shows Fabric OS level*

---

```
itsosan02:admin> version
Kernel:      2.6.14
Fabric OS:   v5.3.0
Made on:     Thu Jun 14 19:04:02 2007
Flash:       Mon Oct 20 12:14:10 2008
BootProm:    4.5.3
```

---

**Note:** You can find the firmware for IBM SAN switches at:

<http://www-03.ibm.com/systems/storage/san/index.html>

Click **Support** and select **Storage are network (SAN)** in the Product family and then select your SAN product.

- b. After a successful firmware update, you have to enable the NPIV capability on each port of the SAN switch. Run the **portCfgNPIVPort** command to enable NPIV on port 16:

```
itsosan02:admin> portCfgNPIVPort 16, 1
```

The **portcfgshow** command list information for all ports, as shown in Example 2-31:

*Example 2-31 List port configuration*

---

```
itsosan02:admin> portcfgshow
Ports of Slot 0  0  1  2  3   4  5  6  7   8  9 10 11  12 13 14 15
-----+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---
Speed          AN AN AN AN  AN AN AN AN  AN AN AN AN  AN AN AN AN
Trunk Port      ON ON ON ON  ON ON ON ON  ON ON ON ON  ON ON ON ON
Long Distance   .. .. .. ..  .. .. .. ..  .. .. .. ..  .. .. .. ..
VC Link Init    .. .. .. ..  .. .. .. ..  .. .. .. ..  .. .. .. ..
Locked L_Port   .. .. .. ..  .. .. .. ..  .. .. .. ..  .. .. .. ..
Locked G_Port   .. .. .. ..  .. .. .. ..  .. .. .. ..  .. .. .. ..
Disabled E_Port .. .. .. ..  .. .. .. ..  .. .. .. ..  .. .. .. ..
ISL R_RDY Mode  .. .. .. ..  .. .. .. ..  .. .. .. ..  .. .. .. ..
RSCN Suppressed .. .. .. ..  .. .. .. ..  .. .. .. ..  .. .. .. ..
Persistent Disable.. .. .. ..  .. .. .. ..  .. .. .. ..  .. .. .. ..
NPIV capability .. ON ON ON  ON ON ON ON  ON ON .. ..  .. ON ON ON
```

Ports of Slot 0	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
Speed	AN	AN	AN	AN	AN	AN	AN	AN	AN	AN	AN	AN	AN	AN	AN	AN
Trunk Port	ON	ON	ON	ON	ON	ON	ON	ON	ON	ON	ON	ON	ON	ON	ON	ON
Long Distance	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
VC Link Init	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
Locked L_Port	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
Locked G_Port	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
Disabled E_Port	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
ISL R_RDY Mode	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
RSCN Suppressed	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
Persistent Disable	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
<b>NPIV capability</b>	<b>ON</b>	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..

where AN:AutoNegotiate, ..:OFF, ?:INVALID,  
SN:Software controlled AutoNegotiation.

**Note:** Refer to your SAN switch users guide for the command to enable NPIV on your SAN switch.

2. Use the following steps to create virtual Fibre Channel server adapter in the Virtual I/O Server partition.
  - a. On the HMC select the managed server to be configured  
**Systems Management** → **Servers** → <servername>
  - b. Select the Virtual I/O Server partition on which the virtual Fibre Channel server adapter is to be configured. Then select **Tasks** → **Dynamic Logical Partitioning** → **Virtual Adapters** as shown in Figure 2-29:



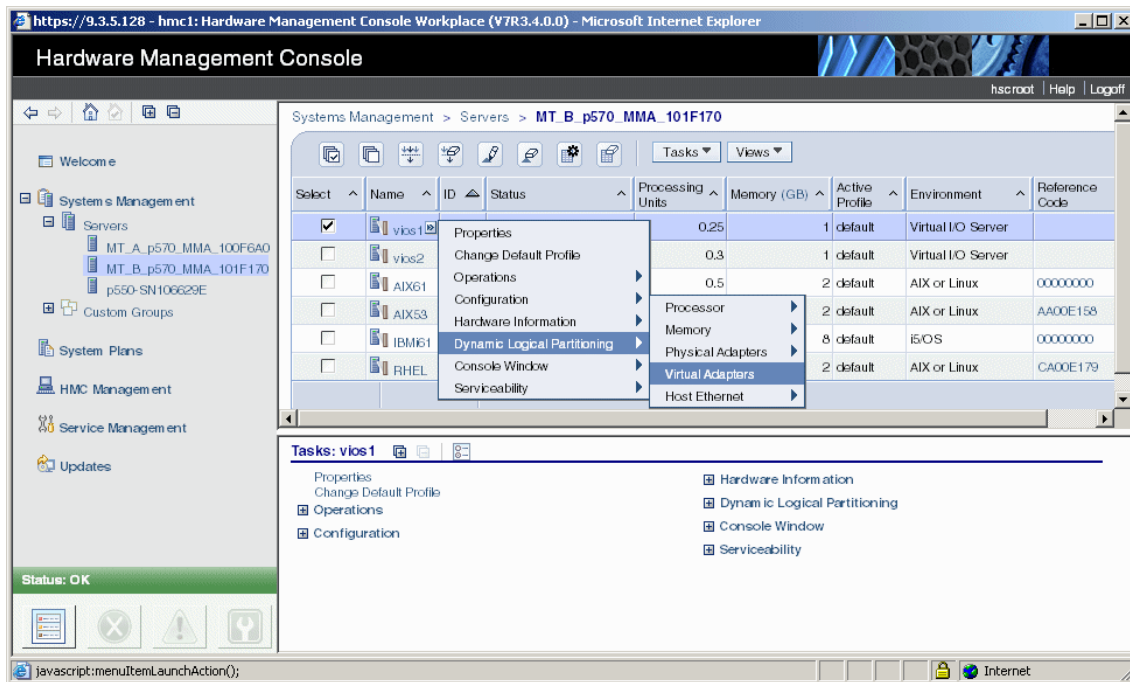


Figure 2-29 Dynamically add virtual adapter.

- c. To create a virtual Fibre Channel server adapter click on **Actions** → **Create** → **Fibre Channel Adapter...** as shown in Figure 2-30.

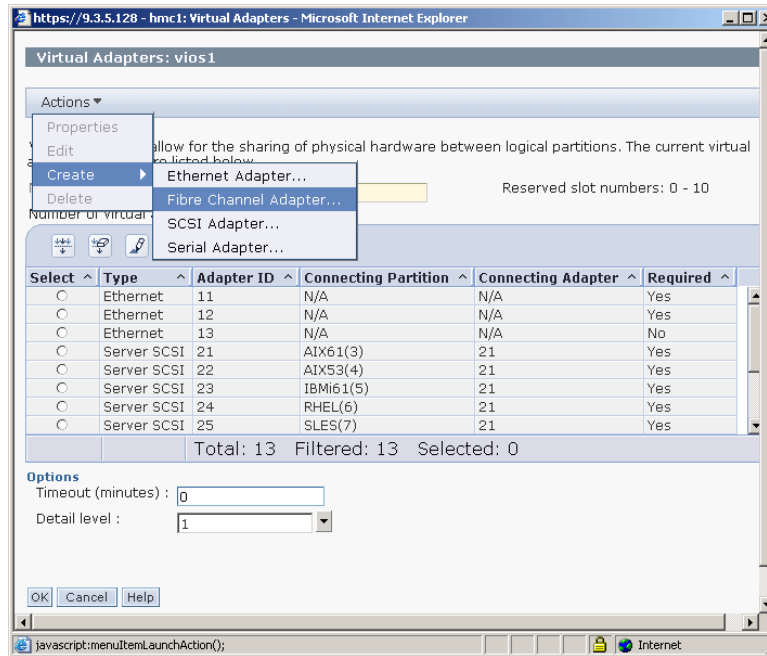


Figure 2-30 Create Fibre Channel server adapter

- d. Enter the virtual slot number for the virtual Fibre Channel server adapter, select Client Partition to which the adapter may be assigned to, and enter the Client adapter ID as shown in Figure 2-31. and click **OK**.

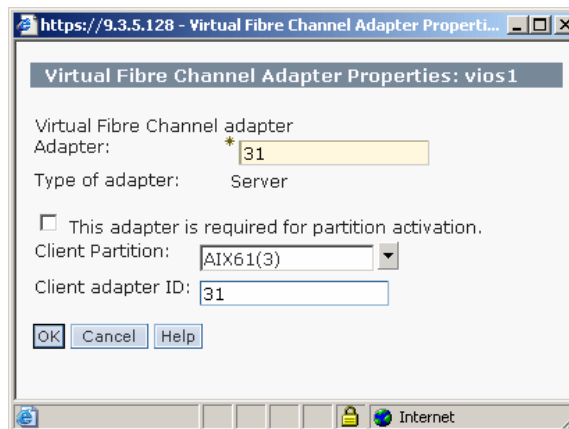


Figure 2-31 Set virtual adapter ID

- e. Click **OK**.

- f. Remember to update the profile of the Virtual I/O Server partition for the change to be reflected across restarts of the partitions. Alternatively, use the **Configuration** → **Save Current Configuration** option to save the changes to the new profile. See Figure 2-32.

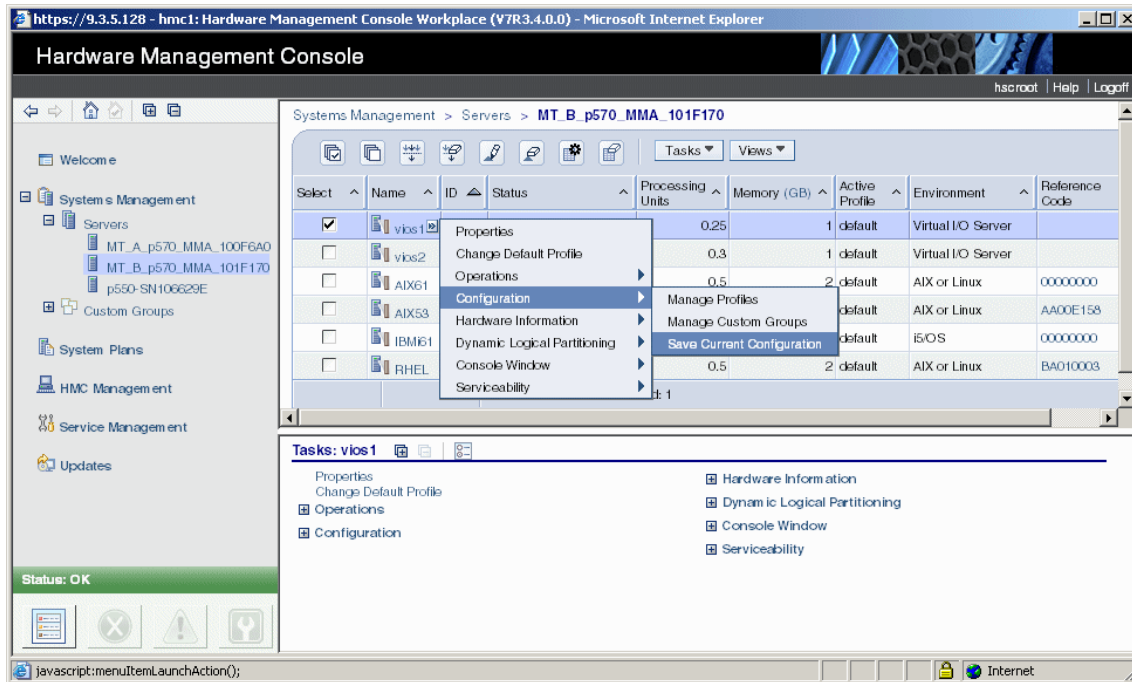


Figure 2-32 Save the Virtual I/O Server partition configuration.

- g. Change the name of the profile if required and click **OK**.

**Note:** To increase the serviceability you can use a second Virtual I/O Server partition. Repeat steps a-e to create virtual Fibre Channel server adapters in the second Virtual I/O Server.

3. Run the next steps to create virtual Fibre Channel client adapter(s) in the virtual I/O client partition.
  - a. Select the virtual I/O client partition on which the virtual Fibre Channel client adapter is to be configured. Then select **Tasks** → **Dynamic Logical Partitioning** → **Virtual Adapters** as shown in Figure 2-33:

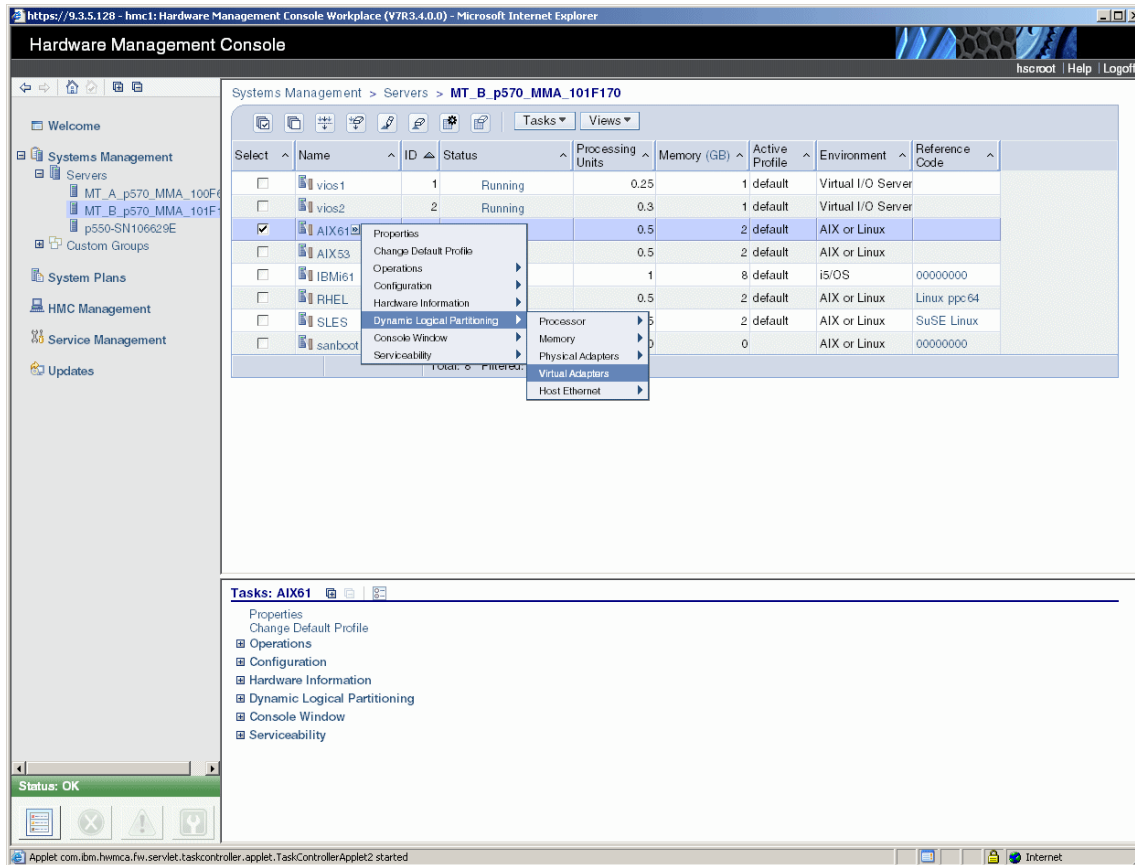


Figure 2-33 Dynamically add virtual adapter.

- b. To create a virtual Fibre Channel client adapter click on **Actions** → **Create** → **Fibre Channel Adapter...** as shown in Figure 2-34.

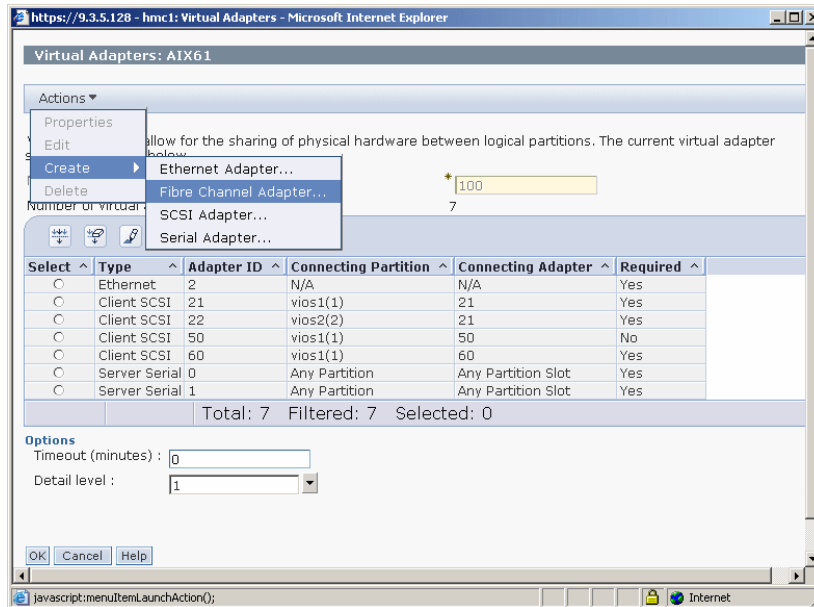


Figure 2-34 Create Fibre Channel client adapter

- c. Enter the virtual slot number for the Virtual Fibre Channel client adapter, select Virtual I/O Server partition to whom the adapter may be assigned to, and enter the Server adapter ID as shown in Figure 2-35. and click **OK**.

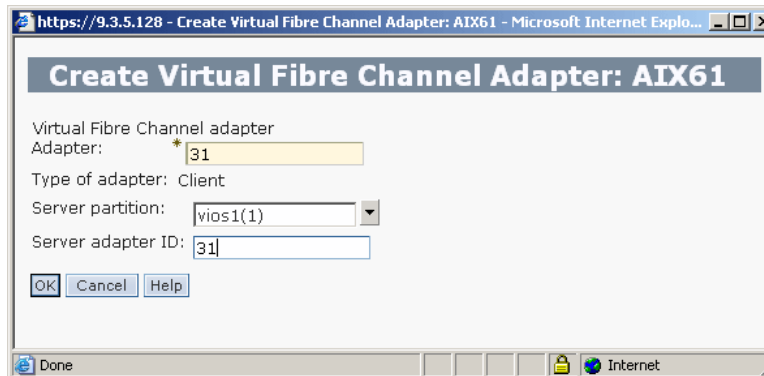


Figure 2-35 Define virtual adapter ID values

- d. Click **OK**.
- e. Remember to update the profile of the virtual I/O client partition for the change to be reflected across restarts of the partitions. Alternatively, use

the **Configuration** → **Save Current Configuration** option to save the changes to the new profile. See Figure 2-36.

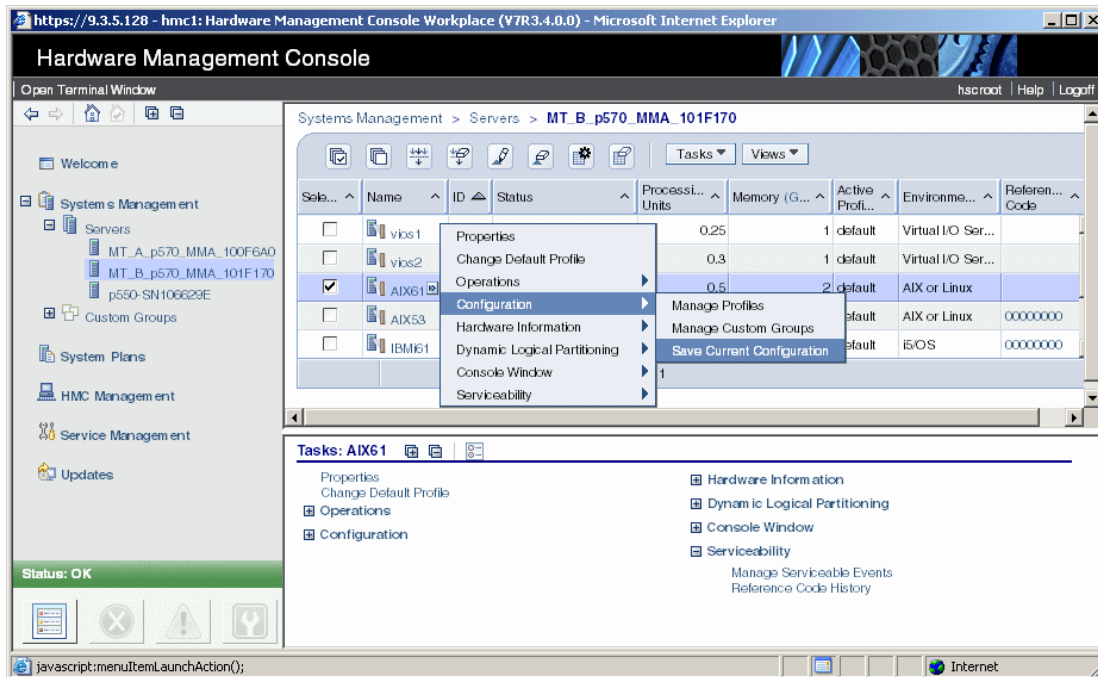


Figure 2-36 Save the virtual I/O client partition configuration.

- f. Change the name of the profile if required and click **OK**.
4. Logon to the Virtual I/O Server partition as user padmin.
5. Run the `cfgdev` command to get the virtual Fibre Channel server adapter(s) configured.
6. The command `lsdev -dev vfchost*` lists all available virtual Fibre Channel server adapters in the Virtual I/O Server partition before mapping to a physical adapter as in Example 2-32.

*Example 2-32* `lsdev -dev vfchost*` command on the Virtual I/O Server

```
$ lsdev -dev vfchost*
name          status      description
vfchost0     Available  Virtual FC Server Adapter
```

7. The `lsdev -dev fcs*` command lists all available physical Fibre Channel server adapters in the Virtual I/O Server partition, as shown in Example 2-33.

*Example 2-33 lsdev -dev fcs\* command on the Virtual I/O Server*


---

```
$ lsdev -dev fcs*
name          status      description
fcs0          Available  4Gb FC PCI Express Adapter (df1000fe)
fcs1          Available  4Gb FC PCI Express Adapter (df1000fe)
fcs2          Available  8Gb PCI Express Dual Port FC Adapter (df1000f114108a03)
fcs3          Available  8Gb PCI Express Dual Port FC Adapter (df1000f114108a03)
```

---

8. Run the **lsnports** command to check the NPIV readiness of the Fibre Channel adapter and the SAN switch. Example 2-34 shows that the *fabric* attribute for the physical Fibre Channel adapter in slot C6 is set to 1. This means the adapter and the SAN switch is NPIV ready. If the value is equal 0, then the adapter or a SAN switch is not NPIV ready and you should check the SAN switch configuration.

*Example 2-34 lsnports command on the Virtual I/O Server*


---

```
$ lsnports
name          physloc          fabric tports aports swwpns awwpns
fcs3          U789D.001.DQDYKYW-P1-C6-T2      1     64     63    2048    2046
```

---

9. Before mapping the virtual FC adapter to a physical adapter, get the *vfchost* name of the virtual adapter you created, and the *fcs* name for the FC adapter from the previous **lsdev** commands output.
10. To map the virtual adapter *vfchost0* to the physical Fibre Channel adapter *fcs3*, use the commands shown in Example 2-35.

*Example 2-35 vfcmmap command with vfchost2 and fcs3*


---

```
$ vfcmmap -vadapter vfchost0 -fcp fcs3
vfchost0 changed
```

---

11. To list the mappings use the **lsmmap -vadapter vfchost0 -npiv** command, as shown in Example 2-36.

*Example 2-36 lsmmap -all -npiv command*


---

```
$ lsmmap -npiv -all
Name          Physloc          CIntID CIntName          CIntOS
=====
vfchost0    U9117.MMA.101F170-V1-C31      3 AIX61          AIX

Status:LOGGED_IN
FC name:fcs3          FC loc code:U789D.001.DQDYKYW-P1-C6-T2
Ports logged in:1
Flags:a<LOGGED_IN,STRIP_MERGE>
```

---

VFC client name: **fcs0**

VFC client DRC:U9117.MMA.101F170-V3-C31-T1

As a result of this example, you can see that the virtual Fibre Channel server adapter `vfchost0` in the Virtual I/O Server is mapped to the physical Fibre Channel adapter `fcs3` in the Virtual I/O Server and appears as `fcs0`, which is a virtual Fibre Channel client adapter in the virtual I/O client partition named `AIX61`, having partition ID 3 and runs AIX as operating system.

12. After you have created the virtual Fibre Channel server adapters in the Virtual I/O server partition and in the virtual I/O client partition, you need to do the correct zoning in the SAN switch. To do so, follow the next steps:
  - a. Get the information about the WWPN of the virtual Fibre Channel client adapter created in the virtual I/O client partition.
    - i. Select the appropriate virtual I/O client partition, then click **Task** → **Properties**, expand **Virtual Adapters** tab, select the Client Fibre Channel client adapter, click on **Actions** → **Properties** to list the properties of the virtual Fibre Channel client adapter, as shown in Figure 2-37.

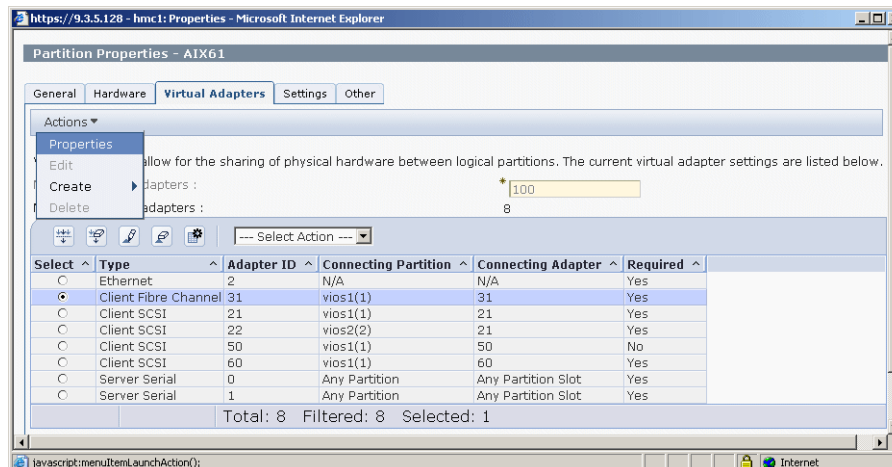


Figure 2-37 Select virtual Fibre Channel client adapter properties

- ii. Figure 2-38 shows the properties of the virtual Fibre Channel client adapter. Here you can get the WWPN which is required for the zoning.



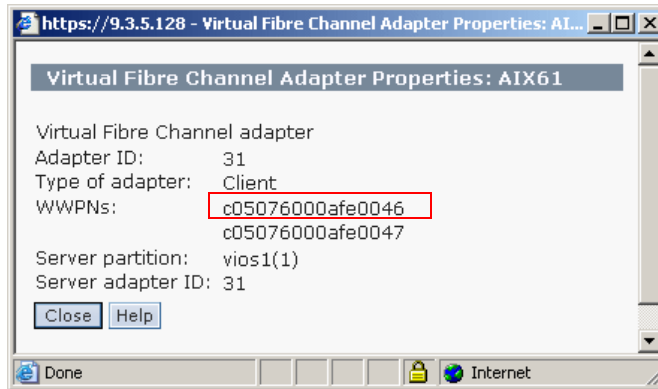


Figure 2-38 Virtual Fibre Channel client adapter Properties

- b. Logon to your SAN switch and create a new zoning or customize an existing one.

The command **zonestow**, available on the IBM 2109-F32 switch lists the existing zones, as shown in Example 2-37.

Example 2-37 *zonestow* command before adding a WWPN

---

```

itsosan02:admin> zonestow
Defined configuration:
cfg:  npiv    vios1; vios2
zone: vios1  20:32:00:a0:b8:11:a6:62; c0:50:76:00:0a:fe:00:14
zone: vios2  C0:50:76:00:0A:FE:00:12; 20:43:00:a0:b8:11:a6:62

Effective configuration:
cfg:  npiv
zone: vios1  20:32:00:a0:b8:11:a6:62
           c0:50:76:00:0a:fe:00:14
zone: vios2  c0:50:76:00:0a:fe:00:12
           20:43:00:a0:b8:11:a6:62

```

---

To add the WWPN c0:50:76:00:0a:fe:00:45 to the zone named vios1 run the command:

```
itsosan02:admin> zoneadd "vios1", "c0:50:76:00:0a:fe:00:45"
```

To save and enable the new zoning, run the **cfgsave** and **cfgenable npiv** commands, as shown in Example 2-38.

*Example 2-38 cfgsave and cfgenable commands*


---

```

itsosan02:admin> cfgsave
You are about to save the Defined zoning configuration. This
action will only save the changes on Defined configuration.
Any changes made on the Effective configuration will not
take effect until it is re-enabled.
Do you want to save Defined zoning configuration only? (yes, y, no, n): [no]
y
Updating flash ...
itsosan02:admin> cfgenable npiv
You are about to enable a new zoning configuration.
This action will replace the old zoning configuration with the
current configuration selected.
Do you want to enable 'npiv' configuration (yes, y, no, n): [no] y
zone config "npiv" is in effect
Updating flash ...

```

---

With the **zonestow** command, you can check if the added WWPN is active, as shown in Example 2-39:

*Example 2-39 zonestow command after adding a new WWPN*


---

```

itsosan02:admin> zonestow
Defined configuration:
cfg:  npiv  vios1; vios2
zone:  vios1  20:32:00:a0:b8:11:a6:62; c0:50:76:00:0a:fe:00:14;
          c0:50:76:00:0a:fe:00:45
zone:  vios2  C0:50:76:00:0A:FE:00:12; 20:43:00:a0:b8:11:a6:62

Effective configuration:
cfg:  npiv
zone:  vios1  20:32:00:a0:b8:11:a6:62
          c0:50:76:00:0a:fe:00:14
          c0:50:76:00:0a:fe:00:45
zone:  vios2  c0:50:76:00:0a:fe:00:12
          20:43:00:a0:b8:11:a6:62

```

---

- c. After you have finished with the zoning, you need to map the LUN device(s) to the WWPN. In our example the LUNs named NPIV\_AIX61 is mapped to the Host Group named VIOS1\_NPIV, as shown in Figure 2-39.

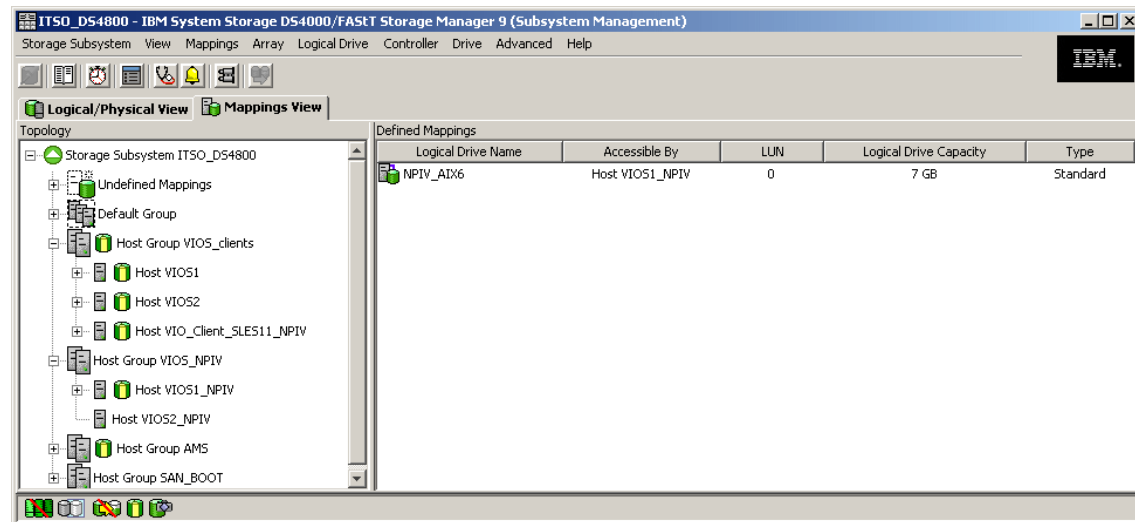


Figure 2-39 LUN mapping within DS4800

13. Login to your AIX client partition. NPV supports a LUN to be mapped as a virtual disk directly from the SAN to the AIX client via a Virtual I/O Server partition using virtual WWPNS. Before a LUN can be seen as a disk you need to run the **cfgmgr** command as shown in Example 2-40. After the **cfgmgr** command was executed, the **lspv** command will list **hdisk2**:

Example 2-40 *cfgmgr* and *lspv* command from the AIX client partition

---

```
# lspv
hdisk0      00c1f170e327afa7      rootvg      active
hdisk1      00c1f170e170fbb2      None

# cfgmgr

# lspv
hdisk0      00c1f170e327afa7      rootvg      active
hdisk1      00c1f170e170fbb2      None
hdisk2      none                      None
```

---

14. You can list all virtual Fibre Channel client adapters in the virtual I/O client partition using the following command:

```
# lsdev -l fcs*
fcs0 Available 31-T1 Virtual Fibre Channel Client Adapter
```

15. To see the available path run the command:

```
# lspath
Enabled hdisk0 vscsi0
```

```
Enabled hdisk1 vscsi0
Enabled hdisk0 vscsi1
Enabled hdisk2 fscsi0
```

The output from the command shows that hdisk2 has one path through the virtual Fibre Channel client adapters fcs0.

16. Use the `mpio_get_config` command to get more detailed information as shown in Example 2-41.

*Example 2-41 mpio\_get\_config command from the AIX client partition*

---

```
# mpio_get_config -A
Storage Subsystem worldwide name: 60ab800114632000048ed17e
Storage Subsystem Name = 'ITS0_DS4800'
  hdisk      LUN #   Ownership           User Label
  hdisk2      0     A (preferred)       NPIV_AIX61
```

---

17. Finally you have to turn on the health check interval, which defines how often the health check is performed on the paths for a device. The attribute supports a range from 0 to 3,600 seconds. When a value of 0 is selected, health checking is disabled. Use the following command:

```
# chdev -l hdisk2 -a hcheck_interval=60 -P
hdisk2 changed
```

## 2.9.6 Redundancy configurations for virtual Fibre Channel adapters

In order to implement the highly reliable configurations, the following redundancy configurations which protect your SAN from physical adapter failures as well as Virtual I/O Server failures are recommended. With NPIV, you can configure the managed system so that multiple logical partitions can independently access physical storage through the same physical Fibre Channel adapter. Each virtual Fibre Channel client adapter in a logical partition is identified by a unique WWPN, which means that you can connect physical storage from a SAN to each logical partition.

Similar to virtual SCSI redundancy, virtual Fibre Channel redundancy can be achieved using MPIIO or mirroring at the client partition. The difference between traditional redundancy with SCSI adapters and the NPIV technology using virtual Fibre Channel client adapters, is that the redundancy occurs on the client, because only the client recognizes the disk. The Virtual I/O Server is essentially just a channel managing the data transfer through the POWER hypervisor. Host bus adapter failover

A Host bus adapter is a physical Fibre Channel adapter which can be assigned to a logical partition. A Host bus adapter (HBA) failover provides a basic level of redundancy for the client logical partition, as shown in Figure 2-40.

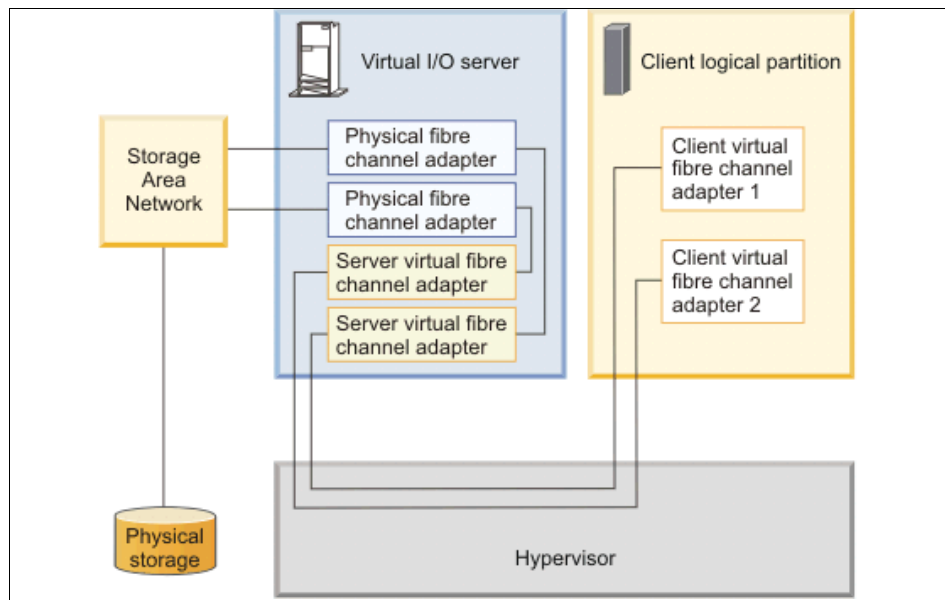


Figure 2-40 Host bus adapter failover

Figure 2-40 shows the following connections:

- ▶ The SAN connects physical storage to two physical Fibre Channel adapters located on the managed system.
- ▶ Two physical Fibre Channel adapters are assigned to the Virtual I/O Server partition and support NPIV.
- ▶ The physical Fibre Channel ports are each connected to a virtual Fibre Channel server adapter on the Virtual I/O Server. The two virtual Fibre Channel server adapters on the Virtual I/O Server are connected to ports on two different physical Fibre Channel adapters in order to provide redundancy for the physical adapters.
- ▶ Each virtual Fibre Channel server adapter in the Virtual I/O Server partition is connected to one virtual Fibre Channel client adapter on a virtual I/O client partition. Each virtual Fibre Channel client adapter on each virtual I/O client partition receives a pair of unique WWPNs. The virtual I/O client partition uses one WWPN to log into the SAN at any given time. The other WWPN is used when you move the client logical partition to another managed system (PowerVM Live Partition Mobility). The virtual Fibre Channel adapters always has a one-to-one relationship between the virtual I/O client partitions and the

virtual Fibre Channel adapters in the Virtual I/O Server partition. That is, each virtual Fibre Channel client adapter that is assigned to a virtual I/O client partition must connect to only one virtual Fibre Channel server adapter in the Virtual I/O Server partition, and each virtual Fibre Channel server adapter in the Virtual I/O Server partition must connect to only one virtual Fibre Channel client adapter in a virtual I/O client partition.

- ▶ Since MPIO is used in the virtual I/O client partition, it can access the physical storage through virtual I/O client virtual Fibre Channel client adapter 1 or 2. If a physical Fibre Channel adapter in the Virtual I/O Server might fail, the virtual I/O client uses the alternate path. This example does not show redundancy in the physical storage, but rather assumes it would be built into the SAN storage device.

### HBA and Virtual I/O Server failover

A Host bus adapter and Virtual I/O Server failover scenario provides a more advanced level of redundancy for the virtual I/O client partition, as shown in Figure 2-41.

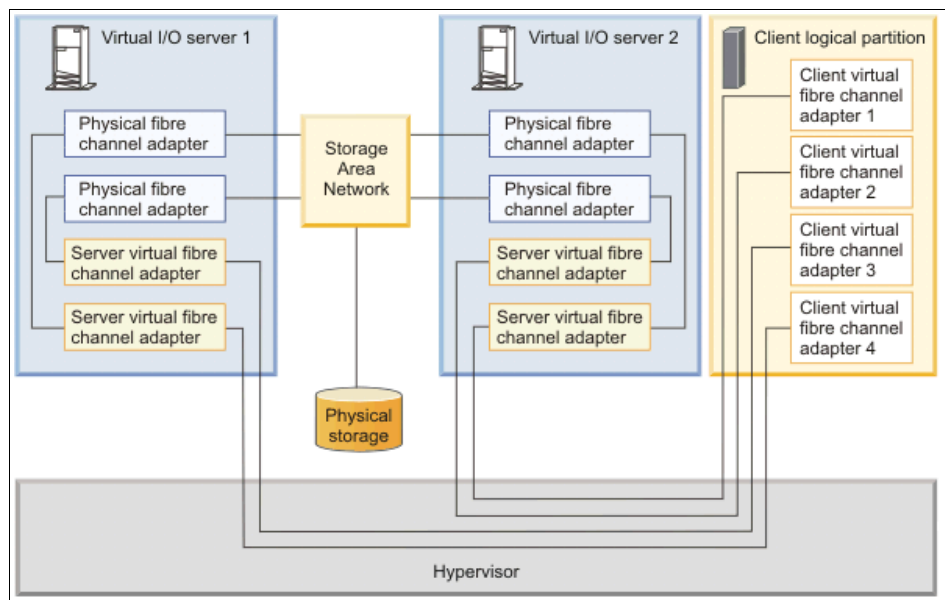


Figure 2-41 Host bus adapter and Virtual I/O Server failover

Figure 2-41 shows the following connections:

- ▶ The SAN connects physical storage to four physical Fibre Channel adapters located on the managed system.

- ▶ There are two Virtual I/O Server partitions to provide redundancy at the Virtual I/O Server level.
- ▶ Two physical Fibre Channel adapters are assigned to their respective Virtual I/O Server partition and support NPIV.
- ▶ The physical Fibre Channel ports are each connected to a virtual Fibre Channel server adapter on the Virtual I/O Server partition. The two virtual Fibre Channel server adapters on the Virtual I/O Server are connected to ports on two different physical Fibre Channel adapters in order to provide the most redundant solution for the physical adapters.
- ▶ Each virtual Fibre Channel server adapter in the Virtual I/O Server partition is connected to one virtual Fibre Channel client adapter in a virtual I/O client partition. Each virtual Fibre Channel client adapter on each virtual I/O client partition receives a pair of unique WWPNs. The client logical partition uses one WWPN to log into the SAN at any given time. The other WWPN is used when you move the client logical partition to another managed system.

The virtual I/O client partition can access the physical storage through virtual Fibre Channel client adapter 1 or 2 on the client logical partition through VIOS 2. The client can also write to physical storage through virtual Fibre Channel client adapter 3 or 4 on the client logical partition through Virtual I/O Server 1. If a physical Fibre Channel adapter fails on Virtual I/O Server 1, the client uses the other physical adapter connected to Virtual I/O Server 1 or uses the paths connected through Virtual I/O Server 2. If you need to shutdown Virtual I/O Server 1 for maintenance reason, then the client uses the path through Virtual I/O Server 2. This example does not show redundancy in the physical storage, but rather assumes it would be built into the SAN.

### ***Considerations for NPIV***

These examples can become more complex as you add physical storage redundancy and multiple clients, but the concepts remain the same. Consider the following points:

- ▶ To avoid configuring the physical Fibre Channel adapter to be a single point of failure for the connection between the virtual I/O client partition and its physical storage on the SAN, do not connect two virtual Fibre Channel client adapters from the same virtual I/O client partition to the same physical Fibre Channel adapter in the Virtual I/O Server partition. Instead, connect each virtual Fibre Channel server adapter to a different physical Fibre Channel adapter.
- ▶ Consider load balancing when mapping a virtual Fibre Channel server adapter in the Virtual I/O Server partition to a physical port on the physical Fiber Channel adapter.

- ▶ Consider what level of redundancy already exists in the SAN to determine whether to configure multiple physical storage units.
- ▶ Consider using two Virtual I/O Server partitions. Since the Virtual I/O Server is central to communication between virtual I/O client partitions and the external network, it is important to provide a level of redundancy for the Virtual I/O Server. Multiple Virtual I/O Server partitions require more resources as well, so you should plan accordingly.
- ▶ NPIV technology is useful when you want to move logical partitions between servers. For example, in an active PowerVM Live Partition Mobility environment, if you use the redundant configurations above, in combination with physical adapters, you can stop all the I/O activity through the dedicated, physical adapter and direct all traffic through a virtual Fibre Channel client adapter until the virtual I/O client partition is successfully moved. The dedicated physical adapter would need to be connected to the same storage as the virtual path. Since you cannot migrate a physical adapter, all I/O activity is routed through the virtual path while you move the partition. After the logical partition is moved successfully, you need to set up the dedicated path (on the destination virtual I/O client partition) if you want to use the same redundancy configuration you had configured on the original logical partition. Then the I/O activity can resume through the dedicated adapter, using the virtual Fibre Channel client adapter as a secondary path.

## 2.9.7 Replacing a Fibre Channel adapter configured with NPIV

This section shows a procedure to deactivate and remove a NPIV Fibre Channel adapter. This procedure can be used for removing or replacing such adapters. See Example 2-42 on page 93 for how to remove the adapter in the Virtual I/O Server.

The adapter need to be unconfigured or removed from the operating system before it can be physically removed.

- ▶ First identify the adapter to be removed. For a dual port card, both ports must be removed.
- ▶ In the Virtual I/O Server the mappings must be unconfigured.
- ▶ The Fibre Channel adapters and their child devices must be unconfigured or deleted. If deleted, they are recovered with the **cfgdev** command for the Virtual I/O Server or the **cfgmgr** command in AIX.
- ▶ The adapter can then be removed using the **diagmenu** command in the Virtual I/O Server or the **diag** command in AIX.



*Example 2-42 Removing a NPIV Fibre Channel adapter in the Virtual I/O Server*


---

```

$ lsdev -dev fcs4 -child
name          status      description
fcnet4        Defined    Fibre Channel Network Protocol Device
fscsi4        Available  FC SCSI I/O Controller Protocol Device
$ lsdev -dev fcs5 -child
name          status      description
fcnet5        Defined    Fibre Channel Network Protocol Device
fscsi5        Available  FC SCSI I/O Controller Protocol Device

$ rmdev -dev vfchost0 -ucfg
vfchost0 Defined
$ rmdev -dev vfchost1 -ucfg
vfchost1 Defined

$ rmdev -dev fcs4 -recursive -ucfg
fscsi4 Defined
fcnet4 Defined
fcs4 Defined

$ rmdev -dev fcs5 -recursive -ucfg
fscsi5 Defined
fcnet5 Defined
fcs5 Defined

diagmenu

```

---

In the DIAGNOSTIC OPERATING INSTRUCTIONS menu press Enter and select

**Task Selection** → **Hot Plug Task** → **PCI Hot Plug Manager** →  
**Replace/Remove a PCI Hot Plug Adapter**

Select the adapter to be removed and follow the instructions on the screen.

**Important:** When replacing a physical NPIV adapter in the Virtual I/O Server, the virtual WWPNs are retained and no new mappings, zoning or LUN assignments need to be updated.

## 2.9.8 Migration to virtual Fibre Channel adapter environments

In this section you will find information about the migration of your existing environment to a NPIV based environment.

## Migration from physical to virtual channel adapter

You can migrate any rootvg or non-rootvg disk assigned from a LUN which is mapped through a physical Fibre Channel adapter to a virtual Fibre Channel mapped environment.

The next steps explain explicitly the migration. In the example `vios1` is used as the name for the Virtual I/O Server partition and NPIV is used as the name for the virtual I/O client partition.

1. Example 2-43 shows that in the NPIV partition a physical Fibre Channel adapter with two ports is assigned. A LUN is mapped to this physical Fibre Channel adapter within the IBM DS4800 storage system, as shown in Figure 2-42 on page 94. There is one path available to `hdisk0`.

### Example 2-43 Show available Fibre Channel adapters

```
# lscfg |grep fcs
+ fcs0          U789D.001.DQDYKYW-P1-C1-T1          8Gb PCI
Express Dual Port FC Adapter (df1000f114108a03)
+ fcs1          U789D.001.DQDYKYW-P1-C1-T2          8Gb PCI
Express Dual Port FC Adapter (df1000f114108a03)

# lscfg |grep disk
* hdisk0        U789D.001.DQDYKYW-P1-C1-T2-W203200A0B811A662-L0  MPIIO Other
DS4K Array Disk

# lspath
Enabled hdisk0 fscs11
```

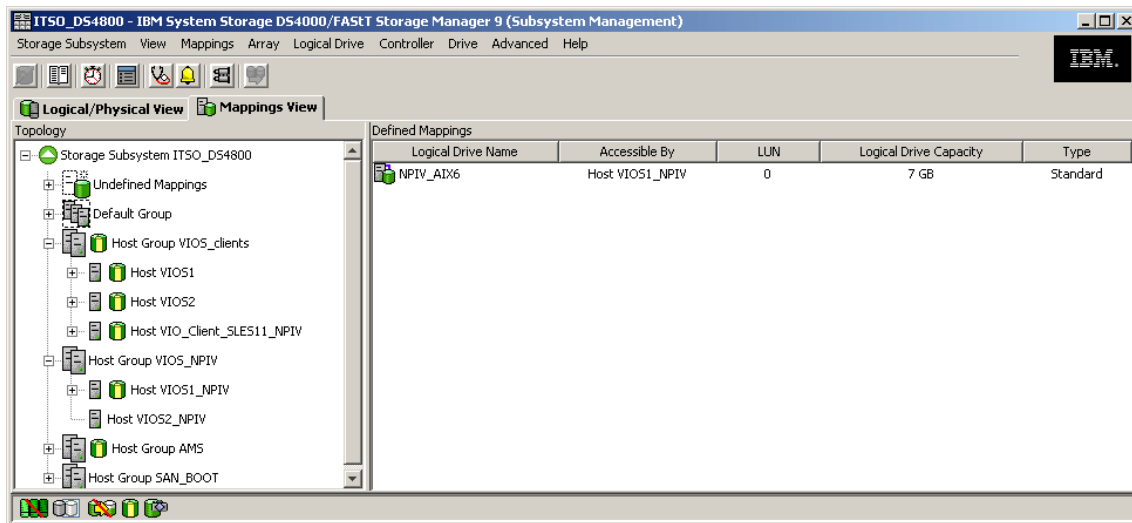


Figure 2-42 LUN mapped to a physical Fibre Channel adapter

2. Add a virtual Fibre Channel server adapter to the Virtual I/O Server partition. In the HMC select your managed server and the Virtual I/O Server partition `vios1`. Click **Tasks** → **Dynamic Logical Partitioning** → **Virtual Adapters** as shown in Figure 2-43.

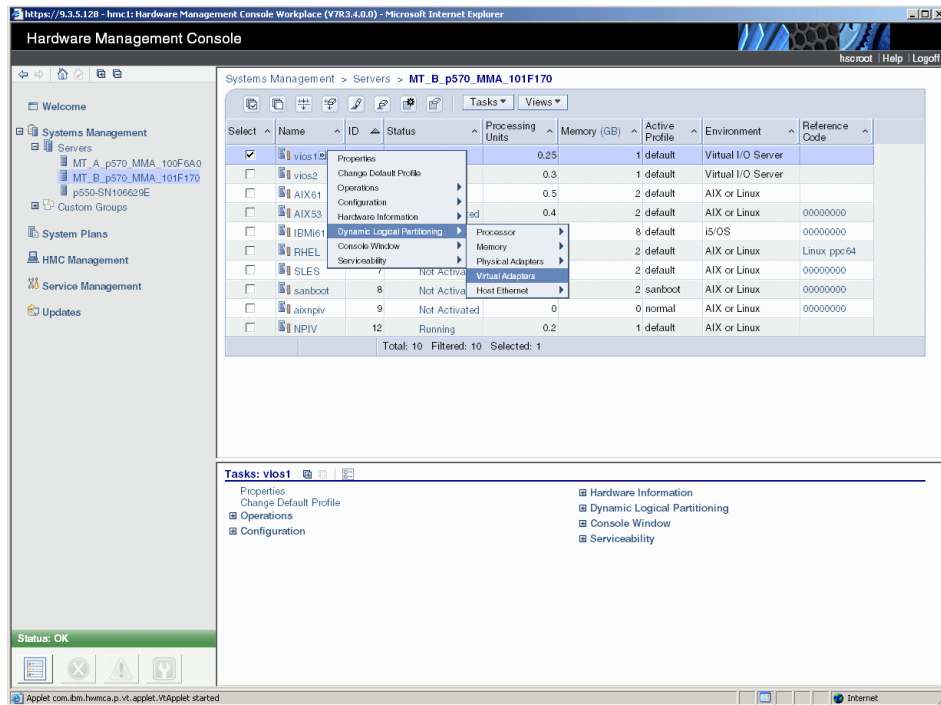


Figure 2-43 Add Virtual Adapter to vios1 partition

3. Click **Actions** → **Create** → **Fibre Channel Adapter...** as shown in Figure 2-44.

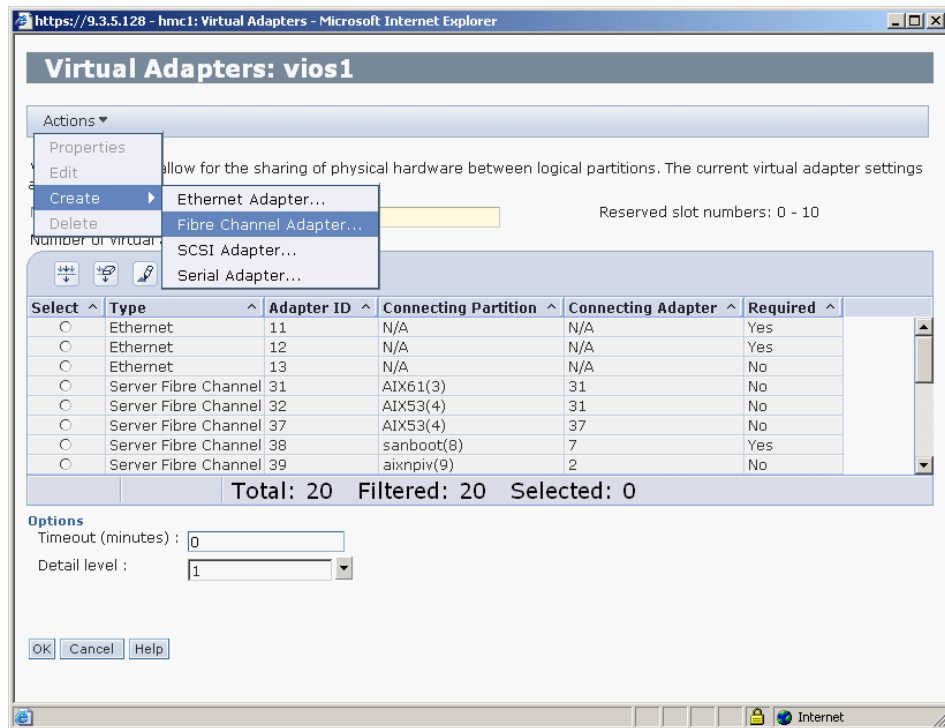


Figure 2-44 Create virtual Fibre Channel server adapter in vios1 partition

4. Enter the Adapter ID for the virtual Fibre Channel server adapter, the name of the partition it should be connected to, and the Client Adapter ID for the slot number of the virtual Fibre Channel client adapter. Then click **OK** as shown in Figure 2-45.

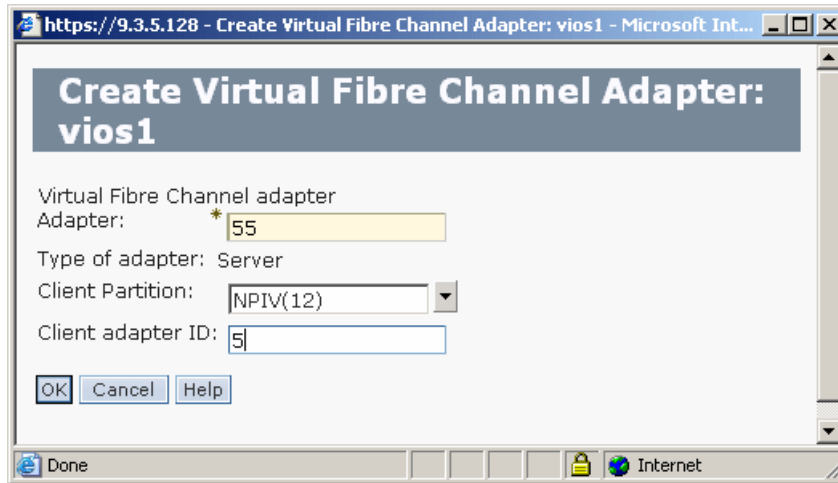


Figure 2-45 Set Adapter IDs in vios1 partition

5. Click **OK**.
6. Add a virtual Fibre Channel client adapter to the virtual I/O client partition. In the HMC select your managed server and the partition NPIV. Click **Tasks** → **Dynamic Logical Partitioning** → **Virtual Adapters** as shown in Figure 2-46.

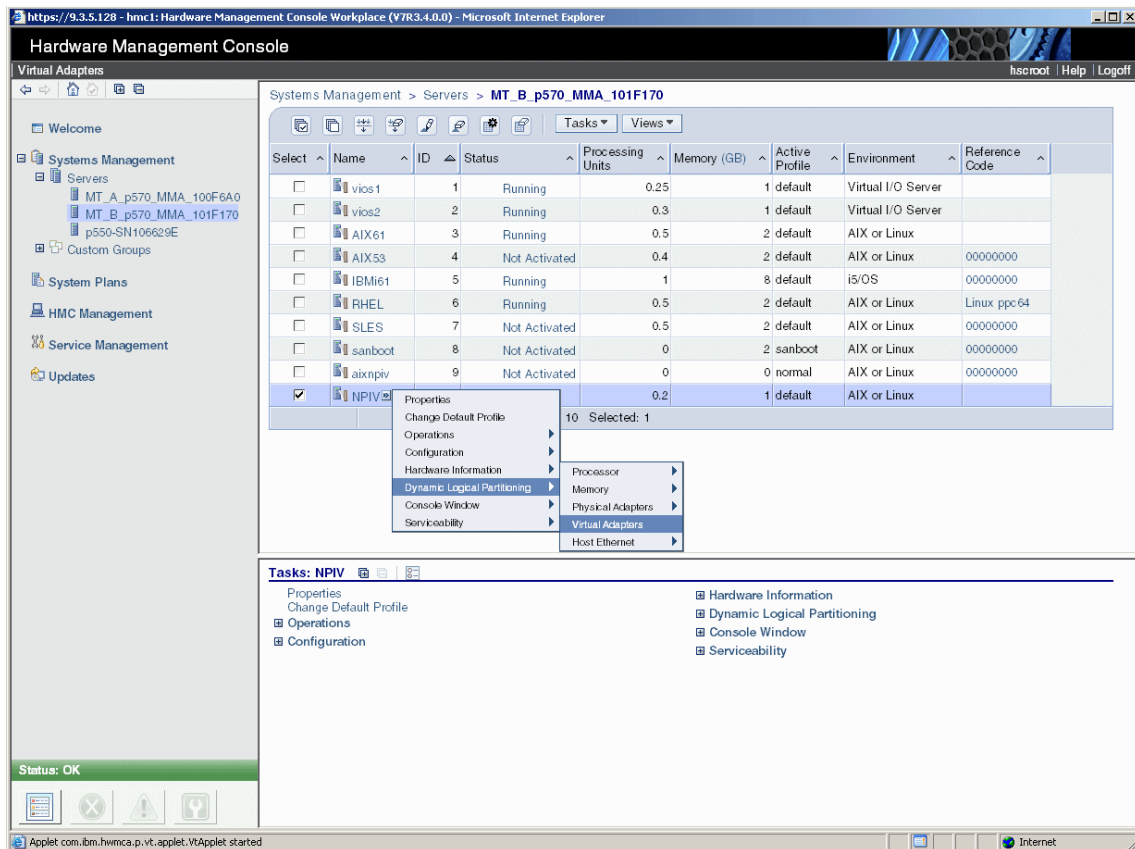


Figure 2-46 Add a virtual adapter to NPIV partition

- Click **Actions** → **Create** → **Fibre Channel Adapter...** as shown in Figure 2-47.

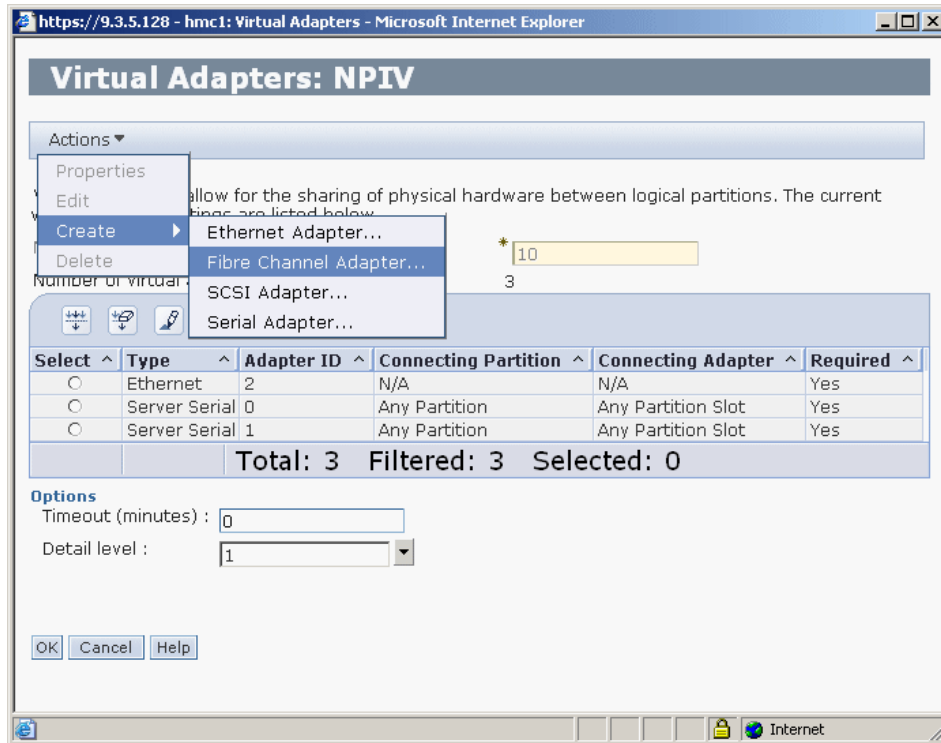


Figure 2-47 Create virtual Fibre Channel client adapter in NPV partition

8. Enter the Adapter ID for the virtual Fibre Channel client adapter, the name of the Virtual I/O Server partition it should be connected and the Server adapter ID for the slot number of the virtual Fibre Channel server adapter and click **OK** as shown in Figure 2-48.

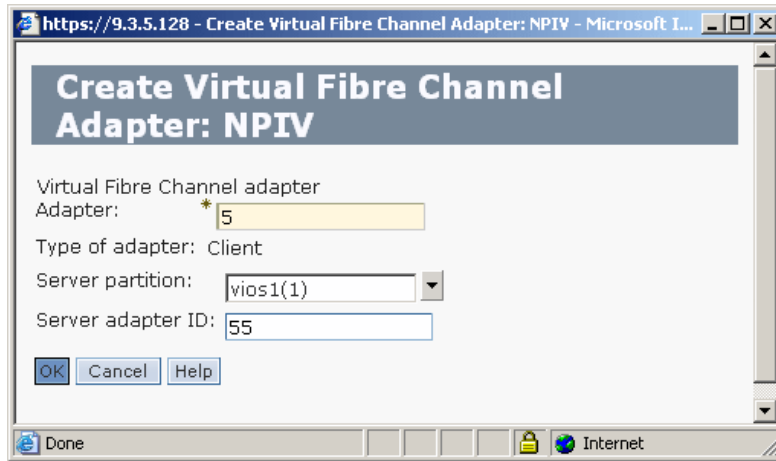


Figure 2-48 Set Adapter IDs in NPIV partition

9. Click **OK**.
10. Log in as `padmin` to the Virtual I/O Server partition `vios1`.
11. Check all available virtual Fibre Channel server adapters with the `lsdev` command:

```
$ lsdev -dev vfchost*
name          status      description
vfchost0     Available  Virtual FC Server Adapter
vfchost1     Available  Virtual FC Server Adapter
vfchost2     Available  Virtual FC Server Adapter
vfchost3     Available  Virtual FC Server Adapter
vfchost4     Available  Virtual FC Server Adapter
vfchost5     Available  Virtual FC Server Adapter
```

12. Run the `cfgdev` command to configure the previously added virtual Fibre Channel server adapter:

```
$ cfgdev
```

13. Run the `lsdev` command again to show the newly configured virtual Fibre Channel server adapter `vfchost6`:

```
$ lsdev -dev vfchost*
name          status      description
vfchost0     Available  Virtual FC Server Adapter
vfchost1     Available  Virtual FC Server Adapter
vfchost2     Available  Virtual FC Server Adapter
vfchost3     Available  Virtual FC Server Adapter
vfchost4     Available  Virtual FC Server Adapter
```



```
vfchost5      Available  Virtual FC Server Adapter
vfchost6    Available  Virtual FC Server Adapter
```

14. Double check the vial product data of vfchost6 for the slot numbering using the **lsdev** command:

```
$ lsdev -dev vfchost6 -vpd
vfchost6      U9117.MMA.101F170-V1-C55  Virtual FC Server
Adapter
```

```
Hardware Location Code.....U9117.MMA.101F170-V1-C55
```

```
PLATFORM SPECIFIC
```

```
Name: vfc-server
Node: vfc-server@30000037
Device Type: fcp
Physical Location: U9117.MMA.101F170-V1-C55
```

**Note:** As previously defined in the HMC, vfchost6 is available in slot 55.

15. Map the virtual Fibre Channel server adapter vfchost6 with the physical Fibre Channel adapter fcs3 running the **vfcmap** command:

```
$ vfcmap -vadapter vfchost6 -fcp fcs3
vfchost6 changed
```

16. Check the mapping with the **lsmap -all -npiv** command:

```
$ lsmap -npiv -vadapter vfchost6
Name          Physloc                                CIntID CIntName          CIntOS
-----
vfchost6      U9117.MMA.101F170-V1-C55              12
```

```
Status:NOT_LOGGED_IN
FC name:fcs3          FC loc code:U789D.001.DQDYKYW-P1-C6-T2
Ports logged in:0
Flags:4<NOT_LOGGED>
VFC client name:      VFC client DRC:
```

17. Log in to the virtual I/O client partition named NPV.

18. Run the **cfgmgr** command to configure the previously defined virtual Fibre Channel client adapter and check all available Fibre Channel adapters with the **lsdev** command:

```
# lscfg |grep fcs
+ fcs2          U9117.MMA.101F170-V12-C5-T1          Virtual
Fibre Channel Client Adapter
```

```

+ fcs0          U789D.001.DQDYKYW-P1-C1-T1          8Gb PCI
Express Dual Port FC Adapter (df1000f114108a03)
+ fcs1          U789D.001.DQDYKYW-P1-C1-T2          8Gb PCI
Express Dual Port FC Adapter (df1000f114108a03)

```

A new virtual Fibre Channel client adapter fcs2 has been added to the operating system.

19. Get the WWPN of this Virtual Fibre Channel client adapter with the `lscfg` command as shown in Example 2-44.

*Example 2-44 WWPN of the virtual Fibre Channel client adapter in NPIV partition*

---

```

# lscfg -vl fcs2
  fcs2          U9117.MMA.101F170-V12-C5-T1  Virtual Fibre Channel
Client Adapter

Network Address.....C05076000AFE0034
ROS Level and ID.....
Device Specific.(Z0).....
Device Specific.(Z1).....
Device Specific.(Z2).....
Device Specific.(Z3).....
Device Specific.(Z4).....
Device Specific.(Z5).....
Device Specific.(Z6).....
Device Specific.(Z7).....
Device Specific.(Z8).....C05076000AFE0034
Device Specific.(Z9).....
Hardware Location Code.....U9117.MMA.101F170-V12-C5-T1

```

---

20. Log in to your SAN switch and zone the WWPN of the virtual Fibre Channel client adapter as shown in Example 2-45 for the IBM 2109-F32.

*Example 2-45 Zoning of WWPN for fcs2*

---

```

itsosan02:admin> portloginshow 15
Type  PID      World Wide Name      credit df_sz cos
=====
fe  660f02  c0:50:76:00:0a:fe:00:34  40  2048  c  scr=3
fe  660f01  c0:50:76:00:0a:fe:00:14  40  2048  c  scr=3
fe  660f00  10:00:00:00:c9:74:a4:75  40  2048  c  scr=3
ff  660f02  c0:50:76:00:0a:fe:00:34  12  2048  c  d_id=FFFFFFC
ff  660f01  c0:50:76:00:0a:fe:00:14  12  2048  c  d_id=FFFFFFC
ff  660f00  10:00:00:00:c9:74:a4:75  12  2048  c  d_id=FFFFFFC

itsosan02:admin> zoneadd "vios1", "c0:50:76:00:0a:fe:00:34"
itsosan02:admin> cfgsave
You are about to save the Defined zoning configuration. This

```

```
action will only save the changes on Defined configuration.
Any changes made on the Effective configuration will not
take effect until it is re-enabled.
Do you want to save Defined zoning configuration only? (yes, y, no, n): [no] y
Updating flash ...
itsosan02:admin> cfgenable npiv
You are about to enable a new zoning configuration.
This action will replace the old zoning configuration with the
current configuration selected.
Do you want to enable 'npiv' configuration (yes, y, no, n): [no] y
zone config "npiv" is in effect
Updating flash ...
itsosan02:admin> zoneshow
Defined configuration:
  cfg:  npiv   vios1; vios2
  zone:  vios1  20:32:00:a0:b8:11:a6:62; c0:50:76:00:0a:fe:00:14;
           10:00:00:00:c9:74:a4:95; c0:50:76:00:0a:fe:00:34
  zone:  vios2  C0:50:76:00:0A:FE:00:12; 20:43:00:a0:b8:11:a6:62

Effective configuration:
  cfg:  npiv
  zone:  vios1  20:32:00:a0:b8:11:a6:62
           c0:50:76:00:0a:fe:00:14
           10:00:00:00:c9:74:a4:95
           c0:50:76:00:0a:fe:00:34
  zone:  vios2  c0:50:76:00:0a:fe:00:12
           20:43:00:a0:b8:11:a6:62
```

---

21. Add a new host port for the WWPN to the DS4800 as shown in Figure 2-49.

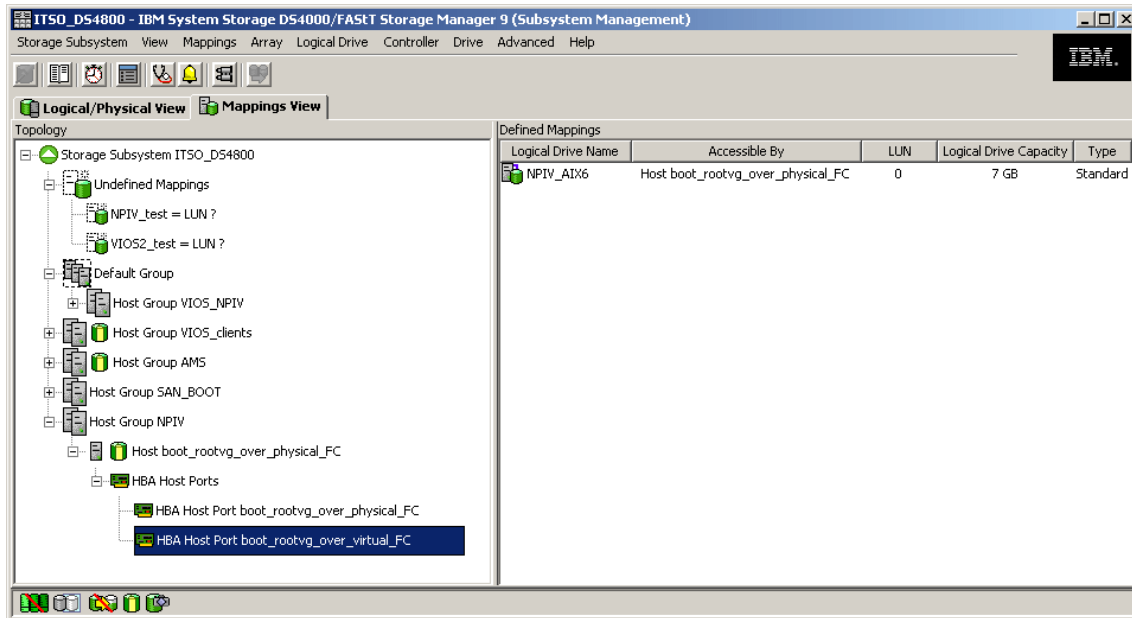


Figure 2-49 Add new host port

22. Run the **cfgmgr** command again to define a second path to the existing disk:

```
# lspath
Enabled hdisk0 fcs11
Enabled hdisk0 fcs12
```

23. Before you can remove the physical Fibre Channel adapter, you have to remove the device from the operating system using the **rmdev** command:

```
# rmdev -d1 fcs0 -R
fcnet0 deleted
fcs10 deleted
fcs0 deleted
```

```
# rmdev -d1 fcs1 -R
fcnet1 deleted
fcs11 deleted
fcs1 deleted
```

**Note:** You have to remove two fcs devices, because this is a two port Fibre Channel adapter.

24. Run the **bosboot** and **bootlist** command to update your boot record and boot list.

```
# bosboot -ad /dev/hdisk0
```

```
bosboot: Boot image is 39570 512 byte blocks.
```

```
# bootlist -m normal hdisk0
```

```
# bootlist -m normal -o  
hdisk0 blv=hd5
```

25. On the HMC, select partition NPIV and then **Tasks** → **Dynamic Logical Partitioning** → **Physical Adapters** → **Move or Remove** as shown in Figure 2-50.

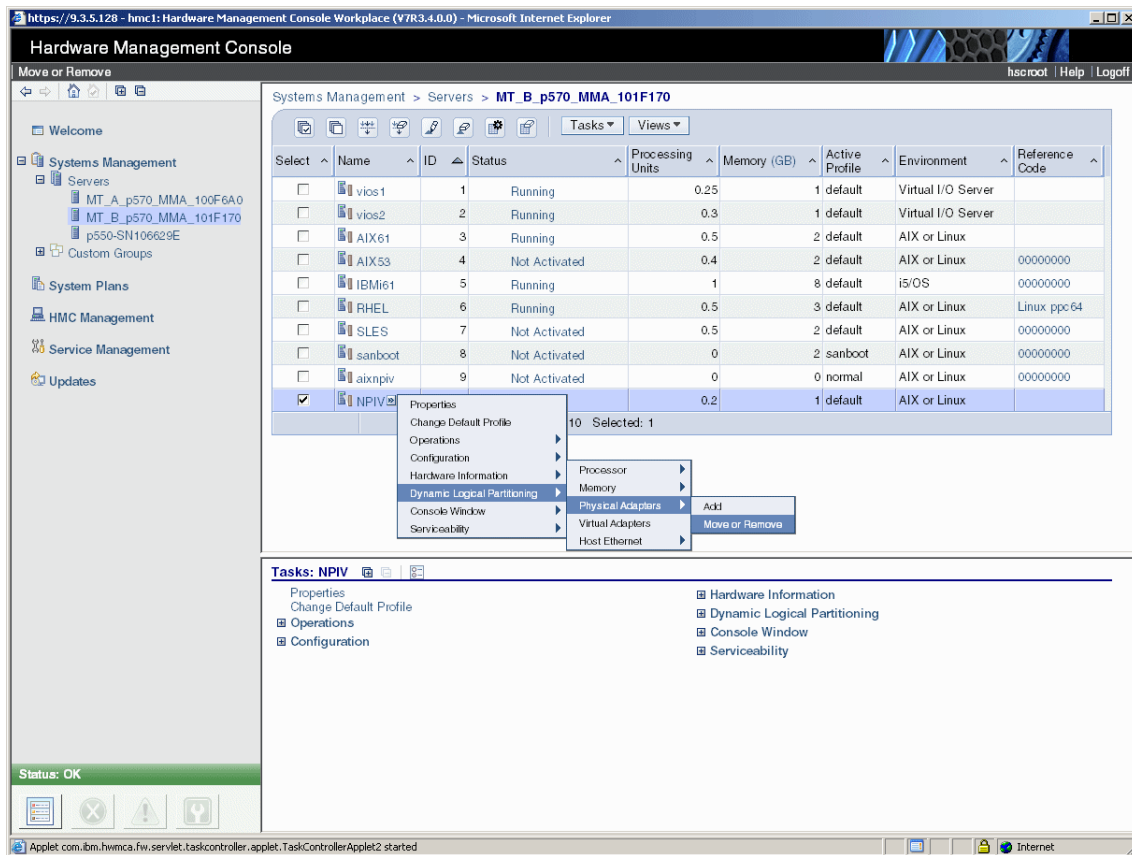


Figure 2-50 Remove a physical Fibre Channel adapter

26. Select the physical adapter which you want to remove from the list and click **OK**, as shown in Figure 2-51.

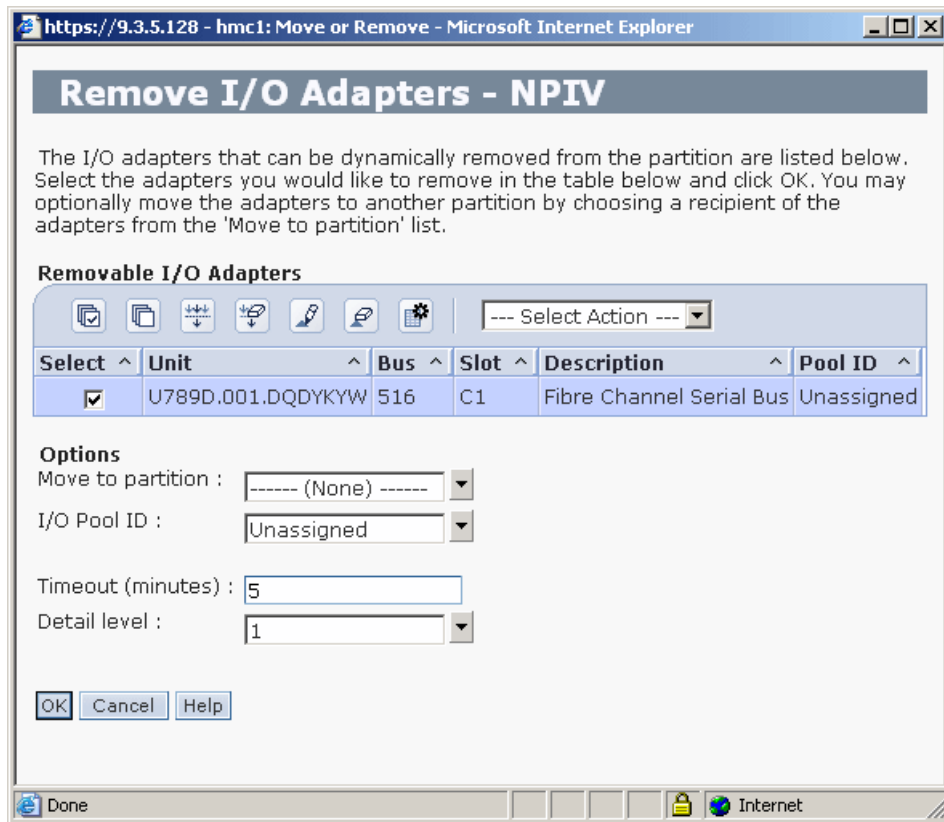


Figure 2-51 Select adapter to be removed

**Note:** Make sure that the adapter to be removed, is defined as desired in the partition profile. Otherwise it can not be removed.

### Migration from vSCSI to NPIV

The migration from a LUN which is mapped to a virtual I/O client through the Virtual I/O Server is not supported. You can not remove the vSCSI mapping and then remap it to a WWPN coming from the virtual Fibre Channel adapter.

If you want to migrate rootvg disks, you have four options:

1. Mirroring

You can create additional disks mapped over NPIV. Then you can add this disks to the rootvg using the command:

```
$ lspv
```

```

hdisk0          00ca58bd2ed277ef          rootvg          active
hdisk1          00ca58bd2f512b88          None
$ extendvg -f rootvg hdisk1
$ lspv
hdisk0          00ca58bd2ed277ef          rootvg          active
hdisk1          00ca58bd2f512b88          rootvg          active
$

```

After you have added hdisk1 to the rootvg, you can mirror hdisk0 to hdisk1 using the **mirrorvg** command and boot from hdisk1.

## 2. Alternate disk installation

Create additional disks mapped over NPIV, and then use alternate disk installation method.

**Note:** For more information search for *alternate disk installation* in the IBM AIX Information Center at:

<http://publib.boulder.ibm.com/infocenter/systems/scope/aix/index.jsp>

## 3. migratepv command

If you want to migrate a disk, you can create additional disks mapped over NPIV. Then you can migrate the disks running the **migratepv** command, which moves physical partitions from one physical volume to one or more physical volumes.

## 4. NIM backup and restore

You can backup the rootvg onto a NIM server and then restore it to a new disk mapped over NPIV. Detailed information on NIM backup and restore can be found in *NIM from A to Z in AIX 5L*, SG24-7296.

## 2.9.9 Heterogeneous configuration with NPIV

It is supported to combine virtual Fibre Channel client adapters with physical adapters using native MPIO as shown in Figure 2-52. One virtual Fibre Channel client adapter and one physical adapter forms two paths to the same LUN.

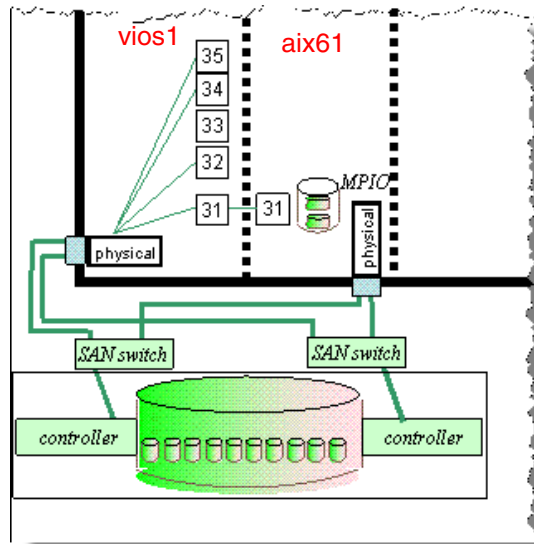


Figure 2-52 Heterogeneous NPIV configuration

### Verifying MPIO in a heterogeneous configuration

Example 2-46 shows how to verify MPIO in a configuration with a physical Fibre Channel adapter in combination with a virtual Fibre Channel client adapter. In this example only one physical port is used on the physical adapter. Using two ports on the physical adapter and one virtual adapter would result in three paths to hdisk2.

Example 2-46 Verifying MPIO in a heterogeneous configuration

```
# lsdev -l fcs*
fcs0 Available 00-00 8Gb PCI Express Dual Port FC Adapter (df1000f114108a03)
fcs1 Available 00-01 8Gb PCI Express Dual Port FC Adapter (df1000f114108a03)
fcs2 Available 31-T1 Virtual Fibre Channel Client Adapter

# lsdev -p fcs0
fcnet0 Defined 00-00-01 Fibre Channel Network Protocol Device
fscsi0 Available 00-00-02 FC SCSI I/O Controller Protocol Device

# lsdev -p fcs1
fcnet1 Defined 00-01-01 Fibre Channel Network Protocol Device
fscsi1 Available 00-01-02 FC SCSI I/O Controller Protocol Device

# lsdev -p fcs2
fscsi2 Available 31-T1-01 FC SCSI I/O Controller Protocol Device
```



```
# cfgmgr

# lspv
hdisk0      00c1f170e327afa7      rootvg      active
hdisk1      00c1f170e170fbb2      None
hdisk2      00c1f17045453fc1      None

# lspath
Enabled hdisk0 vscsi0
Enabled hdisk1 vscsi0
Enabled hdisk0 vscsi1
Enabled hdisk2 fscsi1
Enabled hdisk2 fscsi2

# lscfg -vl fscsi1
fscsi1 U789D.001.DQDYKYW-P1-C1-T2 FC SCSI I/O Controller Protocol Device

# lscfg -vl fscsi2
fscsi2 U9117.MMA.101F170-V3-C31-T1 FC SCSI I/O Controller Protocol Device

# lspath -El hdisk2 -p fscsi1
scsi_id 0x660e00          SCSI ID      False
node_name 0x200200a0b811a662 FC Node Name False
priority 1              Priority      True

# lspath -El hdisk2 -p fscsi2
scsi_id 0x660e00          SCSI ID      False
node_name 0x200200a0b811a662 FC Node Name False
priority 1              Priority      True
```

---













# Virtual network management

Network connectivity in the virtual environment is extremely flexible. This chapter describes best practices for the virtual network configuration. We discuss how to change the IP or the VLAN in a virtualized environment, along with mapping management and tuning packet sizes for best performance. It is assumed you are well versed in setting up virtual network environment. In case not refer to *PowerVM Virtualization on IBM System p: Introduction and Configuration*, SG24-7940.

## 3.1 Changing IP addresses or VLAN

This section describes how to change the IP address and the VLAN within a virtual I/O environment and its impact on the Virtual I/O Server and the virtual I/O client.

### 3.1.1 Virtual I/O Server network address changes

The following sections pertain to the Virtual I/O Server.

#### Changes to the IP address

The IP address on the Shared Ethernet Adapter is used for:

- ▶ RMC communication for dynamic LPAR on the Virtual I/O Server
- ▶ Logon to the Virtual I/O Server
- ▶ NIM installation or restore (the `installios` command) of the Virtual I/O Server
- ▶ Performance monitoring using the `topas` command
- ▶ Update or upgrade the Virtual I/O Server from the NIM server
- ▶ Back up of the Virtual I/O Server to the NIM server or other network servers

The IP address assigned to the Shared Ethernet Adapter is transparent to the virtual I/O client. Therefore, the IP address on the Shared Ethernet Adapter device can be changed without affecting the virtual I/O client using the Shared Ethernet Adapter device.

If the IP address must be changed on en5, for example, from 9.3.5.108 to 9.3.5.109 and the host name from VIO\_Server1 to VIO\_Server2, use the following command:

```
mktcpip -hostname VIO_Server2 -inetaddr 9.3.5.109 -interface en5
```

If you only want to change the IP address or the gateway of a network interface, you can also use the `chtcpip` command:

```
chtcpip -interface en5 -inetaddr 9.3.5.109
```

Note that if you want to change the adapter at the same time, such as from en5 to en8, you have to delete the TCP/IP definitions on en5 first by using the `rmtcpip` command and then running the `mktcpip` command on en8.

#### Changes to the VLAN

You can add or remove the VLANs on an existing tagged virtual Ethernet adapter using dynamic LPAR without interrupting the service running on that Virtual I/O



Server. Remember to change the partition profile to reflect the dynamic change by either using the save option or the properties option on the partition context menu. It will probably also need a network switch change, and should be coordinated with the network admins of your company.

Another possibility is to use the `smitty vlan` command to create a VLAN AIX device on top of the existing network adapter to place tags onto packets.

To start bridging a new tagged VLAN ID, you can create a new virtual Ethernet with a temporary PVID and any tagged VIDs you want. Then, use dynamic LPAR to move it into the Virtual I/O Server, and use the `chdev` command to add the new adapter to the list of *virt\_adapters* of the SEA. It will immediately begin bridging the new VLAN ID without an interruption.

You can add a new physical Ethernet adapter, a new Shared Ethernet Adapter, and a new virtual Ethernet adapter to make a tagged virtual Ethernet adapter. Doing this, you can move from an untagged to a tagged virtual Ethernet adapter. This requires a small planned service window as the virtual I/O clients are moved from the untagged to the tagged adapter, such as any change of IP address would require in a non-virtualized environment.

This also applies when you move from tagged to untagged virtual Ethernet adapters. We recommend that you plan and document a change from untagged to tagged virtual Ethernet adapters or vice versa.

### 3.1.2 Virtual I/O client network address changes

The following sections pertain to the virtual I/O client.

#### Changes to the IP address

For an AIX virtual I/O client to change the IP address on a virtual Ethernet adapter use SMIT or the `mktcpip` command:

In this example, we change the IP address from 9.3.5.113 to 9.3.5.112 and the host name from lpar03 to lpar02. The virtual Ethernet adapter can be modified in the same way you modify a physical adapter, using the following command:

```
mktcpip -h lpar02 -a 9.3.5.112 -i en0
```

For an IBM i virtual I/O client the following procedure describes how to change the IP address on a physical or virtual Ethernet adapter:

Add a new TCP/IP interface with the new IP address (9.3.5.123) to an existing Ethernet line description (ETH01) using the `ADDTCPIFC` command as follows:

```
ADDTCPIFC INTNETADR('9.3.5.123') LIND(ETH01)
SUBNETMASK('255.255.254.0')
```

Start the new TCP/IP interface using the STRTCPIFC command as follows:

```
STRTCPIFC INTNETADR('9.3.5.123')
```

The TCP/IP interface with the old IP address (9.3.5.119) can now be ended and removed using the ENDTCPIFC and RMVTCPIFC command as follows:

```
ENDTCPIFC INTNETADR('9.3.5.119')
RMVTCPIFC INTNETADR('9.3.5.119')
```

Alternatively the **CFGTCP** command choosing option **1. Work with TCP/IP interfaces** allows a menu-based change of TCP/IP interfaces.

To change the hostname for an IBM i virtual I/O client use the **CHGTCDDMN** command.

## Changes to the VLAN

If you want to change the VLAN information at the Virtual I/O Server, it is possible to add or remove the VLANs on an existing tagged virtual Ethernet adapter using dynamic LPAR without interrupting the network service running to the virtual I/O clients. Adding additional IP addresses at the virtual I/O clients can be done as an alias IP address, which will not interrupt the network service on that virtual I/O client. Keep in mind, as with all dynamic LPAR changes, to change the partition profile by either using the save option or the properties option on the partition context menu.

With virtual I/O clients, you cannot change from an untagged to a tagged virtual Ethernet adapter without interrupting that network service. You can add a new virtual Ethernet adapter and make that a tagged virtual Ethernet adapter. In this way, you can move from an untagged to tagged virtual Ethernet adapter requiring a small planned service window as the virtual I/O client is moved from the untagged to the tagged adapter. This is the same interruption as a change of IP address would require in a non-virtualized environment.

This also applies when you move from tagged to untagged virtual Ethernet adapters. We recommend that you plan and document a change from untagged to tagged virtual Ethernet adapters or tagged to untagged.

## 3.2 Managing the mapping of network devices

One of the keys to managing a virtual environment is keeping track of what virtual objects correspond to what physical objects. In the network area, this can involve physical and virtual network adapters, and VLANs that span across hosts

and switches. This mapping is critical for managing performance and to understand what systems will be affected by hardware maintenance.

In environments that require redundant network connectivity, this section focuses on the SEA failover method in preference to the Network Interface Backup method of providing redundancy.

Depending on whether you choose to use 802.1Q tagged VLANs, you might need to track the following information:

- ▶ Virtual I/O Server
  - Server host name
  - Physical adapter device name
  - Switch port
  - SEA device name
  - Virtual adapter device name
  - Virtual adapter slot number
  - Port virtual LAN ID (in tagged and untagged usages)
  - Additional virtual LAN IDs
- ▶ Virtual I/O client
  - Client host name
  - Virtual adapter device name
  - Virtual adapter slot number
  - Port virtual LAN ID (in tagged and untagged usages)
  - Additional virtual LAN IDs

Because of the number of fields to be tracked, we recommend the use of a spreadsheet or database program to track this information. Record the data when the system is installed, and track it over time as the configuration changes.

### 3.2.1 Virtual network adapters and VLANs

Virtual network adapters operate at memory speed. In many cases where additional physical adapters are needed, there is no need for additional virtual adapters. However, transfers that remain inside the virtual environment can benefit from using large MTU sizes on separate adapters. This can lead to improved performance and reduced CPU utilization for transfers that remain inside the virtual environment.

The POWER Hypervisor™ supports tagged VLANs that can be used to separate traffic in the system. Separate adapters can be used to accomplish the same goal. Which method you choose, or a combination of both, should be based on common networking practice in your data center.

## 3.2.2 Virtual device slot numbers

Virtual storage and virtual network devices have a unique slot number. In complex systems, there tend to be far more storage devices than network devices, because each virtual SCSI device can only communicate with one server or client. We recommend that the slot numbers through 20 be reserved for network devices on all LPARs in order to keep the network devices grouped together. In some complex network environments with many adapters, more slots might be required for networking.

The maximum number of virtual adapter slots per LPAR should be increased above the default value of 10 when you create an LPAR. The appropriate number for your environment depends on the number of LPARs and adapters expected on each system. Each unused virtual adapter slot consumes a small amount of memory, so the allocation should be balanced with expected requirements. To plan memory requirements for your system configuration, use the System Planning Tool available at:

<http://www.ibm.com/servers/eserver/series/lpar/systemdesign.html>

## 3.2.3 Tracing a configuration

Despite the best intentions in record keeping, it sometimes becomes necessary to manually trace virtual network connections back to the physical hardware.

### AIX Virtual Ethernet configuration tracing

For an AIX virtual I/O client partition with multiple virtual network adapters, the slot number of each adapter can be determined using the adapter physical location from the `lscfg` command. In the case of virtual adapters, this field includes the card slot following the letter C as shown in the following example:

```
# lscfg -l ent*
ent0          U9117.MMA.101F170-V3-C2-T1  Virtual I/O Ethernet
Adapter (1-lan)
```

You can use the slot numbers from the physical location field to trace back via the HMC Virtual Network Management option and determine what connectivity and VLAN tags are available on each adapter:

From the HMC **Systems Management** → **Servers** view select your Power Systems server and choose **Configuration** → **Virtual Network Management** as shown in Figure 3-1.

The screenshot shows the Hardware Management Console (HMC) interface in a Microsoft Internet Explorer browser window. The address bar displays the URL: `https://9.3.5.128 - hmc1: Hardware Management Console Workplace (V7R3.4.0.0) - Microsoft Internet Explorer`. The page title is "Hardware Management Console".

The main content area is titled "Systems Management > Servers > MT\_B\_p570\_MMA\_101F170". It features a table with the following columns: Select, Name, ID, Status, Processing Units, Memory (GB), Active Profile, Environment, and Reference Code. The table contains seven rows of data:

Select	Name	ID	Status	Processing Units	Memory (GB)	Active Profile	Environment	Reference Code
<input type="checkbox"/>	vibs1	1	Running	0.25	1	default	Virtual I/O Server	
<input type="checkbox"/>	vibs2	2	Running	0.3	1	default	Virtual I/O Server	
<input type="checkbox"/>	AIX61	3	Open Firmware	0.5	2	default	AIX or Linux	CA00E100
<input type="checkbox"/>	AIX53	4	Open Firmware	0.5	2	default	AIX or Linux	CA00E100
<input type="checkbox"/>	IBM61	5	Running	1	8	default	iS/OS	00000000
<input type="checkbox"/>	RHEL	6	Running	0.5	2	default	AIX or Linux	Linux_ppc64
<input type="checkbox"/>	SLES	7	Running	0.5	2	default	AIX or Linux	SuSE Linux

Below the table, it indicates "Total 7 Filtered: 7 Selected: 0".

The "Tasks: MT\_B\_p570\_MMA\_101F170" section is expanded, showing a list of tasks. The "Virtual Network Management" task is highlighted with a red circle. Other tasks include Properties, Operations, Configuration, Create Logical Partition, System Plans, Shared Processor Pool Management, Partition Availability Priority, Memory Pool Management, View Workload Management Groups, Manage Custom Groups, Manage Partition Data, Manage System Profiles, Connections, Hardware Information, Updates, Serviceability, and Capacity On Demand (CoD).

The status bar at the bottom left shows "Status: OK". The bottom right of the browser window shows the address bar with the URL `Applet.com.ibm.hwmca.fw.servlet.taskcontroller.applet.TaskControllerApplet2 started` and the Internet Explorer logo.

Figure 3-1 HMC Virtual Network Management

Select a VLAN to find out where your AIX virtual I/O client Ethernet adapter is connected to. In our example, the AIX61 LPAR virtual I/O client adapter ent0 in slot2 is on VLAN1 as shown in Figure 3-2.

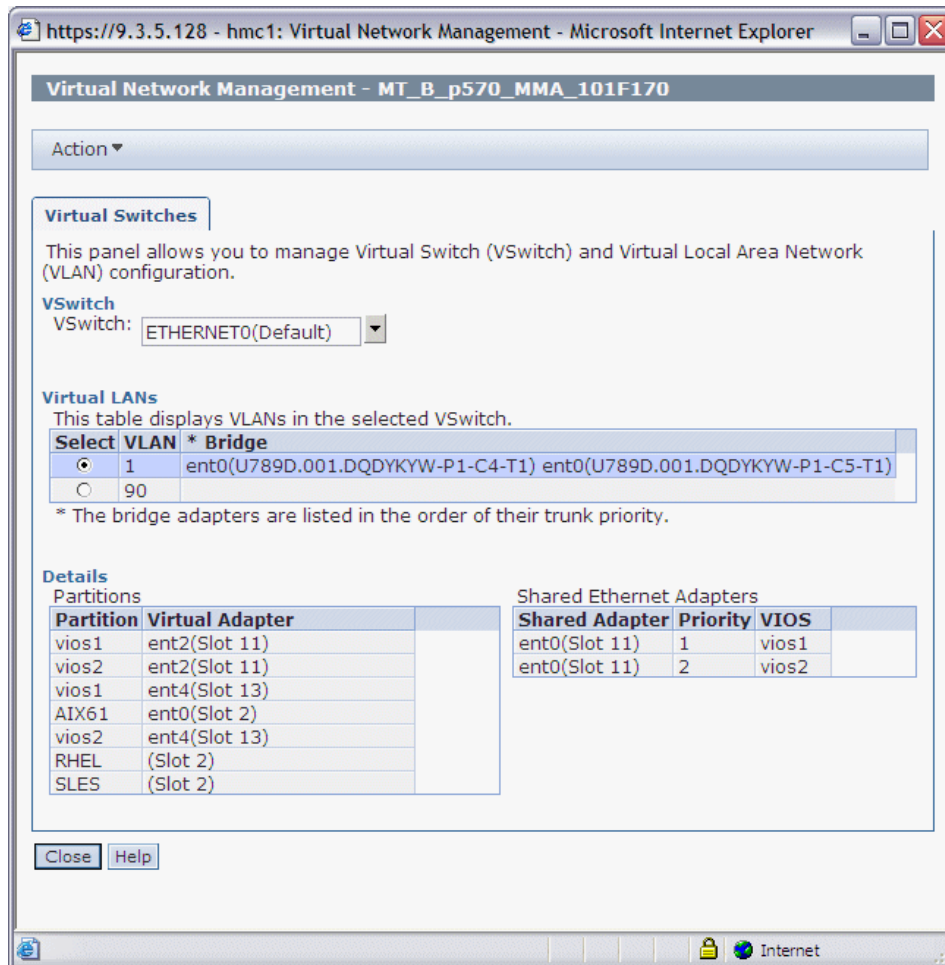


Figure 3-2 Virtual Ethernet adapter slot assignments

**Note:** The HMC Virtual Network Management function currently does not support IBM i partitions.

### IBM i Virtual Ethernet configuration tracing

For an IBM i virtual I/O client partition with virtual Ethernet adapters the slot number of each adapter can be determined using the adapter location information. To display the adapter location information use the `WRKHDWRSC *CMN` command choosing option **7=Display resource detail** for the virtual Ethernet

adapter (type 268C) as shown in Figure 3-3. The location field includes the card slot following the letter C as shown with slot 2 in Figure 3-4.

```

Work with Communication Resources                                     System:E101F170
Type options, press Enter.
  5=Work with configuration descriptions  7=Display resource detail

Opt Resource      Type Status      Text
  CMB06          6B03 Operational Comm Processor
    LIN03          6B03 Operational Comm Adapter
      CMN02          6B03 Operational Comm Port
  CMB07          6B03 Operational Comm Processor
    LIN01          6B03 Operational Comm Adapter
      CMN03          6B03 Operational Comm Port
  CMB08          268C Operational Comm Processor
    LIN02          268C Operational LAN Adapter
  7   CMB01          268C Operational Ethernet Port

Bottom
F3=Exit  F5=Refresh  F6=Print  F12=Cancel

```

Figure 3-3 IBM i Work with Communication Resources screen

```
Display Resource Detail
                                                                    System:E101F170
Resource name . . . . . : CMN01
Text . . . . . : Ethernet Port
Type-model . . . . . : 268C-002
Serial number . . . . . : 00-00000
Part number . . . . . :

Location : U9117.MMA.101F170-V5-C2-T1

Logical address:
SPD bus:
  System bus                255
  System board              128

More...
Press Enter to continue.

F3=Exit  F5=Refresh  F6=Print  F12=Cancel
```

Figure 3-4 IBM i Display Resource Details screen

You can use this slot number from the physical location field to trace back via the HMC partition properties and determine what connectivity and VLAN tags are available on each adapter:

From the HMC **Systems Management** → **Servers** view select your Power Systems server, select your IBM i partition and choose **Properties** to open the partition properties screen as shown in Figure 3-5.



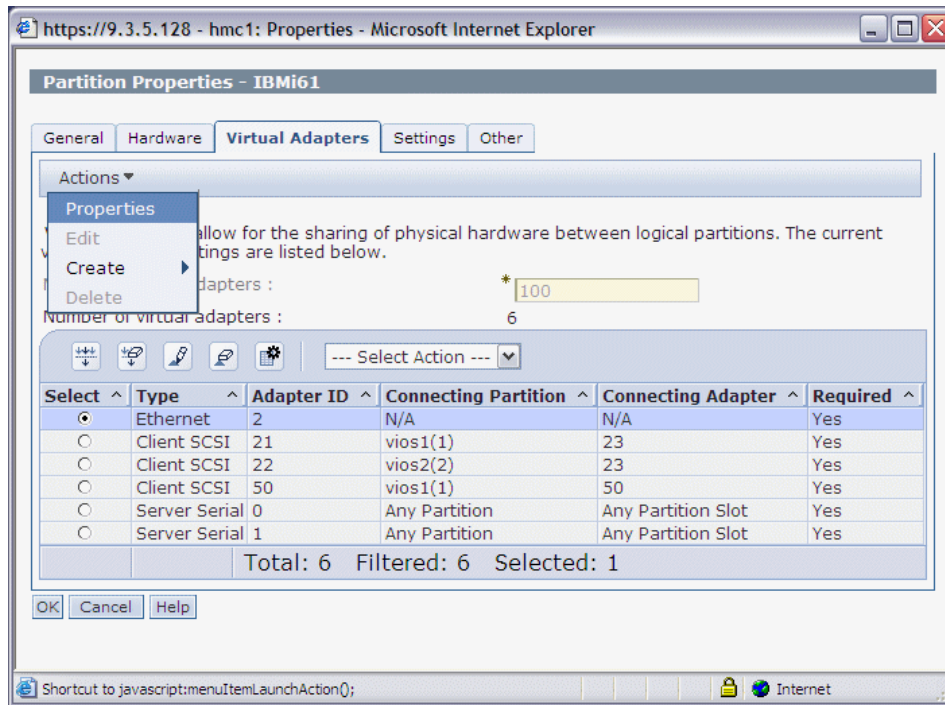


Figure 3-5 HMC IBMi61 partition properties screen

Selecting the IBM i client virtual Ethernet adapter and choosing **Actions** → **Properties** shows for our example that the IBM i client virtual Ethernet adapter in slot 2 is on VLAN1 as shown in Figure 3-6.

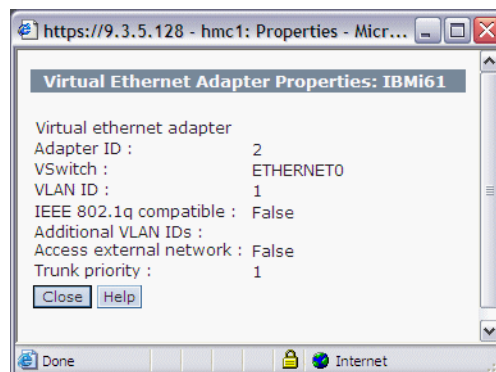


Figure 3-6 HMC virtual Ethernet adapter properties screen

### 3.3 SEA threading on the Virtual I/O Server

The Virtual I/O Server enables you to virtualize both disk and network traffic for IBM AIX, IBM i, and Linux operating system-based clients. The main difference between these types of traffic is their persistence. If the Virtual I/O Server has to move network data around, it must do this immediately because network data has no persistent storage. For this reason, the network services provided by the Virtual I/O Server (such as the Shared Ethernet Adapter) run with the highest priority. Disk data for virtual SCSI devices is run at a lower priority than the network because the data is stored on the disk and there is less of a danger of losing data due to timeouts. The devices are also normally slower.

The shared Ethernet process of the Virtual I/O Server prior to Version 1.3 runs at the interrupt level that was optimized for high performance. With this approach, it ran with a higher priority than the virtual SCSI if there was very high network traffic. If the Virtual I/O Server did not provide enough CPU resource for both, the virtual SCSI performance could experience a degradation of service.

Starting with Virtual I/O Server Version 1.3, the Shared Ethernet function is implemented using kernel threads. This enables a more even distribution of the processing power between virtual disk and network.

This threading can be turned on and off per Shared Ethernet Adapter (SEA) by changing the thread attribute and can be changed while the SEA is operating without any interruption to service. A value of 1 indicates that threading is to be used and 0 indicates the original interrupt method:

```
$ lsdev -dev ent2 -attr thread
value
0
$ chdev -dev ent2 -attr thread=1
ent2 changed
$ lsdev -dev ent2 -attr thread
value
1
```

Using threading requires a minimal increase of CPU usage for the same network throughput; but with the burst nature of network traffic, we recommend enabling threading (this is now the default). By this, we mean that network traffic will come in spikes, as users log on or as Web pages load, for example. These spikes might coincide with disk access. For example, a user logs on to a system, generating a network activity spike, because during the logon process some form

of password database stored on the disk will most likely be accessed or the user profile read.

The one scenario where you should consider disabling threading is where you have a Virtual I/O Server dedicated for network and another dedicated for disk. This is only recommended when mixing extreme disk and network loads together on a CPU constricted server.

Usually the network CPU requirements will be higher than those for disk. In addition, you will probably have the disk Virtual I/O Server setup to provide a network backup with SEA failover if you want to remove the other Virtual I/O Server from the configuration for scheduled maintenance. In this case, you will have both disk and network running through the same Virtual I/O Server, so threading is recommended.

## 3.4 Jumbo frame and path MTU discovery

This section provides information about maximum transfer unit (MTU) and how to use jumbo frames. We also describe the path MTU discovery changes in AIX Version 5.3 and provide recommendations about virtual Ethernet tuning with path MTU discovery.

We discuss the following topics:

- ▶ Maximum transfer unit
- ▶ Path MTU discovery
- ▶ How to use jumbo frame and recommendations for virtual Ethernet

### 3.4.1 Maximum transfer unit

There is a limit on the frame size for Ethernet, IEEE 802.x local area networks, and other networks. The maximum length of an Ethernet frame is 1526 bytes, so it can support a data field length of up to 1500 bytes. Table 3-1 provides typical maximum transmission units (MTUs).

*Table 3-1 Typical maximum transmission units (MTUs)*

Network	MTU (bytes)
Official maximum MTU	65535
Ethernet (10 or 100 MBps)	1500
Ethernet (802.3)	1492

Network	MTU (bytes)
Ethernet (gigabit)	9000
FDDI	4352
TokenRing (802.5)	4464
X.25	576
Official minimum MTU	68

If you send data through a network, and the data is greater than the network's MTU, it can become fragmented because it gets broken up into smaller pieces. Each fragment must be equal to or smaller than the MTU.

The MTU size can affect the network performance between source and target systems. The use of large MTU sizes allows the operating system to send fewer packets of a larger size to reach the same network throughput. The larger packets reduce the processing required in the operating system, because each packet requires the same amount of processing. If the workload is only sending small messages, the larger MTU size will not help.

The maximum segment size (MSS) corresponds to the MTU size minus TCP and IP header information which are 40 bytes for IPv4 and 60 bytes for IPv6. The MSS is the largest data or payload that the TCP layer can send to the destination IP address. When a connection is established, each system announces an MSS value. If one system does not receive an MSS from the other system, it uses the default MSS value.

In AIX Version 5.2 or earlier, the default MSS value was 512 bytes, but starting with AIX Version 5.3 1460 bytes is supported as the default value. If you apply APAR IY57637 in AIX Version 5.2, the default MSS value is changed to 1460 bytes.

The `no -a` command displays the value of the default MSS as `tcp_mssdf1t`. On AIX 6 you receive the information shown in Example 3-1.

*Example 3-1 The default MSS value in AIX 6.1*

---

```
# no -a |grep tcp
tcp_bad_port_limit = 0
tcp_ecn = 0
tcp_ephemeral_high = 65535
tcp_ephemeral_low = 32768
tcp_finwait2 = 1200
tcp_icmpsecure = 0
tcp_init_window = 0
```

```
tcp_inpcb_hashtab_siz = 24499
    tcp_keepcnt = 8
    tcp_keeppidle = 14400
    tcp_keepinit = 150
    tcp_keeppintvl = 150
tcp_limited_transmit = 1
    tcp_low_rto = 0
    tcp_maxburst = 0
    tcp_mssdflt = 1460
    tcp_nagle_limit = 65535
    tcp_nagleoverride = 0
    tcp_ndebug = 100
    tcp_newreno = 1
    tcp_nodelayack = 0
    tcp_pmtu_discover = 1
    tcp_recvspace = 16384
    tcp_sendspace = 16384
    tcp_tcpsecure = 0
    tcp_timewait = 1
    tcp_ttl = 60
tcprexmtthresh = 3
```

---

For IBM i the default MTU size specified, by default, in the Ethernet line description's maximum frame size parameter is 1496 bytes which means 1500 bytes for non encapsulated TCP/IP packets.

If the source network does not receive an MSS when the connection is first established, the system uses the default MSS value. Most network environments are Ethernet, and this can support at least a 1500-byte MTU.

For example, if you execute an FTP application when the MSS value is not received, in AIX Version 5.2 or earlier, the application only uses a 512-byte MSS during the first connection because the default MSS value is 512 bytes, and this can cause degradation in performance. The MSS is negotiated for every connection, so the next connection can use a different MSS.

### 3.4.2 Path MTU discovery

Every network link has a maximum packet size described by the MTU. The datagrams can be transferred from one system to another through many links with different MTU values. If the source and destination system have different MTU values, it can cause fragmentation or dropping of packets while the smallest MTU for the link is selected. The smallest MTU for all the links in a path is called

the path MTU, and the process of determining the smallest MTU along the entire path from the source to the destination is called path MTU discovery (PMTUD).

With AIX Version 5.2 or earlier, the Internet Control Message Protocol (ICMP) echo request and ICMP echo reply packets are used to discover the path MTU using IPv4. The basic procedure is simple. When one system tries to optimize its transmissions by discovering the path MTU, it sends packets of its maximum size. If these do not fit through one of the links between the two systems, a notification from this link is sent back saying what maximum size this link will support. The notifications return an ICMP “Destination Unreachable” message to the source of the IP datagram, with a code indicating “fragmentation needed and DF set” (type 3, type 4).

When the source receives the ICMP message, it lowers the send MSS and tries again using this lower value. This is repeated until the maximum possible value for all of the link steps is found.

Possible outcomes during the path MTU discovery procedure include:

- ▶ The packet can get across all the links to the destination system without being fragmented.
- ▶ The source system can get an ICMP message from any hop along the path to the destination system, indicating that the MSS is too large and not supported by this link.

This ICMP echo request and reply procedure has a few considerations. Some system administrators do not use path MTU discovery because they believe that there is a risk of denial of service (DoS) attacks.

Also, if you already use the path MTU discovery, routers or fire walls can block the ICMP messages being returned to the source system. In this case, the source system does not have any messages from the network environment and sets the default MSS value, which might not be supported across all links.

The discovered MTU value is stored in the routing table using a cloning mechanism in AIX Version 5.2 or earlier, so it cannot be used for multipath routing. This is because the cloned route is always used instead of alternating between the two multipath network routes. For this reason, you can see the discovered MTU value using the **netstat -rn** command.

Beginning with AIX Version 5.3, there are some changes in the procedure for path MTU discovery. Here the ICMP echo reply and request packets are not used anymore. AIX Version 5.3 uses TCP packets and UDP datagrams rather than ICMP echo reply and request packets. In addition, the discovered MTU will not be stored in the routing table. Therefore, it is possible to enable multipath routing to work with path MTU discovery.

When one system tries to optimize its transmissions by discovering the path MTU, a pmtu entry is created in a Path MTU (PMTU) table. You can display this table using the **pmtu display** command, as shown in Example 3-2. To avoid the accumulation of pmtu entries, unused pmtu entries will expire and be deleted when the pmtu\_expire time is exceeded.

*Example 3-2 Path MTU display*

---

```
# pmtu display
```

dst	gw	If	pmtu	refcnt	redisc_t	exp
9.3.4.148	9.3.5.197	en0	1500	1	22	0
9.3.4.151	9.3.5.197	en0	1500	1	5	0
9.3.4.154	9.3.5.197	en0	1500	3	6	0
9.3.5.128	9.3.5.197	en0	1500	15	1	0
9.3.5.129	9.3.5.197	en0	1500	5	4	0
9.3.5.171	9.3.5.197	en0	1500	1	1	0
9.3.5.197	127.0.0.1	lo0	16896	18	2	0
192.168.0.1	9.3.4.1	en0	1500	0	1	0
192.168.128.1	9.3.4.1	en0	1500	0	25	5
9.3.5.230	9.3.5.197	en0	1500	2	4	0
9.3.5.231	9.3.5.197	en0	1500	0	6	4
127.0.0.1	127.0.0.1	lo0	16896	10	2	0

---

Path MTU table entry expiration is controlled by the pmtu\_expire option of the **no** command. The pmtu\_expire option is set to 10 minutes by default.

For IBM i path MTU discovery is enabled by default for negotiation of larger frame transfers. To change the IBM i path MTU discovery setting use the **CHGTCPA** command.

IPv6 never sends ICMPv6 packets to detect the PMTU. The first packet of a connection always starts the process. In addition, IPv6 routers are designed to never fragment packets and always return an ICMPv6 Packet too big message if

they are unable to forward a packet because of a smaller outgoing MTU. Therefore, for IPv6, no changes are necessary to make PMTU discovery work with multipath routing.

### 3.4.3 Using jumbo frames

Jumbo frame support on physical Ethernet adapters under the AIX 6 operating system has a simple design. It is controlled with an attribute on the physical adapter. Virtual Ethernet adapters support all possible MTU sizes automatically.

There is no attribute for jumbo frames on a virtual Ethernet adapter. If an interface is configured on top of a virtual Ethernet adapter, there is an MTU value on the virtual Ethernet interface. Sending jumbo frames from the Shared Ethernet Adapter (SEA) interface is not available on Virtual I/O Server Version 1.5, but bridging jumbo packets is. At the time of writing, packets to and from the SEA interface itself use an MTU of 1500.

However, the primary purpose of SEA is to bridge network communication between the virtual I/O clients and the external network. If the virtual adapter in the virtual I/O clients and the physical Ethernet adapter in the Virtual I/O Server associated with the SEA are all configured to MTU 9000 or have jumbo frames enabled, respectively, the traffic from the virtual I/O clients to the external network can have an MTU of 9000. Although the SEA cannot initiate network traffic using jumbo frames, it is able to bridge this traffic.

To configure jumbo frame communications between AIX or IBM i virtual I/O clients and an external network, use the following steps:

1. For the Virtual I/O Server set an MTU value of 9000 for the physical adapter by enabling jumbo frames with the following command:

```
$ chdev -dev ent0 -attr jumbo_frames=yes
ent0 changed
```

If this physical adapter is in use by the SEA, remove the SEA and then recreate it. You can use the `lsdev -dev ent0 -attr` command to check the jumbo frame attribute of the physical adapter.

2. For the virtual I/O client enable jumbo frames as follows:

For an AIX virtual I/O client change the virtual Ethernet adapter MTU value using the following `chdev` command:

```
# chdev -l en0 -a mtu=9000
en0 changed
# lsattr -El en0
alias4                IPv4 Alias including Subnet Mask      True
alias6                IPv6 Alias including Prefix Length    True
arp                    on                                     Address Resolution Protocol (ARP)     True
```



authority		Authorized Users	True
broadcast		Broadcast Address	True
mtu	9000	Maximum IP Packet Size for This Device	True
netaddr	9.3.5.197	Internet Address	True
netaddr6		IPv6 Internet Address	True
netmask	255.255.254.0	Subnet Mask	True
prefixlen		Prefix Length for IPv6 Internet Address	True
remmtu	576	Maximum IP Packet Size for REMOTE Networks	True
rfc1323		Enable/Disable TCP RFC 1323 Window Scaling	True
security	none	Security Level	True
state	up	Current Interface Status	True
tcp_mssdflt		Set TCP Maximum Segment Size	True
tcp_nodelay		Enable/Disable TCP_NODELAY Option	True
tcp_recvspace		Set Socket Buffer Space for Receiving	True
tcp_sendspace		Set Socket Buffer Space for Sending	True

**Note:** For an AIX partition, virtual Ethernet adapter does not have any attribute for jumbo frame.

For an IBM i virtual I/O client with the default setting of the MTU size defined in the Ethernet line description use the following procedure to enable jumbo frames:

End the TCP/IP interface for the virtual Ethernet adapter using the **ENDTCPIFC** command as follows:

```
ENDTCPIFC INTNETADR('9.3.5.119')
```

Vary off the virtual Ethernet adapter line description using the **VRYCFG** command as follows:

```
VRYCFG CFGOBJ(ETH01) CFGTYPE(*LIN) STATUS(*OFF)
```

Change the corresponding virtual Ethernet adapter line description using the **CHGLIND** command as follows:

```
CHGLINETH LIND(ETH01) MAXFRAME(8996)
```

Vary on the virtual Ethernet adapter line description again using the **VRYCFG** command as follows:

```
VRYCFG CFGOBJ(ETH01) CFGTYPE(*LIN) STATUS(*ON)
```

Start the TCP/IP interface again using the **STRTCPIFC** command as follows:

```
STRTCPIFC INTNETADR('9.3.5.119')
```

Verify jumbo frames are enabled on the IBM i virtual I/O client using the **WRKTCPSTS \*IFC** command and selecting **F11=Display interface status** as shown in Figure 3-7.

```

Work with TCP/IP Interface Status
                                                    System:  E101F170

Type options, press Enter.
  5=Display details  8=Display associated routes  9=Start  10=End
 12=Work with configuration status  14=Display multicast groups

      Internet      Subnet      Type of      Line
Opt  Address      Mask      Service      MTU  Type
   9.3.5.119      255.255.254.0  *NORMAL      8992 *ELAN
   127.0.0.1      255.0.0.0     *NORMAL      576  *NONE

Bottom
F3=Exit  F9=Command line  F11=Display interface status  F12=Cancel
F13=Sort by column      F20=Work with IPv6 interfaces  F24=More
keys

```

Figure 3-7 IBM i Work with TCP/IP Interface Status screen

**Note:** The maximum frame sizes like 1496 or 8996 specified on the IBM i Ethernet line description account for different protocols like SNA or TCP being supported when using the default setting of \*ALL for the Ethernet standard parameter. Setting the line description Ethernet standard to \*ETHV2 allows using the full Ethernet MTU size of 1500 or 9000.

**Important:** Check the port on the switch that is connected to the real adapter associated with the SEA. This must have jumbo frames enabled so jumbo frames send by the virtual I/O clients through the Virtual I/O Server do not get defragmented by the network switch.

### 3.4.4 Virtual Ethernet tuning with path MTU discovery

The MTU and ISNO settings can be tuned to provide better network performance. Note the following considerations to get the best performance in a virtual network:

- ▶ Virtual Ethernet performance is based on CPU capacity entitlement and TCP/IP parameters such as MTU size, buffer size, and rfc1323 settings.
- ▶ If you have large data packets, selecting a high MTU size improves performance because more data per packet can be sent, and therefore, the data is sent using fewer packets.
- ▶ Keep the attributes `tcp_pmtu_discover` and `chksum_offload` set to their default values.
- ▶ Do not turn off simultaneous multithreading unless your applications require it.

### 3.4.5 TCP checksum offload

The TCP checksum offload option enables the network adapter to verify the TCP checksum on transmit and receive, which saves the host CPU from having to compute the checksum. This feature is used to detect a corruption of data in the packet on physical adapters, so virtual Ethernet adapters do not need the checksum process. The checksum offload option in virtual Ethernet adapters is enabled by default, as in physical Ethernet adapters. If you want the best performance between virtual Ethernet adapters, enable the checksum offload option on the source and destination system. If it is disabled in the source or destination system, the hypervisor detects the state and validates the checksum when it is needed. It can be enabled or disabled using the attribute `chksum_offload` of the adapter.

### 3.4.6 Largesend option

The Gigabit or higher Ethernet adapters for IBM Power Systems support TCP segmentation offload (also called largesend). This feature extends the TCP largesend feature to virtual Ethernet adapters and Shared Ethernet Adapters (SEA). In largesend environments, TCP sends a big *chunk* of data to the adapter when TCP knows that the adapter supports largesend. The adapter will break this big TCP packet into multiple smaller TCP packets that will fit the outgoing MTU of the adapter, saving system CPU load and increasing network throughput.

The TCP largesend feature is extended from Virtual I/O client all the way up to the real adapter of VIOS. The TCP stack on the Virtual I/O client will determine

whether the Virtual I/O server supports largesend. If Virtual I/O server supports TCP largesend, the Virtual I/O client sends a big TCP packet directly to Virtual I/O server.

If Virtual Ethernet adapters are used in an LPAR-LPAR environment, however, the large TCP packet does not need to be broken into multiple smaller packets. This is because the underlying hypervisor will take care of sending the big chunk of data from one client to another client.

**Tip:** Largesend is enabled by default on Virtual I/O client for virtual adapters.

To enable largesend on a virtual I/O client adapter use following command on the interface and not adapter:

```
# ifconfig en0 largesend
```

And it can be disabled using:

```
# ifconfig en0 -largesend
```

This feature allows the use of a large MTU for LPAR-LPAR communication, resulting in significant CPU savings and increasing network throughput.

If Virtual I/O Client wants to send large data outside machine then attribute **largesend** needs to be enabled on SEA bridge device and attribute **large\_send** needs to be enabled on physical adapter connected to the SEA. If **large\_send** is not enabled on physical adapter but **largesend** is enabled on SEA bridge adapter then Virtual I/O clients can not send big TCP packets outside of machine. But it can still send packets to another Virtual I/O clients on the same machine.

The largesend option for packets originating from the SEA interface is not available using Virtual I/O Server Version 1.5 (packets coming from the Virtual I/O Server itself).

You can check the largesend option on the SEA using Virtual I/O Server Version 1.5, as shown in Example 3-3. In this example, it is set to off.

#### *Example 3-3 Largesend option for SEA*

```
$ lsdev -dev ent6 -attr
attribute   value   description
user_settable

ctl_chan           Control Channel adapter for SEA failover           True
gvrp               no      Enable GARP VLAN Registration Protocol (GVRP)     True
ha_mode            disabled High Availability Mode                             True
jumbo_frames       no      Enable Gigabit Ethernet Jumbo Frames              True
large_receive      no      Enable receive TCP segment aggregation            True
largesend         0      Enable Hardware Transmit TCP Resegmentation       True
netaddr           0      Address to ping                                    True
```

pvid	1	PVID to use for the SEA device	True
pvid_adapter	ent4	Default virtual adapter to use for non-VLAN-tagged packets	True
real_adapter	ent0	Physical adapter associated with the SEA	True
thread	1	Thread mode enabled (1) or disabled (0)	True
virt_adapters	ent4	List of virtual adapters associated with the SEA (comma separated)	True

It can be enabled using following command:

```
$ chdev -dev ent6 -attr largesend=1
ent6 changed
```

Similarly, **large\_send** can be enabled on physical adapter too on Virtual I/O server. If the physical adapter is connected to SEA then you need to bring it down and then change the attribute otherwise it will fail with device busy message.

### 3.5 SEA enhancements

Few enhancements have been done on Shared ethernet adapter (SEA). IEEE standard 802.1q, Quality of Service (QoS) has been added to SEA and performance metrics can be calculated for SEA based on various parameters. This section explains how QoS works for SEA and how it can be configured.

QoS allows user to divide the network bandwidth on a packet-by-packet basis by setting a user priority when sending the packet.

As explained in Chapter 3.3, “SEA threading on the Virtual I/O Server” on page 126 each SEA instance has certain thread (currently 7) for multiprocessing. Each thread will have 9 queues to take care of network jobs at different priority level. Currently total number of queues are 9. Each queue will take care of the jobs at a priority level and one queue is kept aside which will be used when QoS is disabled.

**Important:** QoS works only for tagged packets, which means all packets emanating from VLAN pseudo device of virtual I/O client. Since virtual Ethernet does not tag a packet so its network traffic cannot be prioritized. It will however go at queue 0 which is default queue at priority level 1.

**Note to Reviewer:**

Each thread will independently follow the same algorithm to determine what queue to send a packet from. A thread will sleep when there are no packets on

any of the nine queues. Once QoS is enabled, SEA will check the priority value of all tagged packets and put that in the corresponding queue.

If QoS is not enabled, then regardless if the packet is tagged or untagged, SEA will ignore the priority value and place all packets in the disabled queue. This will ensure that the packets being enqueued while QoS is disabled will not be sent out of order when QoS is enabled.

When SEA is enabled, there are 2 algorithms to schedule jobs:

### 3.5.1 Strict mode

In this all packets from higher priority queues will be sent before any from a lower priority queue will be sent. SEA adapter will examine the highest priority queue for any packets to send out. If there are any packets to send, the SEA adapter will send that packet. If there are no packets to send in a higher priority queue, the SEA will then check the next highest priority queue for any packets to send out. After sending out a packet from the highest priority queue with packets, the SEA will start the algorithm over again. This does allow for starvation of the lower priority queues.

### 3.5.2 Loose mode

In strict mode its possible lower priority packets starve. Hence loose mode algorithm was devised. If the number of bytes allowed has already been sent out from one priority queue, SEA will check all lower priorities at least once for packets to send before sending out packets from the higher priority again.

When initially sending out packets, SEA will check its highest priority queue. It will continue to send packets out from the highest priority queue until either the queue is empty or the cap is reached. Once either of those two conditions has been met, SEA will then move on to service the next priority queue and will continue using the same algorithm until either of the two conditions mentioned previously have been met, at that point it would move on to the next priority queue. On a fully saturated network, this would allocate certain percentages of bandwidth to each priority. The caps for each priority will be distinct and non-configurable.

### 3.5.3 Setting up QoS

QoS for SEA can be configured using command **chdev**. Attribute to be configured is **qos\_mode** and its value can be **disabled**, **loose** or **strict**. In Example 3-4 on page 139 ent5 is an SEA and its been enabled for loose mode QoS monitoring.

*Example 3-4 Configuring QoS for an SEA*

```

# lsattr -El ent5
accounting enabled Enable per-client accounting of network statistics True
ctl_chan ent3 Control Channel adapter for SEA failover True
gvrp no Enable GARP VLAN Registration Protocol (GVRP) True
ha_mode auto High Availability Mode True
jumbo_frames no Enable Gigabit Ethernet Jumbo Frames True
large_receive no Enable receive TCP segment aggregation True
largesend 0 Enable Hardware Transmit TCP Resegmentation True
netaddr 0 Address to ping True
pvid 1 PVID to use for the SEA device True
pvid_adapter ent2 Default virtual adapter to use for non-VLAN-tagged packets True
qos_mode disabled N/A True
real_adapter ent0 Physical adapter associated with the SEA True
thread 1 Thread mode enabled (1) or disabled (0) True
virt_adapters ent2 List of virtual adapters associated with the SEA (comma separated) True

# chdev -l ent5 -a qos_mode=loose
ent5 changed

# lsattr -El ent5
accounting enabled Enable per-client accounting of network statistics True
ctl_chan ent3 Control Channel adapter for SEA failover True
gvrp no Enable GARP VLAN Registration Protocol (GVRP) True
ha_mode auto High Availability Mode True
jumbo_frames no Enable Gigabit Ethernet Jumbo Frames True
large_receive no Enable receive TCP segment aggregation True
largesend 0 Enable Hardware Transmit TCP Resegmentation True
netaddr 0 Address to ping True
pvid 1 PVID to use for the SEA device True
pvid_adapter ent2 Default virtual adapter to use for non-VLAN-tagged packets True
qos_mode loose N/A True
real_adapter ent0 Physical adapter associated with the SEA True
thread 1 Thread mode enabled (1) or disabled (0) True
virt_adapters ent2 List of virtual adapters associated with the SEA (comma separated) True

```

Then priority can be set for existing Vlan device via **smitty vlan** and selecting the desired vlan device. You will see a screen like Example 3-5 on page 139 where you can set VLAN priority level.

*Example 3-5 Configuring VLAN for existing vlan device*

Change / Show Characteristics of a VLAN

Type or select values in entry fields.  
Press Enter AFTER making all desired changes.

	[Entry Fields]	
VLAN Name	ent1	
VLAN Base Adapter	[ent0]	+
VLAN Tag ID	[20]	+#
VLAN Priority	[0]	+#

---

F1=Help	F2=Refresh	F3=Cancel	F4=List
F5=Reset	F6=Command	F7=Edit	F8=Image
F9=Shell	F10=Exit	Enter=Do	

---

### 3.5.4 Best practices in setting Mode for QoS

#### When to use strict mode:

- ▶ Maintaining priority is more important than preventing starvation.
- ▶ Strict mode should be used when the network administrator has a thorough understanding of the network traffic.
- ▶ The network administrator understands the possibility of overhead and bandwidth starvation; and knows how to prevent this from occurring.

#### When to use loose mode:

- ▶ Preventing starvation is more important than maintaining priority.

## 3.6 DoS Hardening

VIOS/AIX was vulnerable to Denial of Service (DoS) attacks like other operating systems. In large corporations network security is paramount and it was unacceptable to have servers bogged down by DoS attacks. A DoS attack targets a machine and makes it unavailable. The target machine is bombarded with fake network communication requests for a service (like ftp, telnet, etc) which causes it to allocate resources for each of the request. For every request a port is held busy and process resources are allocated. Eventually target machine is exhausted of all its resources and becomes unresponsive.

**Tip:** User can set DoS Hardening rules on default ports using **viosecure -level high** command too. See section 4.1.4 for more details.



### 3.6.1 Solution

One probable solution adopted from z/OS is to limit the total number of active connections an application has at a time. This will put a restriction on the number of address spaces created by forking applications such as ftpd, telnetd etc. A fair share algorithm is also provided based on the percentage of remaining available connections already held by a source IP address. Fair share algorithm will enforce TCP traffic regulations policies.

In order to utilize Network traffic regulation you need to enable it first. Example 3-6 on page 141 shows how to do that.

*Example 3-6 Enabling network traffic regulation*

---

```
# no -p -p tcptr_enable=1

# no -a |grep tcptr
      tcptr_enable = 1
```

---

**tcptr** can be used to display current policy for various services and modify it. For the Virtual I/O Server you need to execute it from the root shell. Syntax for it is:

```
tcptr -add <start_port> <end_port> <max> <div>
```

```
tcptr -delete <start_port> <end_port>
```

```
tcptr -show
```

Where,

<start\_port> is the starting TCP port for this policy.

<end\_port> is the ending TCP port for this policy.

<max> is maximum pool of connections for this policy.

<div> is divisor (<32) governing available pool.

Example 3-7 on page 141 shows how to regulate network traffic for port 25 (sendmail service).

*Example 3-7 Using tcptr for Network traffic regulation for sendmail service*

---

```
# tcptr -show
policy: Error failed to allocate memory
```

```
(1) root @ core13: 6.1.2.0 (0841A_61D) : /
```

```
# tcptr -add 25 25 1000
StartPort=25    EndPort=25    MaxPool=1000    Div=0

(0) root @ core13: 6.1.2.0 (0841A_61D) : /
# tcptr -show
TCP Traffic Regulation Policies:
StartPort=25    EndPort=25    MaxPool=1000    Div=0    Used=0
```

---



# Virtual I/O Server security

This chapter describes how to harden Virtual I/O Server security using the **viosecure** command provided since version 1.3. We discuss the following topics here:

- ▶ Network security and firewall
- ▶ LDAP client on Virtual I/O Server
- ▶ Kerberos client on Virtual I/O Server

## 4.1 Network security

If your Virtual I/O Server has an IP address assigned after installation, some network services are running and open by default. The services in the listening open state are listed in Table 4-1.

Table 4-1 Default open ports on Virtual I/O Server

Port number	Service	Purpose
21	FTP	Unencrypted file transfer
22	SSH	Secure shell and file transfer
23	Telnet	Unencrypted remote login
111	rpcbind	NFS connection
657	RMC	RMC connections (used for dynamic LPAR operations)

In most cases the secure shell (SSH) service for remote login and the secure copy (SCP) for copying files should be sufficient for login and file transfer. Telnet and FTP are not using encrypted communication and should be disabled. Port 657 for RMC has to be left open if you consider using any dynamic LPAR operations. This port is used for the communication between the logical partition and the Hardware Management Console.

### 4.1.1 Stopping network services

In order to stop Telnet and FTP and prevent them from starting automatically after reboot, the **stopnetsvc** command can be used as shown in Example 4-1.

Example 4-1 Stopping network services

---

```
$ stopnetsvc telnet
0513-127 The telnet subserver was stopped successfully.
$ stopnetsvc ftp
0513-127 The ftp subserver was stopped successfully.
```

---

### 4.1.2 Setting up the firewall

A common approach to designing a firewall or IP filter is to determine ports that are necessary for operation, to determine sources from which those ports will be accessed, and to close everything else.

Firewall rules are parsed in the reverse order they are set. The first matching latest set rule is used. In order to keep it simple, we will deny everything first and then allow the desired traffic from certain addresses. In most cases there is no need to use explicit deny rules on the port or source address level. A combination of *deny all* and allow rules is sufficient in most scenarios.

An *Allow rule* will enable a restrictions.

A *Deny rule* will remove a restriction.

By default Firewall is set to *deny all*, which means there are no restrictions.

More *Allow rule* indicates more restrictions.

**Note:** Configure the firewall using the virtual terminal connection, not the network connection, in order to avoid lockout. This can happen if you restrict (add an *Allow rule* for) the protocol through which you are connected to the machine.

Assume we have some hosts on our network, as listed in Table 4-2.

Table 4-2 Hosts on the network

Host	IP Address	Comment
VIO Server	9.3.5.197	
Hardware Management Console	9.3.5.128	For dynamic LPAR and for monitoring RMC communication should be allowed to VIOS.
NIM Server, Management server	9.3.5.200	For administration SSH communication should be allowed to VIOS.
Administrators workstation	9.3.4.148	SSH communication can be allowed from the administrator's workstation, but better use a "jump" to the management server.

Therefore, our firewall would consist of the following rules:

1. Allow RMC from the Hardware Management console.
2. Allow SSH from NIM and/or the administrator's workstation.

### 3. Deny anything else.

To deploy this scenario we will issue the command in reverse order in order to get the rules inserted in the right order.

Example 4-2 shows default firewall settings, which just means there is no firewall active and no rules set.

*Example 4-2 Firewall view*

---

```
$ viosecure -firewall view
Firewall      OFF
```

Interface	Local		Remote		ALLOWED	PORTS	Expiration
	Port	Port	Port	Port	Service	IPAddress	
Time(seconds)	-----	-----	-----	-----	-----	-----	-----

---

**Tip:** Turn the firewall off, set up the rules, then turn it back on again—otherwise you can be locked out by any accidental change.

#### 1. Turn off the firewall first in order not to accidentally lock yourself out:

```
$ viosecure -firewall off
```

#### 2. Set the “deny all” rule first:

```
viosecure -firewall deny -port 0
```

#### 3. Now put your allow rules; they are going to be inserted above the “deny all” rule and be matched first. Change the IP addresses used to match your network.

```
$ viosecure -firewall allow -port 22 -address 9.3.5.200
$ viosecure -firewall allow -port 22 -address 9.3.4.148
$ viosecure -firewall allow -port 657 -address 9.3.5.128
```

#### 4. Check your rules. Your output should look like Example 4-3.

*Example 4-3 Firewall rules output*

---

```
$ viosecure -firewall view
Firewall      OFF
```

Interface	Local		Remote		ALLOWED	PORTS
	Port	Port	Port	Port	Service	IPAddress
Time(seconds)	-----	-----	-----	-----	-----	-----

Interface Time(seconds)	Port	Port	Service	IPAddress	Expiration
all	657	any	rmc	9.3.5.128	0
all	22	any	ssh	9.3.4.148	0
all	22	any	ssh	9.3.5.200	0

5. To disable any kind of connection irrespective of service from host 9.3.5.180 you need to execute:

```
$ viosecure -firewall allow -port 0 -address 9.3.5.180
```

6. Turn your firewall on and test connections:

```
viosecure -firewall on
```

Our rule set allows the desired network traffic only and blocks any other requests. The rules set with the **viosecure** command only apply to inbound traffic. However, this setup will also block any ICMP requests, making it impossible to ping the Virtual I/O Server or get any ping responses. This may be an issue if you are using the **ping** command to determine Shared Ethernet Adapter (SEA) failover or for EtherChannel.

**Note:** Setting up a firewall is also available in the configuration menu accessed by the **cfgassist** command.

### 4.1.3 Enabling ping through the firewall

As described in 4.1.2, “Setting up the firewall” on page 144 our sample firewall setup also blocks all incoming ICMP requests. If you need to enable ICMP for a Shared Ethernet Adapter configuration or Monitoring, use the **oem\_setup\_env** command and root access to define ICMP rules.

We can create additional ICMP rules that will allow pings by using two commands:

```
/usr/sbin/genfilt -v 4 -a P -s 0.0.0.0 -m 0.0.0.0 -d 0.0.0.0 -M
0.0.0.0 -g n -c icmp -o eq -p 0 -0 any -P 0 -r L -w I -l N -t 0 -i
all -D echo_reply
```

and:

```
/usr/sbin/genfilt -v 4 -a P -s 0.0.0.0 -m 0.0.0.0 -d 0.0.0.0 -M
0.0.0.0 -g n -c icmp -o eq -p 8 -0 any -P 0 -r L -w I -l N -t 0 -i
all -D echo_request
```

## 4.1.4 Security Hardening Rules

**viosecure** command can be used to configure Security hardening rules too. User can enforce either the preconfigured security levels or choose customize them based on her requirements. Currently preconfigured rules are high, medium and low. Each rule will have number of security policies which can be enforced as shown in example below:

```
$ viosecure -level low -apply
Processedrules=44      Passedrules=42  Failedrules=2
Level=AllRules
      Input file=/home/ios/security/viosecure.xml
```

Alternatively user can choose the policies she wants as shown in example 4.4 (the command has been truncated because of its length)

### Example 4-4 High level Firewall settings

```
$ viosecure -level high
AIX: "/usr/sbin/aixpert -l high -n -o /home/ios/security/viosecure.out.ab"

1. hls_tcptr:TCP Traffic Regulation High - Enforces denial-of-service mitigation on popular ports.
2. hls_rootpwdintchk:Root Password Integrity Check: Makes sure that the root password being set is not weak
3. hls_sedconfig:Enable SED feature: Enable Stack Execution Disable feature
4. hls_removeguest:Remove guest account: Removes guest account and its files
5. hls_chetcftusers:Add root user in /etc/ftpusers file: Adds root username in /etc/ftpusers file
6. hls_xhost:Disable X-Server access: Disable access control for X-Server
7. hls_rmdotfrmpathroot:Remove dot from non-root path: Removes dot from PATH environment variable from files .profile,
.kshrc, .cshrc and .login in user's home directory
8. hls_rmdotfrmpathroot:Remove dot from path root: Remove dot from PATH environment variable from files .profile, .kshrc,
.cshrc and .login in root's home directory
9. hls_loginherald:Set login herald: Set login herald in default stanza
10. hls_crontabperm:Crontab permissions: Ensures root's crontab jobs are owned and writable only by root

? 1,2

11. hls_limitsysacc:Limit system access: Makes root the only user in cron.allow file and removes the cron.deny file
12. hls_core:Set core file size: Specifies the core file size to 0 for root
13. hls_umask:Object creation permissions: Specifies default object creation permissions to 077
14. hls_ipsecshunports:Guard host against port scans: Shuns vulnerable ports for 5 minutes to guard the host against port
scans
15. hls_ipsecshunhost:Shun host for 5 minutes: Shuns the hosts for 5 minutes, which tries to access un-used ports
16. hls_sockthresh:Network option sockthresh: Set network option sockthresh's value to 60
17. hls_tcp_tcpsecure:Network option tcp_tcpsecure: Set network option tcp_tcpsecure's value to 7
18. hls_sb_max:Network option sb_max: Set network option sb_max's value to 1MB
19. hls_tcp_mssdflt:Network option tcp_mssdflt: Set network option tcp_mssdflt's value to 1448
20. hls_rfc1323:Network option rfc1323: Set network option rfc1323's value to 1

? ALL

21. hls_tcp_recvspace:Network option tcp_recvspace: Set network option tcp_recvspace's value to 262144
22. hls_tcp_sendspace:Network option tcp_sendspace: Set network option tcp_sendspace's value to 262144
23. hls_udp_pmtu_discover:Network option udp_pmtu_discover: Set network option udp_pmtu_discover's value to 0
24. hls_tcp_pmtu_discover:Network option tcp_pmtu_discover: Set network option tcp_pmtu_discover's value to 0
25. hls_nonlocsrcroute:Network option nonlocsrcroute: Set network option nonlocsrcroute's value to 0
26. hls_ip6srcrouteforward:Network option ip6srcrouteforward: Set network option ip6srcrouteforward's value to 0
27. hls_ipsrcroutesend:Network option ipsrcroutesend: Set network option ipsrcroutesend's value to 0
28. hls_ipsrcrouterrecv:Network option ipsrcrouterrecv: Set network option ipsrcrouterrecv's value to 0
29. hls_ipsrcrouteforward:Network option ipsrcrouteforward: Set network option ipsrcrouteforward's value to 0
30. hls_ipsendredirects:Network option ipsendredirects: Set network option ipsendredirects's value to 0
```



⋮  
⋮

---

In order to view current security rules use command **viosecure -view**.

To undo all security policies use **viosecure -undo** command.

### 4.1.5 DoS Hardening

To overcome Denial of Service attacks new feature has been implemented in AIX. For more information see “DoS Hardening” on page 140.

## 4.2 The Virtual I/O Server as an LDAP client

The Lightweight Directory Access Protocol defines a standard method for accessing and updating information about a directory (a database) either locally or remotely in a client-server model. The LDAP method is used by a cluster of hosts to allow centralized security authentication as well as access to user and group information.

Virtual I/O Server Version 1.4 introduced LDAP authentication for the Virtual I/O Server’s users and with Version 1.5 of Virtual I/O Server a secure LDAP authentication is now supported, using a secure sockets layer (SSL).

The steps necessary to create an SSL certificate, set up a server and then configure the Virtual I/O Server as a client are described in the following sections.

### 4.2.1 Creating a key database file

All the steps described here suppose that an IBM Tivoli Directory Server is installed on one server in the environment as well as the GSKit file sets. More information about the IBM Tivoli Directory Server can be found at:

<http://www-306.ibm.com/software/tivoli/resource-center/security/code-directory-server.jsp>

To create the key database file and certificate (self-signed for simplicity in this example), follow these steps:

1. Ensure that the GSKit and gsk7ikm are installed on the LDAP server, as follows:

```
# lsipp -l |grep gsk
gskjs.rte          7.0.3.30 COMMITTED AIX Certificate and SSL Java
gksa.rte           7.0.3.30 COMMITTED AIX Certificate and SSL Base
gskta.rte          7.0.3.30 COMMITTED AIX Certificate and SSL Base
```

2. Start the gsk7ikm utility, which is located on /usr/bin/gsk7ikm, which is a symbolic link to /usr/opt/ibm/gskta/bin/gsk7ikm. A window like the one shown in Figure 4-1 will appear.

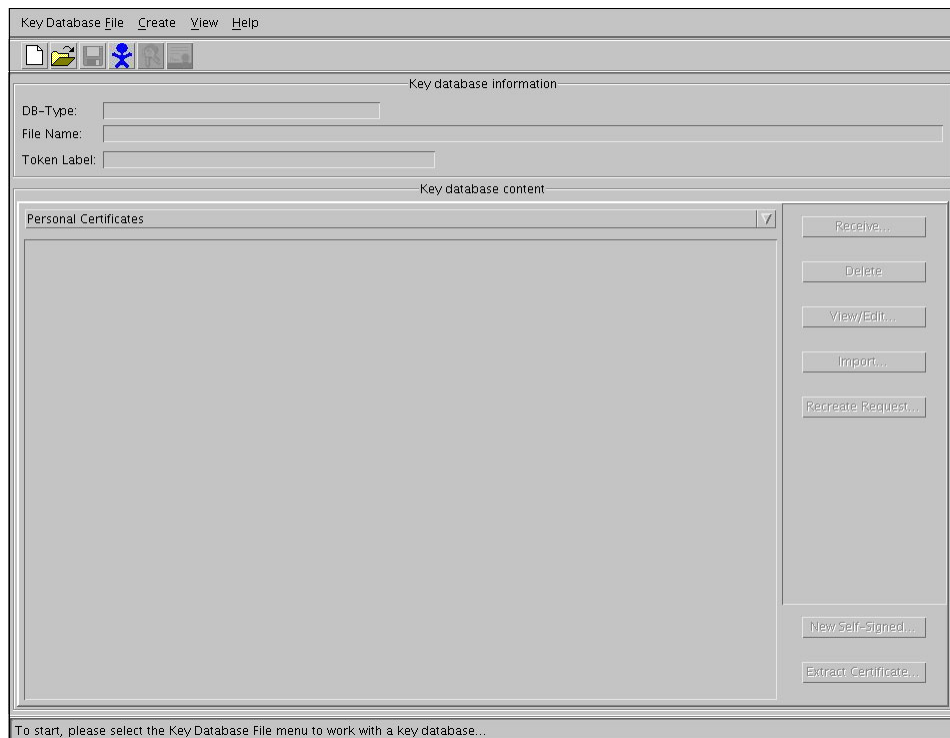


Figure 4-1 ikeyman program initial window

3. Select **Key Database File** and then **New**. A window similar to the one in Figure 4-2 will appear.

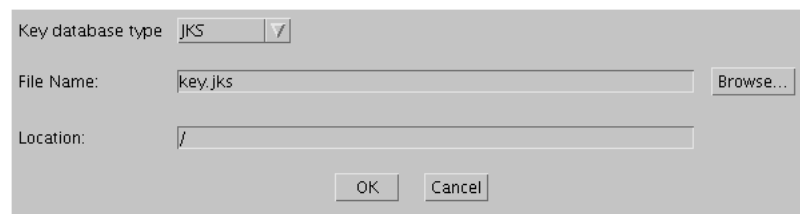


Figure 4-2 Create new key database screen

4. On the same screen change the Key database type to CMS, the File Name to ldap\_server.kdb (on this example) and then set the Location to one directory where the keys can be stored, /etc/ldap in this example. The final screen will be similar to Figure 4-3.

Key database type: CMS

File Name: ldap\_server.kdb [Browse...]

Location: /etc/ldap

OK Cancel

Figure 4-3 Creating the ldap\_server key

5. Click **OK**.
6. A new screen will appear. Enter the key database file password, and confirm it. Remember this password because it is required when the database file is edited. In this example the key database password was set to passw0rd.
7. Accept the default expiration time.
8. If you want the password to be masked and stored into a stash file, select **Stash the password to a file**.

A stash file can be used by some applications so that the application does not have to know the password to use the key database file. The stash file has the same location and name as the key database file and has an extension of \*.sth.

The screen should be similar to the one in Figure 4-4.

Password: \*\*\*\*\*

Confirm Password: \*\*\*\*\*

Set expiration time? 60 Days

Stash the password to a file?

Password Strength:

OK Reset Cancel

Figure 4-4 Setting the key database password

## 9. Click **OK**.

This completes the creation of the key database file. There is a set of default signer certificates. These are the default certificate authorities that are recognized. This is shown in Figure 4-5.

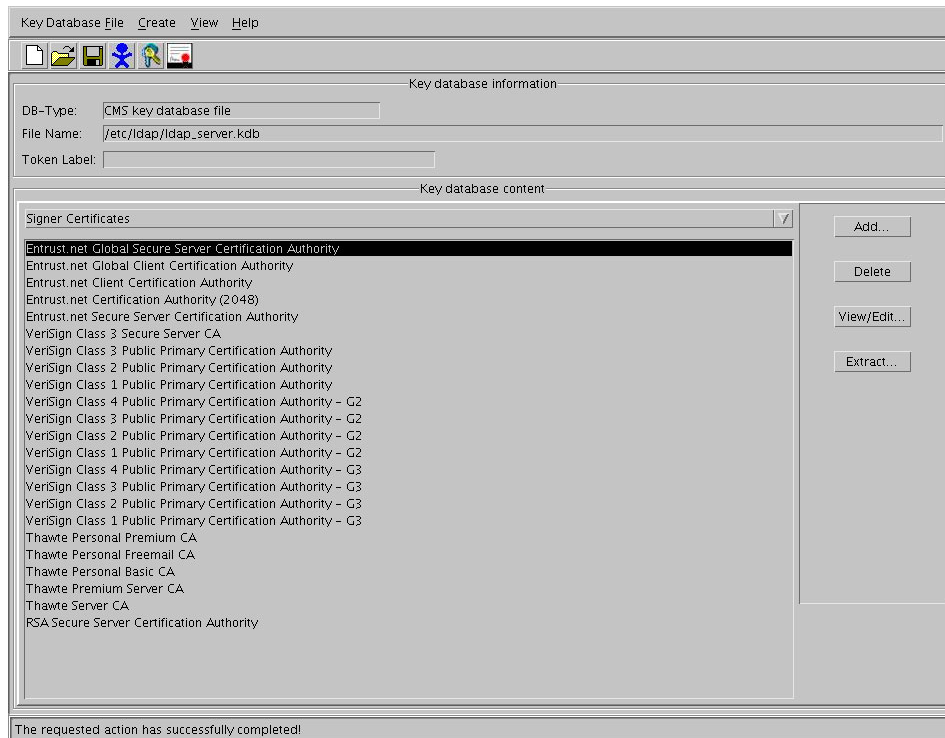
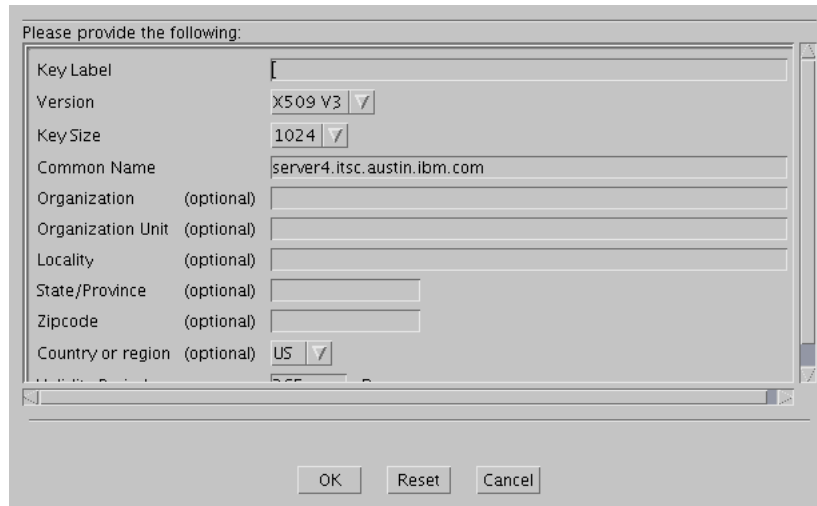


Figure 4-5 Default certificate authorities available on the *keyman* program

10. At this time, the key could be exported and sent to a certificate authority to be validated and then used. In this example, for simplicity reasons, the key is signed using a self-signed certificate. To create a self-signed certificate, select **Create** and then **New Self-Signed Certificate**. A window similar to the one in Figure 4-6 will appear.



Please provide the following:

Key Label	<input type="text"/>
Version	X509 V3 ▾
Key Size	1024 ▾
Common Name	server4.itsc.austin.ibm.com
Organization (optional)	<input type="text"/>
Organization Unit (optional)	<input type="text"/>
Locality (optional)	<input type="text"/>
State/Province (optional)	<input type="text"/>
Zipcode (optional)	<input type="text"/>
Country or region (optional)	US ▾

OK Reset Cancel

Figure 4-6 Creating a self-signed certificate initial screen

11. Type a name in the Key Label field that GSKit can use to identify this new certificate in the key database. In this example the key is labeled `ldap_server`.
12. Accept the defaults for the Version field (X509V3) and for the Key Size field.
13. Enter a company name in the Organization field.

14. Complete any optional fields or leave them blank: the default for the Country field and 365 for the Validity Period field. The window should look like the one in Figure 4-7.

Please provide the following:

Key Label	ldap_server
Version	X509 V3
Key Size	1024
Common Name	ldap_server.itsc.austin.ibm.com
Organization (optional)	IBM
Organization Unit (optional)	ITSO
Locality (optional)	Austin
State/Province (optional)	Texas
Zipcode (optional)	
Country or region (optional)	US

OK Reset Cancel

Figure 4-7 Self-signed certificate information

15. Click **OK**. GSKit generates a new public and private key pair and creates the certificate.

This completes the creation of the LDAP client's personal certificate. It is displayed in the Personal Certificates section of the key database file.

Next, the LDAP Server's certificate must be extracted to a Base64-encoded ASCII data file.

16. Highlight the self-signed certificate that was just created.
17. Click **Extract Certificate**.
18. Click **Base64-encoded ASCII data** as the type.
19. Type a certificate file name for the newly extracted certificate. The certificate file's extension is usually \*.arm.
20. Type the location where you want to store the extracted certificate and then click **OK**.
21. Copy this extracted certificate to the LDAP server system.

This file will only be used if the key database is going to be used as an SSL in a Web Server. This can happen when the LDAP administrator decides to manage the LDAP through its Web interface. Then this \*.arm file can be transferred to your PC and imported to the Web browser.

You can find more about the GSKit at:

[http://publib.boulder.ibm.com/infocenter/tivihelp/v2r1/index.jsp?topic=/com.ibm.itame.doc\\_5.1/am51\\_webinstall223.htm](http://publib.boulder.ibm.com/infocenter/tivihelp/v2r1/index.jsp?topic=/com.ibm.itame.doc_5.1/am51_webinstall223.htm)

## 4.2.2 Configuring the LDAP server

Since the key database was generated, it can now be used to configure the LDAP server.

The **mksecldap** command is used to set up an AIX system as an LDAP server or client for security authentication and data management.

A description of how to set up the AIX system as an LDAP server is provided in this section. Remember that all file sets of the IBM Tivoli directory Server 6.1 have to be installed before configuring the system as an LDAP server. When installing the LDAP server file set, the LDAP client file set and the backend DB2® software are automatically installed as well. No DB2 preconfiguration is required to run this command for the LDAP server setup. When the **mksecldap** command is run to set up a server, the command does the following:

1. Creates the DB2 instance with `ldapdb2` as the default instance name.
2. Because in this case the IBM Directory Server 6.1 is being configured, an LDAP server instance with the default name of `ldapdb2` is created. A prompt is displayed for the encryption seed to create the key files. The input encryption seed must be at least 12 characters.
3. Creates a DB2 database with `ldapdb2` as the default database name.
4. Creates the base DN (`o=ibm` in this example). The directory information tree that will be created in this example by default is shown in Figure 4-8.

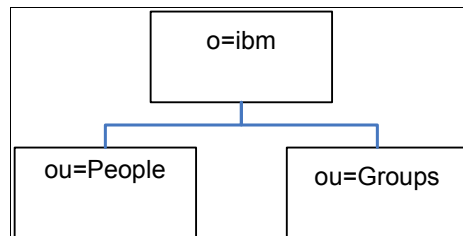


Figure 4-8 Default directory information tree created by the **mksecldap** command

1. Because the `-u NONE` flag was not specified, the data from the security database from the local host is exported into the LDAP database. Because the `-S` option was used and followed by `rfc2307aix`, the **mksecldap** command exports users or groups using this schema.



2. The LDAP administrator DN is set to cn=admin and the password is set to passw0rd.
3. Because the -k flag was used, the server will use SSL (secure socket layer).
4. The plugin libldapaudit.a is installed. This plugin supports an AIX audit of the LDAP server.
5. The LDAP server is started after all the above steps are completed.
6. The LDAP process is added to /etc/inittab to have the LDAP server start after a reboot.

The command and its output are shown here:

```
# mksecldap -s -a cn=admin -p passw0rd -S rfc2307aix -d o=ibm -k
/etc/ldap/ldap_server.kdb -w passw0rd
ldapdb2's New password:
Enter the new password again:
Enter an encryption seed to generate key stash files:
You have chosen to perform the following actions:

GLPICR020I A new directory server instance 'ldapdb2' will be created.
GLPICR057I The directory server instance will be created at: '/home/ldapdb2'.
GLPICR013I The directory server instance's port will be set to '389'.
GLPICR014I The directory server instance's secure port will be set to '636'.
GLPICR015I The directory instance's administration server port will be set to '3538'.
GLPICR016I The directory instance's administration server secure port will be set to
'3539'.
GLPICR019I The description will be set to: 'IBM Tivoli Directory Server Instance V6.1'.
GLPICR021I Database instance 'ldapdb2' will be configured.
GLPICR028I Creating directory server instance: 'ldapdb2'.
GLPICR025I Registering directory server instance: 'ldapdb2'.
GLPICR026I Registered directory server instance: : 'ldapdb2'.
GLPICR049I Creating directories for directory server instance: 'ldapdb2'.
GLPICR050I Created directories for directory server instance: 'ldapdb2'.
GLPICR043I Creating key stash files for directory server instance: 'ldapdb2'.
GLPICR044I Created key stash files for directory server instance: 'ldapdb2'.
GLPICR040I Creating configuration file for directory server instance: 'ldapdb2'.
GLPICR041I Created configuration file for directory server instance: 'ldapdb2'.
GLPICR034I Creating schema files for directory server instance: 'ldapdb2'.
GLPICR035I Created schema files for directory server instance: 'ldapdb2'.
GLPICR037I Creating log files for directory server instance: 'ldapdb2'.
GLPICR038I Created log files for directory server instance: 'ldapdb2'.
GLPICR088I Configuring log files for directory server instance: 'ldapdb2'.
GLPICR089I Configured log files for directory server instance: 'ldapdb2'.
GLPICR085I Configuring schema files for directory server instance: 'ldapdb2'.
GLPICR086I Configured schema files for directory server instance: 'ldapdb2'.
GLPICR073I Configuring ports and IP addresses for directory server instance: 'ldapdb2'.
GLPICR074I Configured ports and IP addresses for directory server instance: 'ldapdb2'.
GLPICR077I Configuring key stash files for directory server instance: 'ldapdb2'.
GLPICR078I Configured key stash files for directory server instance: 'ldapdb2'.
GLPICR046I Creating profile scripts for directory server instance: 'ldapdb2'.
GLPICR047I Created profile scripts for directory server instance: 'ldapdb2'.
```

GLPICR069I Adding entry to /etc/inittab for the administration server for directory instance: 'ldapdb2'.  
GLPICR070I Added entry to /etc/inittab for the administration server for directory instance: 'ldapdb2'.  
GLPICR118I Creating runtime executable for directory server instance: 'ldapdb2'.  
GLPICR119I Created runtime executable for directory server instance: 'ldapdb2'.  
GLPCTL074I Starting admin daemon instance: 'ldapdb2'.  
GLPCTL075I Started admin daemon instance: 'ldapdb2'.  
GLPICR029I Created directory server instance: : 'ldapdb2'.  
GLPICR031I Adding database instance 'ldapdb2' to directory server instance: 'ldapdb2'.  
GLPCTL002I Creating database instance: 'ldapdb2'.  
GLPCTL003I Created database instance: 'ldapdb2'.  
GLPCTL017I Cataloging database instance node: 'ldapdb2'.  
GLPCTL018I Cataloged database instance node: 'ldapdb2'.  
GLPCTL008I Starting database manager for database instance: 'ldapdb2'.  
GLPCTL009I Started database manager for database instance: 'ldapdb2'.  
GLPCTL049I Adding TCP/IP services to database instance: 'ldapdb2'.  
GLPCTL050I Added TCP/IP services to database instance: 'ldapdb2'.  
GLPICR081I Configuring database instance 'ldapdb2' for directory server instance: 'ldapdb2'.  
GLPICR082I Configured database instance 'ldapdb2' for directory server instance: 'ldapdb2'.  
GLPICR052I Creating DB2 instance link for directory server instance: 'ldapdb2'.  
GLPICR053I Created DB2 instance link for directory server instance: 'ldapdb2'.  
GLPICR032I Added database instance 'ldapdb2' to directory server instance: 'ldapdb2'.  
You have chosen to perform the following actions:

GLPDPW004I The directory server administrator DN will be set.  
GLPDPW005I The directory server administrator password will be set.  
GLPDPW009I Setting the directory server administrator DN.  
GLPDPW010I Directory server administrator DN was set.  
GLPDPW006I Setting the directory server administrator password.  
GLPDPW007I Directory server administrator password was set.  
You have chosen to perform the following actions:

GLPCDB023I Database 'ldapdb2' will be configured.  
GLPCDB024I Database 'ldapdb2' will be created at '/home/ldapdb2'  
GLPCDB035I Adding database 'ldapdb2' to directory server instance: 'ldapdb2'.  
GLPCTL017I Cataloging database instance node: 'ldapdb2'.  
GLPCTL018I Cataloged database instance node: 'ldapdb2'.  
GLPCTL008I Starting database manager for database instance: 'ldapdb2'.  
GLPCTL009I Started database manager for database instance: 'ldapdb2'.  
GLPCTL026I Creating database: 'ldapdb2'.  
GLPCTL027I Created database: 'ldapdb2'.  
GLPCTL034I Updating the database: 'ldapdb2'  
GLPCTL035I Updated the database: 'ldapdb2'  
GLPCTL020I Updating the database manager: 'ldapdb2'.  
GLPCTL021I Updated the database manager: 'ldapdb2'.  
GLPCTL023I Enabling multi-page file allocation: 'ldapdb2'  
GLPCTL024I Enabled multi-page file allocation: 'ldapdb2'  
GLPCDB005I Configuring database 'ldapdb2' for directory server instance: 'ldapdb2'.  
GLPCDB006I Configured database 'ldapdb2' for directory server instance: 'ldapdb2'.  
GLPCTL037I Adding local loopback to database: 'ldapdb2'.  
GLPCTL038I Added local loopback to database: 'ldapdb2'.

GLPCTL011I Stopping database manager for the database instance: 'ldapdb2'.  
GLPCTL012I Stopped database manager for the database instance: 'ldapdb2'.  
GLPCTL008I Starting database manager for database instance: 'ldapdb2'.  
GLPCTL009I Started database manager for database instance: 'ldapdb2'.  
GLPCDB003I Added database 'ldapdb2' to directory server instance: 'ldapdb2'.  
You have chosen to perform the following actions:

GLPCSF007I Suffix 'o=ibm' will be added to the configuration file of the directory server instance 'ldapdb2'.  
GLPCSF004I Adding suffix: 'o=ibm'.  
GLPCSF005I Added suffix: 'o=ibm'.  
GLPSRV034I Server starting in configuration only mode.  
GLPCOM024I The extended Operation plugin is successfully loaded from libevent.a.  
GLPSRV155I The DIGEST-MD5 SASL Bind mechanism is enabled in the configuration file.  
GLPCOM021I The preoperation plugin is successfully loaded from libDigest.a.  
GLPCOM024I The extended Operation plugin is successfully loaded from libevent.a.  
GLPCOM024I The extended Operation plugin is successfully loaded from libtranext.a.  
GLPCOM023I The postoperation plugin is successfully loaded from libpsearch.a.  
GLPCOM024I The extended Operation plugin is successfully loaded from libpsearch.a.  
GLPCOM025I The audit plugin is successfully loaded from libdapaudit.a.  
GLPCOM024I The extended Operation plugin is successfully loaded from libevent.a.  
GLPCOM023I The postoperation plugin is successfully loaded from libpsearch.a.  
GLPCOM024I The extended Operation plugin is successfully loaded from libpsearch.a.  
GLPCOM022I The database plugin is successfully loaded from libback-config.a.  
GLPCOM024I The extended Operation plugin is successfully loaded from liblog.a.  
GLPCOM024I The extended Operation plugin is successfully loaded from libidsfget.a.  
GLPSRV180I Pass-through authentication is disabled.  
GLPCOM003I Non-SSL port initialized to 389.  
Stopping the LDAP server.  
GLPSRV176I Terminated directory server instance 'ldapdb2' normally.  
GLPSRV041I Server starting.  
GLPCTL113I Largest core file size creation limit for the process (in bytes):  
'1073741312'(Soft limit) and '-1'(Hard limit).  
GLPCTL121I Maximum Data Segment(Kbytes) soft ulimit for the process was 131072 and it is modified to the prescribed minimum 262144.  
GLPCTL119I Maximum File Size(512 bytes block) soft ulimit for the process is -1 and the prescribed minimum is 2097151.  
GLPCTL122I Maximum Open Files soft ulimit for the process is 2000 and the prescribed minimum is 500.  
GLPCTL121I Maximum Physical Memory(Kbytes) soft ulimit for the process was 32768 and it is modified to the prescribed minimum 262144.  
GLPCTL121I Maximum Stack Size(Kbytes) soft ulimit for the process was 32768 and it is modified to the prescribed minimum 65536.  
GLPCTL119I Maximum Virtual Memory(Kbytes) soft ulimit for the process is -1 and the prescribed minimum is 1048576.  
GLPCOM024I The extended Operation plugin is successfully loaded from libevent.a.  
GLPCOM024I The extended Operation plugin is successfully loaded from libtranext.a.  
GLPCOM024I The extended Operation plugin is successfully loaded from libldaprepl.a.  
GLPSRV155I The DIGEST-MD5 SASL Bind mechanism is enabled in the configuration file.  
GLPCOM021I The preoperation plugin is successfully loaded from libDigest.a.  
GLPCOM024I The extended Operation plugin is successfully loaded from libevent.a.  
GLPCOM024I The extended Operation plugin is successfully loaded from libtranext.a.  
GLPCOM023I The postoperation plugin is successfully loaded from libpsearch.a.  
GLPCOM024I The extended Operation plugin is successfully loaded from libpsearch.a.

```

GLPCOM025I The audit plugin is successfully loaded from libldapaudit.a.
GLPCOM025I The audit plugin is successfully loaded from
/usr/ccs/lib/libseclldapaudit64.a(shr.o).
GLPCOM024I The extended Operation plugin is successfully loaded from libevent.a.
GLPCOM023I The postoperation plugin is successfully loaded from libpsearch.a.
GLPCOM024I The extended Operation plugin is successfully loaded from libpsearch.a.
GLPCOM022I The database plugin is successfully loaded from libback-config.a.
GLPCOM024I The extended Operation plugin is successfully loaded from libevent.a.
GLPCOM024I The extended Operation plugin is successfully loaded from libtranext.a.
GLPCOM023I The postoperation plugin is successfully loaded from libpsearch.a.
GLPCOM024I The extended Operation plugin is successfully loaded from libpsearch.a.
GLPCOM022I The database plugin is successfully loaded from libback-rdbm.a.
GLPCOM010I Replication plugin is successfully loaded from libldaprepl.a.
GLPCOM021I The preoperation plugin is successfully loaded from libpta.a.
GLPSRV017I Server configured for secure connections only.
GLPSRV015I Server configured to use 636 as the secure port.
GLPCOM024I The extended Operation plugin is successfully loaded from libloga.a.
GLPCOM024I The extended Operation plugin is successfully loaded from libidsfget.a.
GLPSRV180I Pass-through authentication is disabled.
GLPCOM004I SSL port initialized to 636.
Migrating users and groups to LDAP server.
#

```

At this point a query can be issued to the LDAP server in order to test its functionality. The `ldapsearch` command is used to retrieve information from the LDAP server and to execute an SSL search on the server that was just started. It can be used in the following way:

```

/opt/IBM/ldap/V6.1/bin/ldapsearch -D cn=admin -w passwd -h localhost -Z -K
/etc/ldap/ldap_server.kdb -p 636 -b "cn=SSL,cn=Configuration" "(ibm-slapdSslAuth=*)"
cn=SSL, cn=Configuration
cn=SSL
ibm-slapdSecurePort=636
ibm-slapdSecurity=SSLOnly
ibm-slapdSslAuth=serverauth
ibm-slapdSslCertificate=none
ibm-slapdSslCipherSpec=AES
ibm-slapdSslCipherSpec=AES-128
ibm-slapdSslCipherSpec=RC4-128-MD5
ibm-slapdSslCipherSpec=RC4-128-SHA
ibm-slapdSslCipherSpec=TripleDES-168
ibm-slapdSslCipherSpec=DES-56
ibm-slapdSslCipherSpec=RC4-40-MD5
ibm-slapdSslCipherSpec=RC2-40-MD5
ibm-slapdSslFIPSProcessingMode=false
ibm-slapdSslKeyDatabase=/etc/ldap/ldap_server.kdb
ibm-slapdSslKeyDatabasePW={AES256}31Ip2qH5pLx0IPX9NTbgvA==
ibm-slapdSslPKCS11AcceleratorMode=none
ibm-slapdSslPKCS11Enabled=false
ibm-slapdSslPKCS11Keystorage=false
ibm-slapdSslPKCS11Lib=libcknfast.so
ibm-slapdSslPKCS11TokenLabel=none
objectclass=top

```

```
objectclass=ibm-slapdConfigEntry
objectclass=ibm-slapdSSL
```

In this example the SSL configuration is retrieved from the server. Note that the database key password is stored in a cryptographic form, (`{AES256}31Ip2qH5pLx0IPX9NTbgvA==`).

Once the LDAP server has been shown to be working, the Virtual I/O Server can be configured as a client.

### 4.2.3 Configuring the Virtual I/O Server as an LDAP client

The first thing to be checked on the Virtual I/O Server before configuring it as a secure LDAP client is whether the `ldap.max_crypto_client` file sets are installed. To check this, issue the `ls1pp` command on the Virtual I/O Server as root, as follows:

```
# ls1pp -l |grep ldap
ldap.client.adt          5.2.0.0  COMMITTED  Directory Client SDK
ldap.client.rte          5.2.0.0  COMMITTED  Directory Client Runtime (No
ldap.max_crypto_client.adt
ldap.max_crypto_client.rte
ldap.client.rte          5.2.0.0  COMMITTED  Directory Client Runtime (No
```

If they are not installed, proceed with the installation before going forward with these steps. These file sets can be found on the Virtual I/O Server 1.5 Expansion Pack CD. The expansion CD comes with the Virtual I/O Server Version 1.5 CDs.

Transfer the database key from the LDAP server to the Virtual I/O Server. In this example, `ldap_server.kdb` and `ldap_server.sth` were transferred from `/etc/ldap` on the LDAP server to `/etc/ldap` on the Virtual I/O Server.

On the Virtual I/O Server, the `mkldap` command is used to configure it as an LDAP client. To configure the Virtual I/O Server as a secure LDAP client of the LDAP server that was previously configured, use the following command:

```
$ mkldap -bind cn=admin -passwd passw0rd -host NIM_server -base o=ibm -keypath
/etc/ldap/ldap_server.kdb -keypasswd passw0rd -port 636
gskjs.rte
gskjt.rte
gksa.rte
gskta.rte
```

The manual page of the `mkldap` command can be found in Appendix A.

To check whether the secure LDAP configuration is working, create an LDAP user using the `mkuser` command with the `-ldap` flag and then use the `lsuser`

command to check its characteristics, as shown in Example 4-5. Note that the registry of the user is now stored on the LDAP server.

*Example 4-5 Creating an ldap user on the Virtual I/O Server*

---

```
$ mkuser -ldap itso
itso's Old password:
itso's New password:
Enter the new password again:
$ lsuser itso
itso roles=Admin account_locked=false expires=0 histexpire=0 histsize=0
loginretries=0 maxage=0 maxexpired=-1 maxrepeats=8 minage=0 minalpha=0
mindiff=0 minlen=0 minother=0 pwdwarntime=330 registry=LDAP SYSTEM=LDAP
```

---

When the user itso tries to log in, its password has to be changed as shown in Example 4-6.

*Example 4-6 Log on to the Virtual I/O Server using an LDAP user*

---

```
login as: itso
itso@9.3.5.108's password:
[LDAP]: 3004-610 You are required to change your password.
Please choose a new one.
WARNING: Your password has expired.
You must change your password now and login again!
Changing password for "itso"
itso's Old password:
itso's New password:
Enter the new password again:
```

---

Another way to test whether the configuration is working is to use the **ldapsearch** command to do a search on the LDAP directory. In Example 4-7, this command is used to search for the characteristics of the o=ibm object.

*Example 4-7 Searching the LDAP server*

---

```
$ ldapsearch -b o=ibm -h NIM_server -D cn=admin -w passw0rd -s base -p 636 -K
/etc/ldap/ldap_server.kdb -N ldap_server -P passw0rd objectclass=*
o=ibm
objectclass=top
objectclass=organization
o=ibm
```

---

The secure LDAP connection between the LDAP server and the Virtual I/O Server is now configured and operational.

## 4.3 Network Time Protocol configuration

A synchronized time is important for error logging, Kerberos and various monitoring tools. The Virtual I/O Server has an NTP client installed. To configure it you can create or edit the configuration file `/home/padmin/config/ntp.conf` using `vi` as shown in Example 4-8:

```
$ vi /home/padmin/config/ntp.conf
```

*Example 4-8 Content of the `/home/padmin/config/ntp.conf` file*

---

```
server ptbtime1.ptb.de
server ptbtime2.ptb.de
driftfile /home/padmin/config/ntp.drift
logfile /home/padmin/config/ntp.trace
logfile /home/padmin/config/ntp.log
```

---

Once configured, you should start the `xntpd` service using the `startnetsvc` command:

```
$ startnetsvc xntpd
0513-059 The xntpd Subsystem has been started. Subsystem PID is
123092.
```

After the daemon is started, check your `ntp.log` file. If it shows messages similar to that in Example 4-9, you have to set the time manually first.

*Example 4-9 Too large time error*

---

```
$ cat config/ntp.log
5 Dec 13:52:26 xntpd[516180]: SRC stop issued.
5 Dec 13:52:26 xntpd[516180]: exiting.
5 Dec 13:56:57 xntpd[516188]: synchronized to 9.3.4.7, stratum=3
5 Dec 13:56:57 xntpd[516188]: time error 3637.530348 is way too large
(set clock manually)
```

---

In order to set the date on the Virtual I/O Server, use the `chdate` command:

```
$ chdate 1206093607
$ Thu Dec 6 09:36:16 CST 2007
```

If the synchronization is successful, your log in `/home/padmin/config/ntp.log` should look like Example 4-10.

*Example 4-10 Successful ntp synchronization*

---

```
6 Dec 09:48:55 xntpd[581870]: synchronized to 9.3.4.7, stratum=2
6 Dec 10:05:34 xntpd[581870]: time reset (step) 998.397993 s
```

```
6 Dec 10:05:34 xntpd[581870]: synchronisation lost
6 Dec 10:10:54 xntpd[581870]: synchronized to 9.3.4.7, stratum=2
```

---

In the Virtual I/O Server Version 1.5.1.1-FP-10.1 you need to restart the xntpd daemon using the **stopnetsvc xntpd** and **startnetsvc xntpd** commands after every reboot, otherwise it will use /etc/ntp.conf as a configuration file instead of /home/padmin/config/ntp.conf. As a workaround you can set up a symbolic link from /home/padmin/config/ntp.conf to /etc/ntp.conf as shown in Example 4-11.

*Example 4-11 Setting up a symbolic link for ntp.conf*

---

```
# ln -sf /home/padmin/config/ntp.conf /etc/ntp.conf
# ls -al /etc/ntp.conf
lrwxrwxrwx  1 root  staff          28 Dec 05 22:27 /etc/ntp.conf
-> /home/padmin/config/ntp.conf
```

---

## 4.4 Setting up Kerberos on the Virtual I/O Server

In order to use Kerberos on the Virtual I/O Server, you first have to install the Kerberos krb5.client.rte file set from the Virtual I/O Server expansion pack.

You then have to insert the first expansion pack CD in the DVD drive. In case the drive is mapped for the other partitions to access it, you have to unmap it on the Virtual I/O Server with the **rmvdev** command, as follows:

```
$ lsmmap -all | grep cd
Backing device      cd0
$ rmvdev -vdev cd0
vtopt0 deleted
```

You can then run the **installp** command. We use the **oem\_setup\_env** command to do this because **installp** must run with the root login.

```
$ echo "installp -agXYd /dev/cd0 krb5.client.rte" | oem_setup_env
+-----+
                        Pre-deinstall Verification...
+-----+
Verifying selections...done
```

[ output part removed for clarity purpose ]

Installation Summary

```
-----
Name                               Level      Part      Event      Result
-----
```



krb5.client.rte	1.4.0.3	USR	APPLY	SUCCESS
krb5.client.rte	1.4.0.3	ROOT	APPLY	SUCCESS

The Kerberos client file sets are now installed on the Virtual I/O Server. The login process to the operating system remains unchanged. Therefore, you must configure the system to use Kerberos as the primary means of user authentication.

To configure the Virtual I/O Server to use Kerberos as the primary means of user authentication, run the **mkkrb5clnt** command with the following parameters:

```
$ oem_setup_env
# mkkrb5clnt -c KDC -r realm -a admin -s server -d domain -A -i database -K -T
# exit
$
```

The **mkkrb5clnt** command parameters are:

- c** Sets the Kerberos Key Center (KDC) that centralizes authorizations.
- r** Sets the Kerberos realm.
- s** Sets the Kerberos admin server.
- K** Specifies Kerberos to be configured as the default authentication scheme.
- T** Specifies the flag to acquire server admin TGT based admin ticket.

For integrated login, the **-i** flag requires the name of the database being used. For LDAP, use the load module name that specifies LDAP. For local files, use the keyword files.

For example, to configure the VIO\_Server1 Virtual I/O Server to use the ITSC.AUSTIN.IBM.COM realm, the krb\_master admin and KDC server, the itsc.austin.ibm.com domain, and the local database, type the following:

```
$ oem_setup_env
# mkkrb5clnt -c krb_master.itsc.austin.ibm.com -r ITSC.AUSTIN.IBM.COM \
-s krb_master.itsc.austin.ibm.com -d itsc.austin.ibm.com -A -i files -K -T
```

```
Password for admin/admin@ITSC.AUSTIN.IBM.COM:
Configuring fully integrated login
Authenticating as principal admin/admin with existing credentials.
WARNING: no policy specified for host/VIO_Server1@ITSC.AUSTIN.IBM.COM;
defaulting to no policy. Note that policy may be overridden by
ACL restrictions.
Principal "host/VIO_Server1@ITSC.AUSTIN.IBM.COM" created.
```

```
Administration credentials NOT DESTROYED.
Making root a Kerberos administrator
Authenticating as principal admin/admin with existing credentials.
```

```
WARNING: no policy specified for root/VIO_Server1@ITSC.AUSTIN.IBM.COM;
  defaulting to no policy. Note that policy may be overridden by
  ACL restrictions.
Enter password for principal "root/VIO_Server1@ITSC.AUSTIN.IBM.COM":
Re-enter password for principal "root/VIO_Server1@ITSC.AUSTIN.IBM.COM":
Principal "root/VIO_Server1@ITSC.AUSTIN.IBM.COM" created.
```

```
Administration credentials NOT DESTROYED.
Configuring Kerberos as the default authentication scheme
Cleaning administrator credentials and exiting.
# exit
$
```

This example results in the following actions:

1. Creates the `/etc/krb5/krb5.conf` file. Values for realm name, Kerberos admin server, and domain name are set as specified on the command line. Also, this updates the paths for `default_keytab_name`, `kdc`, and `kadmin` log files.
2. The `-i` flag configures fully integrated login. The database entered is the location where AIX user identification information is stored. This is different than the Kerberos principal storage. The storage where Kerberos principals are stored is set during the Kerberos configuration.
3. The `-K` flag configures Kerberos as the default authentication scheme. This allows the users to become authenticated with Kerberos at login time.
4. The `-A` flag adds an entry in the Kerberos database to make root an admin user for Kerberos.
5. The `-T` flag acquires the server admin TGT-based admin ticket.

If a system is installed that is located in a different DNS domain than the KDC, the following additional actions must be performed:

1. Edit the `/etc/krb5/krb5.conf` file and add another entry after `[domain realm]`.
2. Map the different domain to your realm.

For example, if you want to include a client that is in the `abc.xyz.com` domain into your `MYREALM` realm, the `/etc/krb5/krb5.conf` file includes the following additional entry:

```
[domain realm]
  .abc.xyz.com = MYREALM
```

## 4.5 Managing users

When the Virtual I/O Server is installed, the only user type that is active is the prime administrator (padmin), which can create additional user IDs with the following roles:

- ▶ System administrator
- ▶ Service representative
- ▶ Development engineer

**Note:** You cannot create the prime administrator (padmin) user ID. It is automatically created and enabled after the Virtual I/O Server is installed.

Table 4-3 lists the user management tasks available on the Virtual I/O Server, as well as the commands you must run to accomplish each task.

*Table 4-3 Task and associated command to manage Virtual I/O Server users*

Task	Command
Create a system administrator user ID	<code>mkuser</code>
Create a service representative (SR) user ID	<code>mkuser</code> with the <code>-sr</code> flag
Create a development engineer (DE) user ID	<code>mkuser</code> with the <code>-de</code> flag
Create a LDAP user	<code>mkuser</code> with the <code>-ldap</code> flag
List a user's attributes	<code>lsuser</code>
Change a user's attributes	<code>chuser</code>
Switch to another user	<code>su</code>
Remove a user	<code>rmuser</code>

### 4.5.1 Creating a system administrator account

In Example 4-12 we show how to create a system administration account with the default values and then check its attributes.

*Example 4-12 Creating a system administrator user and checking its attributes*

---

```
$ mkuser johndoe
johndoe's New password:
Enter the new password again:
$ lsuser johndoe
```

```
johndoe roles=Admin account_locked=false expires=0 histexpire=0  
histsize=0 loginretries=0 maxage=0 maxexpired=-1 maxrepeats=8 minage=0  
minalpha=0 mindiff=0 minlen=0 minother=0 pldwarntime=330 registry=files  
SYSTEM=compat
```

---

The system administrator account has access to all commands except the following:

- ▶ **cleargcl**
- ▶ **lsfailedlogin**
- ▶ **lsgcl**
- ▶ **mirrorios**
- ▶ **mkuser**
- ▶ **oem\_setup\_env**
- ▶ **rmuser**
- ▶ **shutdown**
- ▶ **unmirrorios**

## 4.5.2 Creating a service representative (SR) account

In Example 4-13, we have created a service representative (SR) account. This type of account enables a service representative to run commands required to service the system without being logged in as root. This includes the following command types:

- ▶ Run diagnostics, including service aids (for example, hot plug tasks, certify, format, and so forth).
- ▶ Run all commands that can be run by a group system.
- ▶ Configure and unconfigure devices that are not busy.
- ▶ Use the service aid to update the system microcode.
- ▶ Perform the shutdown and reboot operations.

The recommended SR login user name is `qserv`.

*Example 4-13 Creating a service representative account*

---

```
$ mkuser -sr qserv  
qserv's New password:  
Enter the new password again:  
$ lsuser qserv
```

```
qserv roles=SRUser account_locked=false expires=0 histexpire=0
histsize=0 loginretries=0 maxage=0 maxexpired=-1 maxrepeats=8 minage=0
minalpha=0 mindiff=0 minlen=0 minother=0 pldwarntime=330 registry=files
SYSTEM=compat
```

---

When the server representative user logs in to the system for the first time, it is asked to change its password. After changing it, the diag menu is automatically loaded. It can then execute any task from that menu or get out of it and execute commands on the command line.

### 4.5.3 Creating a read-only account

The Virtual I/O Server **mkuser** command offers the possibility to create a read-only account. An account like that would be able to view everything a system administrator account can see but it could not change anything. Auditors are usually given an account like this. The creation of this account is accomplished by **padmin** with the following command:

```
$ mkuser -attr pgrp=view auditor
```

**Note:** A read-only account will not be able to even write on its own home directory, but it can view all configuration settings.

### 4.5.4 Checking the global command log (gcl)

Once the users and their roles are set up it is important to periodically check what they have been doing on the Virtual I/O Server. We accomplish this with the **lsgcl** command.

The **lsgcl** command lists the contents of the global command log (gcl). This log contains a listing of all commands that have been executed by all Virtual I/O Server users. Each listing contains the date and time of execution as well as the userid the command was executed from. In Example 4-14, we can see the output of this command on our Virtual I/O Server.

**Note:** The **lsgcl** command can only be executed by the prime administrator (**padmin**) user.

*Example 4-14 lsgcl command output*

---

```
Nov 16 2007, 17:12:26 padmin ioslevel
Nov 16 2007, 17:25:55 padmin updateios -accept -dev /dev/cd0
...
```

```
Nov 20 2007, 15:26:34 padmin  uname -a
Nov 20 2007, 15:29:26 qserv   diagmenu
Nov 20 2007, 16:16:11 padmin  lsfailedlogin
Nov 20 2007, 16:25:51 padmin  lsgcl
Nov 20 2007, 16:28:52 padmin  passwd johndoe
Nov 20 2007, 16:30:40 johndoe lsmapi -all
Nov 20 2007, 16:30:53 johndoe lsmapi -vadapter vhost0
Nov 20 2007, 16:32:11 padmin  lsgcl
```

---



## Virtual I/O Server maintenance

Like all other servers included in an enterprise's data recovery program, you need to back up and update the Virtual I/O Server logical partition.

This chapter first describes the following processes:

- ▶ Install or migrate to a Virtual I/O Server Version 2.1
- ▶ Understand the Virtual I/O Server backup strategy and how to coordinate the Virtual I/O Server strategy with an existing or new backup strategy.
- ▶ Restore the Virtual I/O Server logical partition
- ▶ Understand the complete update process of the Virtual I/O Server

## 5.1 Installation or migration of Virtual I/O Server Version 2.1

There are four different procedures to install or to migrate to Virtual I/O Server Version 2.1:

1. “Installation of Virtual I/O Server Version 2.1”
2. “Migration from an HMC” on page 173
3. “Migration from DVD managed by an HMC” on page 174
4. “Migration from DVD managed by IVM” on page 186

**Note:** A migration to Virtual I/O Server Version 2.1 is only supported, if you run Virtual I/O Server Version 1.3, or later. If you are running Virtual I/O Server Version 1.2 or less, you need to apply the latest Virtual I/O Server Version 1.5 fix pack before migrating.

Before you begin a migration, backup your existing Virtual I/O Server installation and then follow the steps for the installation method that you choose.

**Note:** There are two different DVDs shipped with every new Virtual I/O server 2.1 order:

- ▶ Virtual I/O Server Version 2.1 Migration DVD
- ▶ Virtual I/O Server Version 2.1 Installation DVD

For customers having a Software Maintenance Agreement (SWMA), they can order both sets of the Virtual I/O Server Version 2.1 media from the following web site:

<http://www.ibm.com/servers/eserver/ess/ProtectedServlet.wss>

In a redundant Virtual I/O Server environment, you can install or migrate one Virtual I/O Server at a time to avoid any interruption of service. The client LPAR can be up and running through the migration process of each of the Virtual I/O Servers.

**Note:** Go to the Virtual I/O Server support web site and check for any available Fix Pack:

<http://www14.software.ibm.com/webapp/set2/sas/f/vios/home.html>



The next sections describe more in detail how to install or migrate to a Virtual I/O Server Version 2.1 environment.

### 5.1.1 Installation of Virtual I/O Server Version 2.1

Before you start make sure that you have the Virtual I/O Server Version 2.1 Installation DVD available and then follow the Virtual I/O Server installation procedure described in *PowerVM Virtualization on IBM System p: Introduction and Configuration*, SG24-7940.

### 5.1.2 Migration from an HMC

Before you begin the migration from an HMC make sure that the following requirements are fulfilled:

- ▶ An HMC is attached to the system and the HMC version is at a minimum level of 7.3.4 or later and the server firmware is at the according level.
- ▶ You have the Virtual I/O Server Version 2.1 Migration DVD.
- ▶ You have hmcsuperadmin authority.
- ▶ Run the **backupios** command, and save the mksysb image to a safe location.

To start the Virtual I/O Server migration, follow the next steps:

1. Insert the Virtual I/O Server Version 2.1 Migration DVD into the DVD drive of the HMC.
2. If the HMC is communicating with the Flexible Service Processor (FSP) through a private network channel (for example, HMC (eth0) = 192.168.0.101) and the installation is over a private network, the `INSTALLIOS_PRIVATE_IF=eth0` variable will need to be exported to force the network installation over that private interface. Not doing so will prevent the client logical partition from successfully running a BOOTP from the HMC. The `INSTALLIOS_PRIVATE_IF` variable is not required for all public network installation.

To use the variable, type the following command from the HMC command line:

```
export INSTALLIOS_PRIVATE_IF=interface
```

where *interface* is the network interface through which the installation should take place.

3. To begin the migration installation, enter the following command from the HMC command line:

```
installios
```

4. Choose the server where your Virtual I/O Server partition is located.
5. Select the Virtual I/O Server partition you want to migrate.
6. Select the partition profile.
7. Enter the source of the installation images. The default installation media is /dev/cdrom.
8. Enter the IP address of the Virtual I/O Server partition.
9. Enter the subnet mask of the Virtual I/O Server partition.
10. Enter the IP address of the gateway.
11. Enter the speed for the Ethernet interface.
12. Enter the information whether it is full duplex or half duplex.

**Note:** The `installios` command defaults the network setting of 100 Mbps/full duplex for its speed and duplex setting. Please check the network switch configuration or consult the network administrator to see what is the correct speed/duplex setting in your environment.

13. Enter **no** if prompted for the client's network configured after the installation.
14. The information for all available network adapters is being retrieved. At that point the Virtual I/O Server partition reboots. Choose the correct physical Ethernet adapter.
15. Enter the appropriate language and locale.
16. Verify that your settings are correct. If so, press Enter and proceed with the installation.

After the migration is complete, the Virtual I/O Server partition is restarted to the configuration that it had before the migration installation. Run the `ioslevel` command and verify that the migration was successful. The results should indicate the similar level:

```
$ ioslevel
2.1.0.0
```

### 5.1.3 Migration from DVD managed by an HMC

Before you begin the migration from an DVD make sure that the following requirements are fulfilled:

- ▶ An HMC is attached to the system and the HMC version is at a minimum level of 7.3.4 or later and the server firmware is at the according level.

- ▶ A DVD drive is assigned to the Virtual I/O Server partition and you have the Virtual I/O Server Version 2.1 Migration DVD.
- ▶ The Virtual I/O Server is currently at version 1.3, or later.
- ▶ Run the **backupios** command, and save the mksysb image to a safe location.

**Note:** Do not use the **updateios** command to migrate the Virtual I/O Server.

To start the Virtual I/O Server migration, follow the next steps:

1. Insert the Virtual I/O Server Version 2.1 Migration DVD into the DVD drive assigned to the Virtual I/O Server partition.
2. Shutdown the Virtual I/O Server partition doing the following:
  - On a Virtual I/O Server command line run the command: **shutdown -force** and wait for the shutdown completed.
  - or
  - Check the Virtual I/O Server partition on the HMC menu **Systems Management** → **Servers** → **<name\_of\_server>**.
  - Select **Tasks** → **Operations** → **Shutdown**.
  - In the Shutdown menu, select **delayed**, click on **OK**, and wait for the shutdown completed.
3. Activate the Virtual I/O Server partition and boot it into SMS menu using **Tasks** → **Operations** → **Activate**.
4. In the appearing window select the correct profile, check *Open a terminal window or console session* and then select SMS as the Boot mode in the Advanced selection. Click on **OK**.
5. A console window opens and the partition starts the SMS main menu.
6. In the SMS menu enter 5 for option 5. *Select Boot Options* and then press Enter.
7. Enter 1 for option 1. *Select Install/Boot device* and press Enter.
8. Enter 4 for option 3. *CD/DVD* and press Enter.
9. Select 6 for option 6. *List all devices* and press Enter.
10. Select the installation drive and press Enter.
11. Enter 2 for option 2. *Normal Mode Boot* and press Enter.
12. Enter 1 for option *Yes* and press Enter.

13. The partition will now boot from the Migration DVD. Figure 5-1 shows the menu appearing after a few moments. Select the desired console and press Enter.

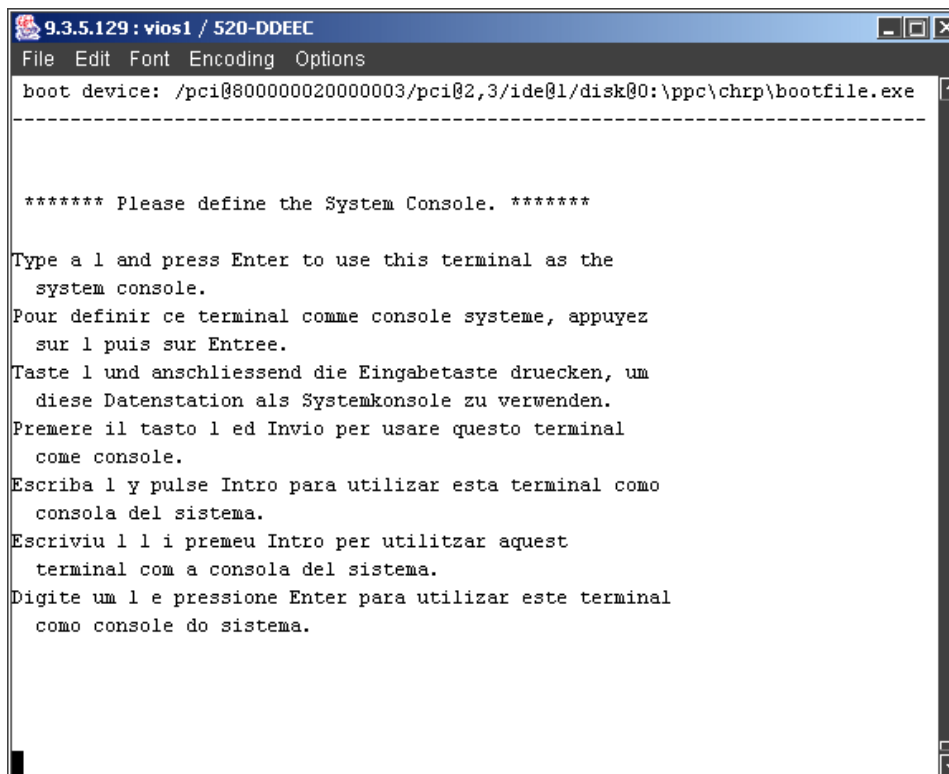


Figure 5-1 Define the System Console

14. Type 1 in the next window and press Enter to have English during install.
15. The migration proceeds and the main menu as shown in Figure 5-2 will come up:

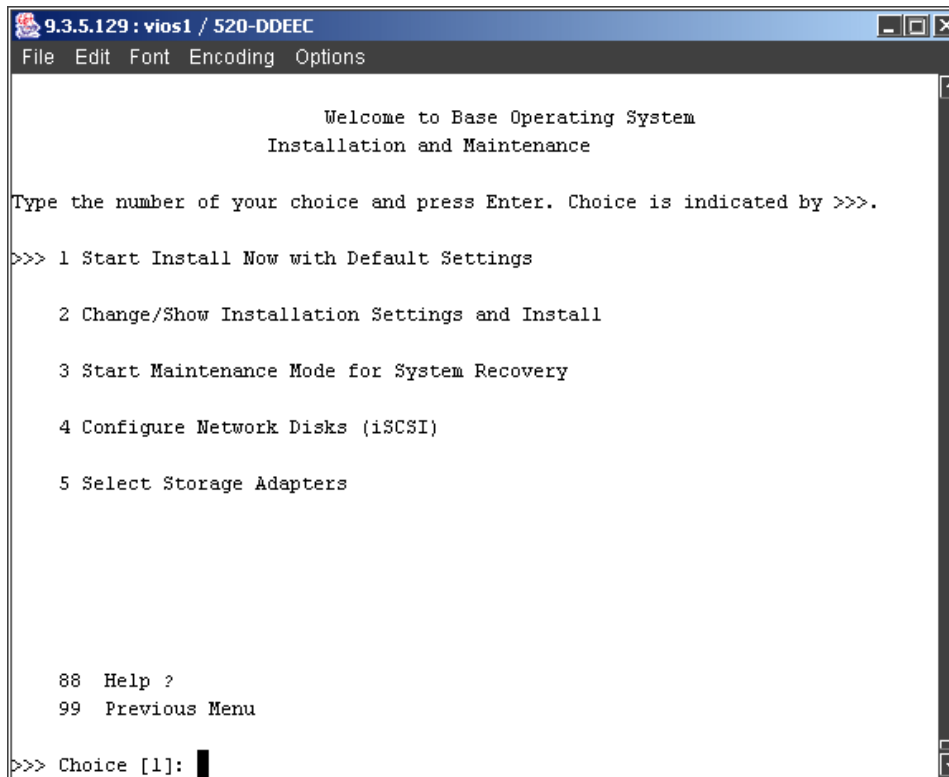


Figure 5-2 Installation and Maintenance main menu

16. Select option *1 Start Install Now with Default Settings* or verify the installation settings by choosing option *2 Change/Show Installation Settings and Install* and press Enter.
17. Figure 5-3 shows the Virtual I/O Server Installation and Settings menu:

```
9.3.5.129 : vios1 / 520-DDEEC
File Edit Font Encoding Options

          VIOS Migration Installation and Settings

Either type 0 and press Enter to install with current settings, or type the
number of the setting you want to change and press Enter.

  1 System Settings:
    Disk Where You Want to Install.....hdisk0...

>>> 0 Install with the settings listed above.

      88 Help ?          | +-----+
      99 Previous Menu | |WARNING: Base Operating System Installation will
                        | |destroy or impair recovery of SOME data on the
                        | |destination disk hdisk0.
>>> Choice [0]: 1
```

Figure 5-3 Virtual I/O Server Migration Installation and Settings

Select option 1 to verify the system settings.

18. Figure 5-4 shows the menu where you can select the disks for migration. In our example we had a mirrored Virtual I/O Server Version 1.5 environment and therefore we use option 1:

```
9.3.5.129 : vios1 / 520-DDEEC
File Edit Font Encoding Options

Change Disks Where You Want to Install

Type the number for the disks to be used for installation and press Enter.

Level   Disks In Rootvg   Location Code   Size(MB)
-----
1 5.3   hdisk0            04-08-00-3,0   34715
        hdisk1            04-08-00-4,0   34715
2 5.3   hdisk4            08-08-00-3,0   34715

88 Help ?
99 Previous Menu

>>> Choice []: 1
```

Figure 5-4 Change Disk Where You Want to Install

**Note:** Here you can see that the existing Virtual I/O Server is reported as an AIX 5.3 system. Note that other disks, not part of the Virtual I/O Server rootvg, can also have AIX 5.3 installed on. The first two disks shown in Figure 5-4 are internal SAS disks where the existing Virtual I/O Server 1.x resides on. hdisk4 is another Virtual I/O Server installation on a SAN LUN.

19. Select 0 to continue and start the migration as shown in Figure 5-5:

```
9.3.5.129 : vios1 / 520-DDEEC
File Edit Font Encoding Options

          VIOS Migration Installation and Settings

Either type 0 and press Enter to install with current settings, or type the
number of the setting you want to change and press Enter.

  1 System Settings:
    Disk Where You Want to Install....hdisk0

>>> 0 Install with the settings listed above.

      88 Help ?           | +-----+
      99 Previous Menu  | | WARNING: Base Operating System Installation will
                        | | destroy or impair recovery of SOME data on the
                        | | destination disk hdisk0.
>>> Choice [0]: █
```

Figure 5-5 Virtual I/O Server Migration Installation and Settings - start migration

20. The migration will start and then prompt you for a final confirmation as shown in Figure 5-6. At this point you can still stop the migration and boot up your existing Virtual I/O Server Version 1.x environment.



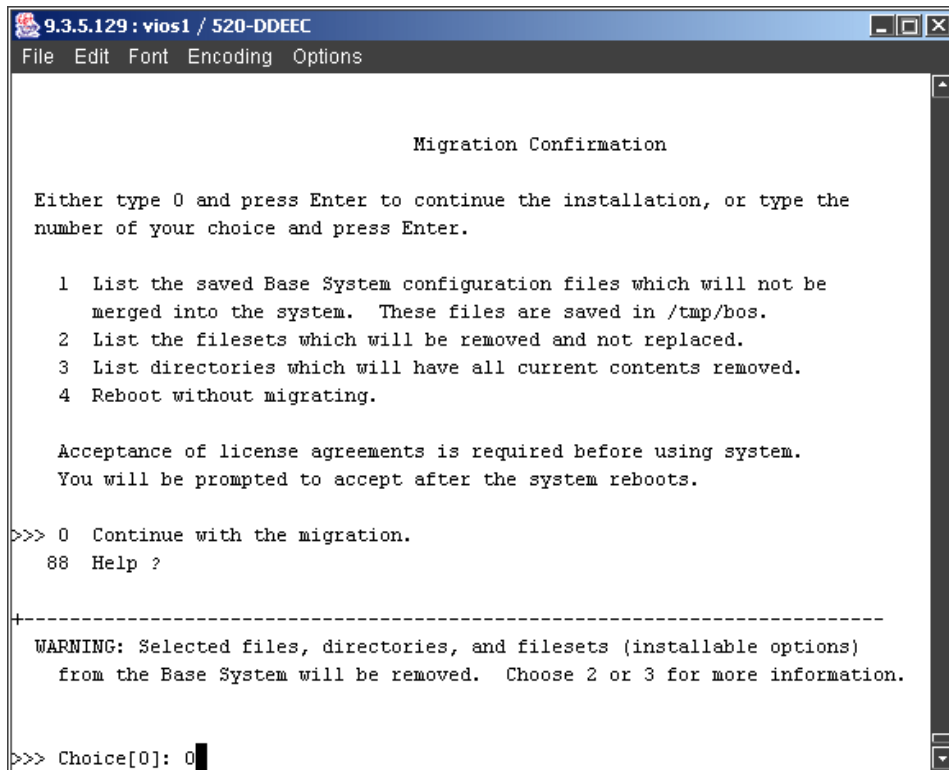


Figure 5-6 Migration Confirmation

Select 0 to continue with the migration. After a few seconds the migration will start as shown in Figure 5-7:

```
9.3.5.129: vios1 / 520-DDEEC
File Edit Font Encoding Options

installp: APPLYING software for:
      sysmgtlib.framework.core 6.1.2.0

. . . . . << Copyright notice for sysmgtlib.framework >> . . . . .
Licensed Materials - Property of IBM

5765G6200
  Copyright International Business Machines Corp. 1997, 2008.

All rights reserved.
US Government Users Restricted Rights - Use, duplication or disclosure
restricted by GSA ADP Schedule Contract with IBM Corp.
. . . . . << End of copyright notice for sysmgtlib.framework >>. . . . .

Filesets processed:  3 of 788
System Installation Time: 4 minutes      Tasks Complete: 18%

installp: APPLYING software for:
      sysmgt.websm.icons 6.1.2.0
      sysmgt.websm.apps 6.1.2.0
```

Figure 5-7 Running migration

Be patient while the migration is ongoing. This may take a while.

21. After the migration has finished you need to set the terminal type. Enter vt320 and press Enter as shown in Figure 5-8:

```
9.3.5.129 : vios1 / 520-DDEEC
File Edit Font Encoding Options

                          Set Terminal Type

The terminal is not properly initialized. Please enter a terminal type
and press Enter. Some terminal types are not supported in
non-English languages.

      ibm3101          tv1912          vt330          aixterm
      ibm3151          tv1920          vt340          dtterm
      ibm3161          tv1925          wyse30         xterm
      ibm3162          tv1950          wyse50         lft
      ibm3163          vs100          wyse60         sun
      ibm3164          vt100          wyse100
      ibmpc            vt320          wyse350

      +-----Messages-----
      | If the next screen is unreadable, press Break (Ctrl-c)
      | to return to this screen.
      |

88 Help ?

>>> Choice []: vt320
```

Figure 5-8 Set Terminal Type

22. Finally you need to accept the license agreements as shown in Figure 5-9:



Figure 5-9 Software License Agreements

Press Enter and then change *ACCEPT Installed License Agreements?* to YES using the Tab key in the Accept License Agreements menu, as shown in Figure 5-10, and press Enter:

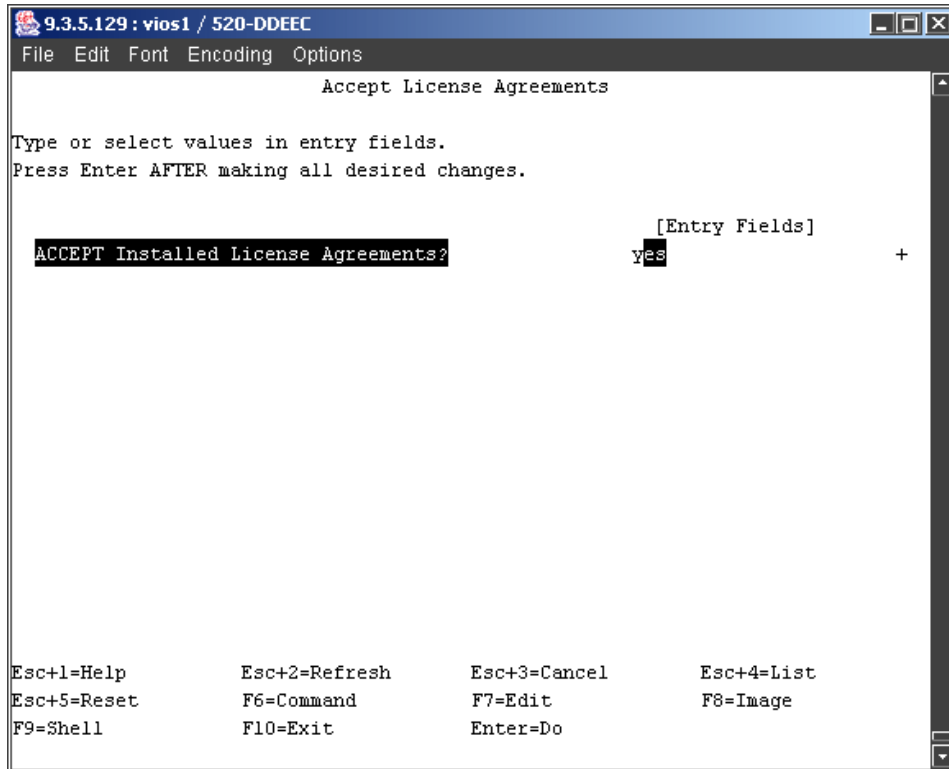


Figure 5-10 Accept License Agreements

After you have accepted the license agreements exit the menu by pressing F10 (or ESC+0) and you will see the Virtual I/O Server login shown in Figure 5-11:



Figure 5-11 IBM Virtual I/O Server login menu

23. Login as padmin and verify the new Virtual I/O Server Version with the **ioslevel** command.
24. Check also the configuration of all disks and Ethernet adapters on the Virtual I/O Server and the mapping of the virtual resources to the virtual I/O clients. Use the **lsmmap -all** and **lsdev -virtual** command.
25. Start the client partitions.

Verify the Virtual I/O Server environment, document the update, and create a new backup of your Virtual I/O Server.

## 5.1.4 Migration from DVD managed by IVM

Before you begin the migration from an DVD using the Integrated Virtualization Manager make sure that the following requirements are fulfilled:

- ▶ A DVD drive is assigned to the Virtual I/O Server partition and you have the Virtual I/O Server Version 2.1 Migration DVD.
- ▶ The Virtual I/O Server is currently at version 1.3, or later.
- ▶ The partition profile data for the management partition and its clients is backed up before you back up the Virtual I/O Server. Use the **bkprofdata** command to save the partition configuration data to a safe location.

**Note:** The IVM configuration in Virtual I/O Server 2.1 is not backward compatible. If you want to revert back to an earlier version of the Virtual I/O Server, you need to restore the partition configuration data from the backup file.

- ▶ Run the **backupios** command, and save the mksysb image to a safe location.

To start the Virtual I/O Server migration, follow the next steps:

1. Step 1 is for a Blade server environment only:

Access the Virtual I/O Server logical partition using the management module of the blade server:

- a. Verify that all logical partitions except the Virtual I/O Server logical partition are shut down.
  - b. Insert the Virtual I/O Server Migration DVD into the DVD drive assigned to your Virtual I/O Server partition.
  - c. Use telnet to connect to the management module of the Blade server on which the Virtual I/O Server logical partition is located.
  - d. Enter the following command: **env -T system:blade[x]** where x is the specific number of the blade to be migrated.
  - e. Enter the following **console** command.
  - f. Login into the Virtual I/O Server using the padmin user.
  - g. Enter the following **shutdown -restart** command.
  - h. When you see the system management services (SMS) logo appear, select 1 to enter the SMS menu.
  - i. Go to step 3 below.
2. Step 2 is for a non-Blade server environment only:

Access the Virtual I/O Server partition using the Advanced System Management Interface (ASMI) with a Power Systems server that is not managed by an HMC:

- a. Verify that all logical partitions except the Virtual I/O Server logical partition are shut down.
  - b. Insert the Virtual I/O Server Migration DVD into the Virtual I/O Server logical partition.
  - c. Log in to the ASCII terminal to communicate with the Virtual I/O Server. See *Access the ASMI without an HMC* if you need assistance at:  
<http://publib.boulder.ibm.com/infocenter/systems/scope/hw/topic/i-phby/ascii.htm>
  - d. Sign on to the Virtual I/O Server using the padmin user.
  - e. Enter the **shutdown -restart** command.
  - f. When you see the SMS logo appear, select 1 to enter the SMS menu.
3. Select the boot device:
    - a. Select option 5 *Select Boot Options* and press Enter.
    - b. Select option 1 *Select Install/Boot Device* and press Enter.
    - c. Select *IDE* and press Enter.
    - d. Select the device number that corresponds to the DVD and press Enter. You can also select *List all devices* and select the device number from a list and press Enter.
    - e. Select Normal mode boot.
    - f. Select Yes to exit SMS.
  4. Install the Virtual I/O Server:

Follow the steps described in 5.1.3, “Migration from DVD managed by an HMC”, beginning with step 13.

## 5.2 Virtual I/O server backup strategy

A complete disaster recovery strategy for the Virtual I/O Server should include backing up several components such that you can recover the virtual devices and their physical backing devices.

The Virtual I/O Server contains the following types of information that you need to back up:

- ▶ The Virtual I/O Server operating system includes the base code, applied fix packs, custom device drivers to support disk subsystems, Kerberos, and LDAP client configurations. All of this information is backed up when you use the **backupios** command. In situations where you plan to restore the Virtual



I/O Server to the same system from which it was backed up, backing up only the Virtual I/O Server itself is usually sufficient.

- ▶ User-defined virtual devices include metadata, such as virtual device mappings, that define the relationship between the physical environment and the virtual environment. This data can be saved to a location that is automatically backed up when you use the **backupios** command.

In a complete disaster recovery scenario, the Virtual I/O Server may be restored to a new or repaired system. You must then back up both the Virtual I/O Server and user-defined virtual devices.

Furthermore, in this situation, you must also back up the following components of your environment in order to fully recover your Virtual I/O Server configuration:

- ▶ External device configurations, such as Storage Area Network (SAN) devices
- ▶ Resources defined on the Hardware Management Console (HMC) or on the Integrated Virtualization Manager (IVM), such as processor and memory allocations, physical or virtual adapter configuration
- ▶ The operating systems and applications running in the client logical partitions

## 5.2.1 Backup external device configuration

Planning should be included into the end-to-end backup strategy for the event that a natural or man-made disaster destroys a complete site. This is probably part of your disaster recovery strategy, but consider it in the complete backup strategy. The backup strategy for this depends on the hardware specifics of the storage, networking equipment, and SAN devices, to name but a few. Examples of the type of information you will need to record include the network virtual local area network (VLAN) or logical unit number (LUN) information from a storage subsystem.

This information is beyond the scope of this document, but we mention it here to make you aware that a complete disaster recovery solution for a physical or virtual server environment has a dependency on this information. The method to collect and record the information depends not only on the vendor and model of the infrastructure systems at the primary site, but also on what is present at the disaster recovery site.

## 5.2.2 Backup HMC resources

If the system is managed by an HMC, the HMC information needs to be backed up. The definition of the Virtual I/O Server logical partition on the HMC includes, for example, how much CPU and memory and what physical adapters are to be used. In addition to this, you have the virtual device configuration (for example,

virtual Ethernet adapters and to which virtual LAN ID they belong) that needs to be captured. For information on this topic, see the IBM Systems Hardware Information Center under the “Backing up partition profile data” topic at:

<http://publib.boulder.ibm.com/infocenter/eserver/v1r3s/topic/iphai/backupprofdata.htm>

Note that, especially if you are planning for disaster recovery, you might have to rebuild selected HMC profiles from scratch on new hardware. In this case, it is important to have detailed documentation of the configuration, such as how many Ethernet cards are needed. Using the system plans and the viewer can help record such information but you should check that this is appropriate and that it records all the information needed in every case.

Starting with HMC V7, you can save the current system configuration to an HMC system plan. The system plan can be redeployed to rebuild the complete partition configuration.

**Note:** Check that the system plan is valid by viewing the report; look for a message saying that the system plan cannot be deployed (in red).

Refer to the `mksysplan` command and the HMC interface for more information.

### 5.2.3 Backup IVM resources

If the system is managed by the Integrated Virtualization Manager, you need to back up your partition profile data for the management partition and its clients before you back up the Virtual I/O Server operating system.

To do so, from the Service Management menu, click **Backup/Restore**. The Backup/Restore page is displayed. Then click **Generate Backup**.

This operation can also be done from the Virtual I/O Server. To do so, enter this command:

```
bkprofdata -o backup -f /home/padmin/profile.bak
```

### 5.2.4 Backup operating systems from the client logical partitions

Backing up and restoring the operating systems and applications running in the client logical partitions is a separate topic. This is because the Virtual I/O Server manages only the devices and the linking of these devices along with the Virtual I/O operating system itself. The AIX, IBM i or Linux operating system-based

clients of the Virtual I/O Server should have a backup strategy independently defined as part of your existing server backup strategy.

For example, if you have an AIX 6.1 server made up of virtual disks and virtual networks, you would still have an `mksysb`, `savevg`, or equivalent strategy in place to back up the system. This backup strategy can rely on the virtual infrastructure – for example, backing up to an Network Installation Manager (NIM) or IBM Tivoli Storage Manager server over a virtual network interface through a physical Shared Ethernet Adapter.

## 5.2.5 Backup the Virtual I/O Server operating system

The Virtual I/O Server operating system consists of the base code, fix packs, custom device drivers to support disk subsystems, and user-defined customization. An example of user-defined customization can be as simple as the changing of the Message of the Day or the security settings.

These settings, after an initial setup, will probably not change apart from the application of fix packs, so a sensible backup strategy for the Virtual I/O Server is after fix packs have been applied or configuration changes made. Although we discuss the user-defined virtual devices in the next section, it is worth noting that the backup of the Virtual I/O Server will capture some of this data. With this fact in mind, you can define the schedule for the Virtual I/O operating system backups to occur more frequently to cover both the Virtual I/O operating system and the user-defined devices in one single step.

Starting with the release of Virtual I/O Server Version 1.3, you can schedule jobs with the `crontab` command. You can schedule the following backup steps to take place at regular intervals using this command.

The `backupios` command creates a backup of the Virtual I/O Server to a bootable tape, a DVD, or a file system (local or a remotely mounted Network File System).

**Note:** Consider the following:

- ▶ Virtual device mappings (that is, customized metadata) is backed up by default. Nothing special needs to happen.
- ▶ Client data is not backed up.

You can back up and restore the Virtual I/O Server by the means listed in Table 5-1.

Table 5-1 Virtual I/O Server backup and restore methods

Backup method	Media	Restore method
To tape	Tape	From tape
To DVD	DVD-RAM	From DVD
To remote file system	nim_resources.tar image	From an HMC using the Network Installation Management on Linux (NIMOL) facility and the <b>installios</b> command
To remote file system mksysb image	mksysb image	From an AIX NIM server and a standard mksysb system installation
Tivoli Storage Manager	mksysb image	Tivoli Storage Manager

## 5.3 Scheduling backups of the Virtual I/O Server

You can schedule regular backups of the Virtual I/O Server and user-defined virtual devices to ensure that your backup copy accurately reflects the current configuration.

To ensure that your backup of the Virtual I/O Server accurately reflects your current running Virtual I/O Server, you should back up the Virtual I/O Server each time that its configuration changes. For example:

- ▶ Changing the Virtual I/O Server, for example installing a fix pack.
- ▶ Adding, deleting, or changing the external device configuration, such as changing the SAN configuration.
- ▶ Adding, deleting, or changing resource allocations and assignments for the Virtual I/O Server, such as memory, processors, or virtual and physical devices.
- ▶ Adding, deleting, or changing user-defined virtual device configurations, such as virtual device mappings.

You can then back up your Virtual I/O Server manually after any of these modifications.

You can also schedule backups on a regular basis using the crontab function. You therefore create a script for backing up the Virtual I/O Server, and save it in a directory that is accessible to the padmin user ID. For example, create a script called *backup* and save it in the /home/padmin directory. Ensure that your script

includes commands for backing up the Virtual I/O Server and saving information about user-defined virtual devices.

Mounting the image directory of your NIM server and posting the backups to this directory is an approach to quickly deploy a backup if the need arises.

Then create a crontab file entry that runs the backup script on regular intervals. For example, to run backup every Saturday at 2:00 a.m., type the following commands:

```
$ crontab -e
0 2 0 0 6 /home/padmin/backup
```

When you are finished, remember to save and exit.

## 5.4 Backing up the Virtual I/O Server operating system

The following topics explain the options that can be used to back up the Virtual I/O Server.

### 5.4.1 Backing up to tape

You can back up the Virtual I/O Server base code, applied fix packs, custom device drivers to support disk subsystems, and some user-defined metadata to tape.

If the system is managed by the Integrated Virtualization Manager, then you need to back up your partition profile data for the management partition and its clients before you back up the Virtual I/O Server. To do so, refer to 5.2.3, “Backup IVM resources” on page 190.

You can find the device name on the Virtual I/O Server by typing the following command:

```
$ lsdev -type tape
name          status      description
rmt0          Available  Other SCSI Tape Drive
```

If the device is in the *Defined* state, type the following command where *dev* is the name of your tape device:

```
cfgdev -dev dev
```

Run the **backupios** command with the **-tape** option. Specify the path to the device. Use the **-accept** flag to automatically accept licences. For example:

```
backupios -tape /dev/rmt0 -accept
```

Example 5-1 illustrates a **backupios** command execution to back up the Virtual I/O Server on a tape.

*Example 5-1 Backing up the Virtual I/O Server to tape*

---

```
$ backupios -tape /dev/rmt0
```

```
Creating information file for volume group volgrp01.
```

```
Creating information file for volume group storage01.
```

```
Backup in progress. This command can take a considerable amount of time  
to complete, please be patient...
```

```
Creating information file (/image.data) for rootvg.
```

```
Creating tape boot image.....
```

```
Creating list of files to back up.
```

```
Backing up 44950 files.....
```

```
44950 of 44950 files (100%)
```

```
0512-038 mksysb: Backup Completed Successfully.
```

---

## 5.4.2 Backing up to a DVD-RAM

You can back up the Virtual I/O Server base code, applied fix packs, custom device drivers to support disk subsystems, and some user-defined metadata to DVD.

If the system is managed by the Integrated Virtualization Manager, then you need to back up your partition profile data for the management partition and its clients before you back up the Virtual I/O Server. To do so, refer to 5.2.3, “Backup IVM resources” on page 190.

To back up the Virtual I/O Server to one or more DVDs, you generally use DVD-RAM media. Vendor disk drives might support burning to additional disk types, such as CD-RW and DVD-R. Refer to the documentation for your drive to determine which disk types are supported.

DVD-RAM media can support both **-cdformat** and **-udf** format flags, while DVD-R media only supports the **-cdformat**.

The DVD device cannot be virtualized and assigned to a client partition when using the **backupios** command. Remove the device mapping from the client before proceeding with the backup.

You can find the device name on the Virtual I/O Server by typing the following command:

```
$ lsdev -type optical
name          status    description
cd0           Available SATA DVD-RAM Drive
```

If the device is in the *Defined* state, type the following command where *dev* is the name of your CD or DVD device:

```
cfgdev -dev dev
```

Run the **backupios** command with the **-cd** option. Specify the path to the device. Use the **-accept** flag to automatically accept licenses. For example:

```
backupios -cd /dev/cd0 -accept
```

Example 5-2 illustrates a **backupios** command execution to back up the Virtual I/O Server on a DVD-RAM.

*Example 5-2 Backing up the Virtual I/O Server to DVD-RAM*

---

```
$ backupios -cd /dev/cd0 -udf -accept
```

```
Creating information file for volume group volgrp01.
```

```
Creating information file for volume group storage01.
```

```
Backup in progress. This command can take a considerable amount of time
to complete, please be patient...
```

```
Initializing mkcd log: /var/adm/ras/mkcd.log...
```

```
Verifying command parameters...
```

```
Creating image.data file...
```

```
Creating temporary file system: /mkcd/mksysb_image...
```

```
Creating mksysb image...
```

```
Creating list of files to back up.
```

```
Backing up 44933 files.....
```

```
44933 of 44933 files (100%)
```

```
0512-038 mksysb: Backup Completed Successfully.
```

```
Populating the CD or DVD file system...
```

```
Copying backup to the CD or DVD file system...
```

```
.....
```

```
.....
```

```
Building chrp boot image...
```

---

**Note:** If the Virtual I/O Server does not fit on one DVD, then the **backupios** command provides instructions for disk replacement and removal until all the volumes have been created.

### 5.4.3 Backing up to a remote file

The major difference for this type of backup compared to tape or DVD media is that all of the previous commands resulted in a form of bootable media that can be used to directly recover the Virtual I/O Server.

Backing up the Virtual I/O Server base code, applied fix packs, custom device drivers to support disk subsystems, and some user-defined metadata to a file will result in either:

- ▶ A `nim_resources.tar` file that contains all the information needed for a restore. This is the preferred solution if you intend to restore the Virtual I/O Server on the same system. This backup file can both be restored by the HMC or a NIM server.
- ▶ An `mksysb` image. This solution is preferred if you intend to restore the Virtual I/O Server from a Network Installation Management (NIM) server.

**Note:** The `mksysb` backup of the Virtual I/O Server can be extracted from the tar file created in a full backup, so either method is appropriate if the restoration method uses a NIM server.

Whichever method you choose, if the system is managed by the Integrated Virtualization Manager, you need to back up your partition profile data for the management partition and its clients before you back up the Virtual I/O Server. To do so, refer to 5.2.3, “Backup IVM resources” on page 190.

#### Mounting the remote file system

You can use the **backupios** command to write to a local file on the Virtual I/O Server, but the more common scenario is to perform a backup to a remote NFS-based storage. The ideal situation might be to use the NIM server as the destination because this server can be used to restore these backups. In the following example, a NIM server has a host name of `nim_server` and the Virtual I/O Server is `vios1`.

The first step is to set up the NFS-based storage export on the NIM server. Here, we export a file system named `/export/ios_backup`, and in this case, `/etc/exports` looks similar to the following:

```
$ mkdir /export/ios_backup
```



```
$ mknfsxp -d /export/ios_backup -B -S sys,krb5p,krb5i,krb5,dh -t rw -r vios1
$ grep lpar01 /etc/exports
/export/ios_backup -sec=sys:krb5p:krb5i:krb5:dh,rw,root=vios1
```

**Important:** The NFS server must have the root access NFS attribute set on the file system exported to the Virtual I/O Server logical partition for the backup to succeed.

In addition, make sure that the name resolution is functioning from the NIM server to the Virtual I/O Server and back again (reverse resolution) for both the IP and host name. To edit the name resolution on the Virtual I/O Server, use the **hostmap** command to manipulate the `/etc/hosts` file or the **cfgnamesrv** command to change the DNS parameters.

The backup of the Virtual I/O Server can be large, so ensure that the system `ulimits` parameter in the `/etc/security/limits` file on the NIM server is set to `-1` and therefore will allow the creation of large files.

With the NFS export and name resolution set up, the file system needs to be mounted on the Virtual I/O Server. You can use the **mount** command:

```
$ mkdir /mnt/backup
$ mount nim_server:/export/ios_backup /mnt/backup
```

**Note:** We recommend the remote file system should be mounted automatically at bootup of the Virtual I/O Server to simplify the scheduling of regular backups.

## Backing up to a `nim_resources.tar` file

Once the remote file system is mounted, you can start the backup operation to the `nim_resources.tar` file.

Backing up the Virtual I/O Server to a remote file system creates the `nim_resources.tar` image in the directory you specify. The `nim_resources.tar` file contains all the necessary resources to restore the Virtual I/O Server, including the `mksysb` image, the `bosinst.data` file, the network boot image, and the Shared Product Object Tree (SPOT) resource.

The **backupios** command empties the `target_disks_stanza` section of `bosinst.data` and sets `RECOVER_DEVICES=Default`. This allows the `mksysb` file generated by the command to be cloned to another logical partition. If you plan to use the `nim_resources.tar` image to install to a specific disk, then you need to repopulate the `target_disk_stanza` section of `bosinst.data` and replace this file in the `nim_resources.tar` image. All other parts of the `nim_resources.tar` image must remain unchanged.

Run the **backupios** command with the **-file** option. Specify the path to the target directory. For example:

```
backupios -file /mnt/backup
```

Example 5-3 illustrates a **backupios** command execution to back up the Virtual I/O Server on a `nim_resources.tar` file.

*Example 5-3 Backing up the Virtual I/O Server to the `nim_resources.tar` file*

---

```
$ backupios -file /mnt/backup
```

```
Creating information file for volume group storage01.
```

```
Creating information file for volume group volgrp01.
```

```
Backup in progress. This command can take a considerable amount of time  
to complete, please be patient...
```

---

This command created a `nim_resources.tar` file that you can use to restore the Virtual I/O Server from the HMC as described in , “Restoring from a `nim_resources.tar` file with the HMC” on page 207.

**Note:** The argument for the **backupios -file** command is a directory. The `nim_resources.tar` file is stored in this directory.

## Backing up to an `mksysb` file

Alternatively, once the remote file system is mounted, you can start the backup operation to an `mksysb` file. The `mksysb` image is an installable image of the root volume group in a file.

Run the **backupios** command with the **-file** option. Specify the path to the target directory and specify the **-mksysb** parameter. For example:

```
backupios -file /mnt/backup -mksysb
```

Example 5-4 on page 198 illustrates a **backupios** command execution to back up the Virtual I/O Server on a `mksysb` file.

*Example 5-4 Backing up the Virtual I/O Server to the `mksysb` image*

---

```
$ backupios -file /mnt/VIOS_BACKUP_130ct2008.mksysb -mksysb
```

```
/mnt/VIOS_BACKUP_130ct2008.mksysb doesn't exist.
```

```
Creating /mnt/VIOS_BACKUP_130ct2008.mksysb
```

```
Creating information file for volume group storage01.
```

```
Creating information file for volume group volgrp01.  
Backup in progress. This command can take a considerable amount of time  
to complete, please be patient...
```

```
Creating information file (/image.data) for rootvg.
```

```
Creating list of files to back up...  
Backing up 45016 files.....  
45016 of 45016 files (100%)  
0512-038 savevg: Backup Completed Successfully.
```

---

**Note:** If you intend to use a NIM server for the restoration, it must be running a level of AIX that can support the Virtual I/O Server installation. For this reason, the NIM server should be running the very latest technology level and service packs at all times. For the restoration of any backups of a Virtual I/O Server Version 2.1, your NIM server needs to be at the latest AIX Version 6.1 level. For a Virtual I/O Server 1.x environment, your NIM servers needs to be at the latest AIX Version 5.3 level.

#### 5.4.4 Backing up user-defined virtual devices

Once you backed up the Virtual I/O Server operating system, you still need to back up the user-defined virtual devices:

- ▶ If you are restoring to the same server, some information might be available such as data structures (storage pools or volume groups and logical volumes) held on non-rootvg disks.
- ▶ If you are restoring to new hardware, these devices cannot be automatically recovered because the disk structures will not exist.
- ▶ If the physical devices exist in the same location and structures such as logical volumes are intact, the virtual devices such as virtual target SCSI and Shared Ethernet Adapters are recovered during the restoration.

In the disaster recovery situation where these disk structures do not exist and network cards are at different location codes, you need to make sure to back up the following:

- ▶ Any user-defined disk structures such as storage pools or volume groups and logical volumes
- ▶ The linking of the virtual device through to the physical devices

These devices will mostly be created at the Virtual I/O Server build and deploy time, but will change depending on when new clients are added or changes are

made. For this reason, a weekly schedule or manual backup procedure when configuration changes are made is appropriate.

## Backing up disk structures with `savevgstruct`

Use the `savevgstruct` command to back up user-defined disk structures. This command writes a backup of the structure of a named volume group (and therefore storage pool) to the `/home/ios/vgbackups` directory.

For example, assume you have the following storage pools:

```
$ lssp
Pool          Size(mb)  Free(mb)  Alloc  Size(mb)  BDs
rootvg       139776    107136    128    139776    0
storage01    69888    69760    64     69888    1
volgrp01     69888    69760    64     69888    1
```

Then you run the `savevgstruct storage01` command to back up the structure in the `storage01` volume group:

```
$ savevgstruct storage01
```

```
Creating information file for volume group storage01.
```

```
Creating list of files to back up.
```

```
Backing up 6 files
```

```
6 of 6 files (100%)
```

```
0512-038 savevg: Backup Completed Successfully.
```

The `savevgstruct` command is automatically called before the backup commences for all active non-rootvg volume groups or storage pools on a Virtual I/O Server when the `backupios` command is run. Because this command is called before the backup commences, the volume group structures will be included in the system backup. For this reason, you can use the `backupios` command to back up the disk structure as well, so the frequency that this command runs might increase.

**Note:** The volume groups or storage pools need to be activated for the backup to succeed.

Only active volume groups or storage pools are automatically backed up by the `backupios` command. Use the `lsvg` or `lssp` command to list and `activatevg` to activate the volume groups or storage pools if necessary before starting the backup.

## Backing up virtual devices linking information

The last item to back up is the linking information. You can gather this information from the output of the `lsmmap` command, as shown in Example 5-5.

*Example 5-5 Sample output from the lsmmap command*

```
$ lsmmap -net -all
SVEA Physloc
-----
ent2  U9117.MMA.101F170-V1-C11-T1

SEA          ent5
Backing device ent0
Status       Available
Physloc      U789D.001.DQDYKYW-P1-C4-T1

$ lsmmap -all
SVSA          Physloc          Client Partition ID
-----
vhost0       U9117.MMA.101F170-V1-C21  0x00000003

VTD          aix61_rvg
Status       Available
LUN          0x8100000000000000
Backing device hdisk7
Physloc      U789D.001.DQDYKYW-P1-C2-T2-W201300A0B811A662-L1000000000000

SVSA          Physloc          Client Partition ID
-----
vhost1       U9117.MMA.101F170-V1-C22  0x00000004

VTD          aix53_rvg
Status       Available
LUN          0x8100000000000000
Backing device hdisk8
Physloc      U789D.001.DQDYKYW-P1-C2-T2-W201300A0B811A662-L2000000000000
```

From this example, the `vhost0` device in slot 10 on the HMC (the C21 value in the location code) is linked to the `hdisk7` device and named `aix61_rvg`. The `vhost1` device holds the `hdisk8` device and named `aix53_rvg`. For the network, the virtual Ethernet adapter `ent2` is linked to the physical Ethernet adapter `ent0` making the `ent5` Shared Ethernet Adapter.

**Consideration:** The previous output does not gather information such as SEA control channels (for SEA Failover), IP addresses to ping, and whether threading is enabled for the SEA devices. These settings and any other changes that have been made (for example MTU settings) must be documented separately, as explained later in this section.

**Note:** It is also vitally important to use the slot numbers as a reference for the virtual SCSI and virtual Ethernet devices, not the vhost number or ent number.

The vhost and ent devices are assigned by the Virtual I/O Server because they are found at boot time or when the `cfgdev` command is run. If you add in more devices after subsequent boots or with the `cfgdev` command, these will be sequentially numbered.

In the vhost0 example, the important information to note is that it is not vhost0 but that the virtual SCSI server in slot 21 (the C21 value in the location code) is mapped to a LUN disk hdisk7. The vhost and ent numbers are assigned sequentially by the Virtual I/O Server at initial discovery time and should be treated with caution to rebuild user-defined linking devices.

## Backing up additional information

You should also save the information about network settings, adapters, users, and security settings to the `/home/padmin` directory by running each command in conjunction with the `tee` command as follows:

```
command | tee /home/padmin/filename
```

Where:

- `command` is the command that produces the information you want to save.
- `filename` is the name of the file to which you want to save the information.

The `/home/padmin` directory is backed up using the `backupios` command; therefore, it is a good location to collect configuration information prior to a backup. Table 5-2 provides a summary of the commands that help you to save the information.

Table 5-2 Commands to save information about Virtual I/O Server

Command	Information provided (and saved)
<code>cfgnamesrv -ls</code>	Shows all system configuration database entries related to domain name server information used by local resolver routines.
<code>entstat -all devicename</code>  devicename is the name of a device. Run this command for each device whose attributes or statistics you want to save.	Shows Ethernet driver and device statistics for the device specified.

Command	Information provided (and saved)
<b>hostmap -ls</b>	Shows all entries in the system configuration database.
<b>ioslevel</b>	Shows the current maintenance level of the Virtual I/O Server.
<b>lsdev -dev devicename -attr</b>  devicename is the name of a device. Run this command for each device whose attributes or statistics you want to save. You generally want to save the customized devices attributes. Try to keep track of them when managing the Virtual I/O Server.	Shows the attributes of the device specified.
<b>lsdev -type adapter</b>	Shows information about physical and logical adapters.
<b>lsuser</b>	Shows a list of all attributes of all the system users.
<b>netstat -routinfo</b>	Shows the routing tables, including the user-configured and current costs of each route.
<b>netstat -state</b>	Shows the routing tables, including the user-configured and current costs of each route.
<b>optimizenet -list</b>	Shows characteristics of all network tuning parameters, including the current and reboot value, range, unit, type, and dependencies.
<b>viorecure -firewall view</b>	Shows a list of allowed ports.
<b>viorecure -view -nonint</b>	Shows all of the security level settings for non-interactive mode.

## 5.4.5 Backing up using IBM Tivoli Storage Manager

You can use the IBM Tivoli Storage Manager to automatically back up the Virtual I/O Server on regular intervals, or you can perform incremental backups.

### IBM Tivoli Storage Manager automated backup

You can automate backups of the Virtual I/O Server using the **crontab** command and the Tivoli Storage Manager scheduler.

Before you start, complete the following tasks:

- ▶ Ensure that you configured the Tivoli Storage Manager client on the Virtual I/O Server. If it is not configured, refer to 11.2, “Configuring the IBM Tivoli Storage Manager client” on page 380.
- ▶ Ensure that you are logged into the Virtual I/O Server as the administrator (padmin).

To automate backups of the Virtual I/O Server, complete the following steps:

1. Write a script that creates an mksysb image of the Virtual I/O Server and save it in a directory that is accessible to the padmin user ID. For example, create a script called backup and save it in the /home/padmin directory. If you plan to restore the Virtual I/O Server to a different system than the one from which it was backed up, then ensure that your script includes commands for saving information about user-defined virtual devices. For more information, see the following tasks:
  - For instructions about how to create an mksysb image, see “Backing up to an mksysb file” on page 198.
  - For instructions about how to save user-defined virtual devices, see 5.4.4, “Backing up user-defined virtual devices” on page 199.
2. Create a crontab file entry that runs the backup script on a regular interval. For example, to create an mksysb image every Saturday at 2:00 a.m., type the following commands:

```
$ crontab -e
0 2 0 0 6 /home/padmin/backup
```

When you are finished, remember to save and exit.

3. Work with the Tivoli Storage Manager administrator to associate the Tivoli Storage Manager client node with one or more schedules that are part of the policy domain. This task is not performed on the Tivoli Storage Manager client on the Virtual I/O Server, but by the Tivoli Storage Manager administrator on the Tivoli Storage Manager server.
4. Start the client scheduler and connect to the server schedule using the **dsmc** command as follows:

```
dsmc -schedule
```

5. If you want the client scheduler to restart when the Virtual I/O Server restarts, add the following entry to the /etc/inittab file:

```
itsm::once:/usr/bin/dsmc sched > /dev/null 2>&1 # TSM scheduler
```



## IBM Tivoli Storage Manager incremental backup

You can back up the Virtual I/O Server at any time by performing an incremental backup with the Tivoli Storage Manager.

Perform incremental backups in situations where the automated backup does not suit your needs. For example, before you upgrade the Virtual I/O Server, perform an incremental backup to ensure that you have a backup of the current configuration. Then, after you upgrade the Virtual I/O Server, perform another incremental backup to ensure that you have a backup of the upgraded configuration.

Before you start, complete the following tasks:

- ▶ Ensure that you configured the Tivoli Storage Manager client on the Virtual I/O Server. For instructions, see 11.2, “Configuring the IBM Tivoli Storage Manager client” on page 380.
- ▶ Ensure that you have an mksysb image of the Virtual I/O Server. If you plan to restore the Virtual I/O Server to a different system than the one from which it was backed up, then ensure that the mksysb includes information about user-defined virtual devices. For more information, see the following tasks:
  - For instructions about how to create an mksysb image, see “Backing up to an mksysb file” on page 198.
  - For instructions about how to save user-defined virtual devices, see 5.4.4, “Backing up user-defined virtual devices” on page 199.

To perform an incremental backup of the Virtual I/O Server, run the **dsmc** command. For example:

```
dsmc -incremental sourcefilespec
```

Where `sourcefilespec` is the directory path to where the mksysb file is located. For example, `/home/padmin/mksysb_image`.

## 5.5 Restoring the Virtual I/O Server

With all of the different backups described and the frequency discussed, we now describe how to rebuild the server from scratch. The situation we work through is a Virtual I/O Server hosting an AIX operating system-based client partition running on virtual disk and network. We work through the restore from the uninstalled bare metal Virtual I/O Server upward and discuss where each backup strategy will be used.

This complete end-to-end solution is only for this extreme disaster recovery scenario. If you need to back up and restore a Virtual I/O Server onto the same server, the restoration of the operating system is probably of interest.

### 5.5.1 Restoring the HMC configuration

In the most extreme case of a natural or man-made disaster that has destroyed or rendered unusable an entire data center, systems might have to be restored to a disaster recovery site. In this case, you need another HMC and server location to which to recover your settings. You should also have a disaster recovery server in place with your HMC profiles ready to start recovering your systems.

The details of this are beyond the scope of this document but would, along with the following section, be the first steps for a disaster recovery.

### 5.5.2 Restoring other IT infrastructure devices

All other IT infrastructure devices, such as network routers, switches, storage area networks and DNS servers, to name just a few, also need to be part of an overall IT disaster recovery solution. Having mentioned them, we say no more about them apart from making you aware that not just the Virtual I/O Server but the whole IT infrastructure will rely on these common services for a successful recovery.

### 5.5.3 Restoring the Virtual I/O Server operating system

This section details how to restore the Virtual I/O Server. We describe how to recover from a complete disaster.

If you migrated to a different system and if this system is managed by the Integrated Virtualization Manager, you need to restore your partition profile data for the management partition and its clients before you restore the Virtual I/O Server.

To do so, from the Service Management menu, click **Backup/Restore**. The Backup/Restore page is displayed. Then click **Restore Partition Configuration**.

#### Restoring from DVD backup

The backup procedures described in this chapter created bootable media that you can use to restore as stand-alone backups.

Insert the first DVD into the DVD drive and boot the Virtual I/O server partition into SMS mode, making sure the DVD drive is assigned to the partition. Select,

using the SMS menus, to install from the DVD drive and work through the usual installation procedure.

**Note:** If the DVD backup spanned multiple disks during the install, you will be prompted to insert the next disk in the set with a message similar to the following:

Please remove volume 1, insert volume 2, and press the ENTER key.

## Restoring from tape backup

The procedure for the tape is similar to the DVD procedure. Because this is a bootable media, just place the backup media into the tape drive and boot the Virtual I/O Server partition into SMS mode. Select to install from the tape drive and follow the same procedure as previously described.

## Restoring from a `nim_resources.tar` file with the HMC

If you made a full backup of the Virtual I/O Server to a `nim_resources.tar` file, you can use the HMC to restore it using the `installios` command.

To do so, the tar file must be located either on the HMC, an NFS-accessible directory, or a DVD. To make the `nim_resources.tar` file accessible for restore, we performed the following steps:

1. Created a directory named `backup` using the `mkdir /home/padmin/backup` command.
2. Checked that the NFS server was exporting a file system with the `showmount nfs_server` command.
3. Mounted the NFS-exported file system onto the `/home/padmin/backup` directory.
4. Copied the tar file created in “Backing up to a `nim_resources.tar` file” on page 197 to the NFS mounted directory using the following command:

```
$ cp /home/padmin/backup_loc/nim_resources.tar /home/padmin/backup
```

At this stage, the backup is ready to be restored to the Virtual I/O Server partition using the `installios` command on the HMC or an AIX partition that is a NIM server. The restore procedure will shut down the Virtual I/O Server partition if it is still running. The following is an example of the command help:

```
hscroot@hmc1:~> installios -?  
installios: usage: installios [-s managed_sys -S netmask -p partition  
-r profile -i client_addr -d source_dir -m mac_addr  
-g gateway [-P speed] [-D duplex] [-n] [-l language]]  
| -u
```

Using the **installios** command, the **-s** managed\_sys option requires the HMC defined system name, the **-p** partition option requires the name of the Virtual I/O Server partition, and the **-r** profile option requires the partition profile you want to use to boot the Virtual I/O Server partition during the recovery.

If you do not specify the **-m** flag and include the MAC address of the Virtual I/O Server being restored, the restore will take longer because the **installios** command shuts down the Virtual I/O Server and boots it in SMS to determine the MAC address. The following is an example of the use of this command:

```
hscroot@hmc1:~> installios -s MT_B_p570_MMA_101F170 -S 255.255.254.0 -p vios1
-r default -i 9.3.5.111 -d 9.3.5.5:/export_fs -m 00:02:55:d3:dc:34 -g 9.3.4.1
```

**Note:** If you do not input a parameter, the **installios** command will prompt you for one.

```
hscroot@hmc1:~> installios
```

The following objects of type "managed system" were found. Please select one:

1. MT\_B\_p570\_MMA\_101F170
2. MT\_A\_p570\_MMA\_100F6A0
3. p550-SN106629E

Enter a number (1-3): 1

The following objects of type "virtual I/O server partition" were found. Please select one:

1. vios2
2. vios1

Enter a number (1-2):

At this point, open a terminal console on the server to which you are restoring in case user input is required. Then run the **installios** command as described above.

Following this command, NIMOL on the HMC takes over the NIM process and mounts the exported file system to process the **backupios** tar file created on the Virtual I/O Server previously. NIMOL on the HMC then proceeds with the installation of the Virtual I/O Server and a reboot of the partition completes the install.

### Notes:

- ▶ The configure client network setting must be set to no when prompted by the **installios** command. This is because the physical adapter we are installing the backup through might already be used by an SEA and the IP configuration will fail if this is the case. Log in and configure the IP if necessary after the installation using a console session.
- ▶ If the command seems to be taking a long time to restore, this is most commonly caused by a speed or duplex misconfiguration in the network.

## Restoring from a file with the NIM server

The **installios** command is also available on the NIM server, but at present it only supports installations from the base media of the Virtual I/O Server. The method we used from the NIM server was to install the mksysb image. This can either be the mksysb image generated with the **-mksysb** flag in the **backupios** command shown previously or you can extract the mksysb image from the **nim\_resources.tar** file.

Whatever method you use, after you have stored the mksysb file this on the NIM server, you need to create a NIM mksysb resource as shown:

```
# nim -o define -t mksysb -aserver=master
-a location=/export/mksysb/VIOS_BACKUP_130ct2008.mksysb VIOS_mksysb
# lsnim VIOS_mksysb
VIOS_mksysb      resources      mksysb
```

After NIM mksysb resource has been successfully created, generate a SPOT from the NIM mksysb resource or use the SPOT available at the latest AIX technology and service pack level. To create the SPOT from the NIM mksysb resource, run the command:

```
# nim -o define -t spot -a server=master -a location=/export/spot/ -a
source=VIOS_mksysb VIOS_SPOT
```

```
Creating SPOT in "/export/spot" on machine "master" from "VIOS_mksysb" ...
Restoring files from BOS image. This may take several minutes ...
```

```
# lsnim VIOS_SPOT
VIOS_SPOT      resources      spot
```

With the SPOT and the mksysb resources defined to NIM, you can install the Virtual I/O Server from the backup. If the Virtual I/O Server partition you are installing is not defined to NIM, make sure that it is now defined as a machine and enter the **smitty nim\_bosinst** fast path command. Select the NIM mksysb resource and SPOT defined previously.

**Important:** Note that the Remain NIM client after install field must be set to no. If this is not set to no, the last step for the NIM installation is to configure an IP address onto the physical adapter through which the Virtual I/O Server has just been installed. This IP address is used to register with the NIM server. If this is the adapter used by an existing Shared Ethernet Adapter (SEA), it will cause error messages to be displayed.

If this is the case, reboot the Virtual I/O Server if necessary, and then login to it using a terminal session and remove any IP address information and the SEA. After this, recreate the SEA and configure the IP address back for the SEA interface.

Now that you have set up the NIM server to push out the backup image, the Virtual I/O Server partition needs to have the remote IPL setup completed. For this procedure, see the *Installing with Network Installation Management* section under the Installation and Migration category of the IBM System p and AIX Information Center at:

<http://publib16.boulder.ibm.com/pseries/index.htm>

**Tip:** One of the main causes of installation problems using NIM is the NFS exports from the NIM server. Make sure that the `/etc/exports` file is correct on the NIM server.

The installation of the Virtual I/O Server should complete, but here is a big difference between restoring to the existing server and restoring to a new disaster recovery server. One of the NIM install options is to preserve the NIM definitions for resources on the target. With this option, NIM attempts to restore any virtual devices that were defined in the original backup. This depends on the same devices being defined in the partition profile (virtual and physical) such that the location codes have not changed.

This means that virtual target SCSI devices and Shared Ethernet Adapters should all be recovered without any need to recreate them (assuming the logical partition profile has not changed). If restoring to the same machine, there is a dependency that the non-rootvg volume groups are present to be imported and any logical volume structure contained on these is intact. To demonstrate this, we operated a specific test scenario: A Virtual I/O Server was booted from a diagnostics CD and the Virtual I/O Server operating system disks were formatted and certified, destroying all data (this was done for demonstration purposes). The other disks containing volume groups and storage pools were not touched.

Using a NIM server, the backup image was restored to the initial Virtual I/O Server operating system disks. Examining the virtual devices after the

installation, the virtual target devices and Shared Ethernet Adapters are all recovered, as shown in Example 5-6.

*Example 5-6 Restore of Virtual I/O Server to the same logical partition*

```

$ lsdev -virtual
name          status      description
ent2          Available  Virtual I/O Ethernet Adapter (1-lan)
ent3          Available  Virtual I/O Ethernet Adapter (1-lan)
ent4          Available  Virtual I/O Ethernet Adapter (1-lan)
ent6          Available  Virtual I/O Ethernet Adapter (1-lan)
vasi0         Available  Virtual Asynchronous Services Interface (VASI)
vbsd0         Available  Virtual Block Storage Device (VBSD)
vhost0        Available  Virtual SCSI Server Adapter
vhost1        Available  Virtual SCSI Server Adapter
vhost2        Available  Virtual SCSI Server Adapter
vhost3        Available  Virtual SCSI Server Adapter
vhost4        Available  Virtual SCSI Server Adapter
vhost5        Available  Virtual SCSI Server Adapter
vhost6        Available  Virtual SCSI Server Adapter
vhost7        Defined    Virtual SCSI Server Adapter
vhost8        Defined    Virtual SCSI Server Adapter
vhost9        Defined    Virtual SCSI Server Adapter
vsa0          Available  LPAR Virtual Serial Adapter
IBMi61_0      Available  Virtual Target Device - Disk
IBMi61_1      Available  Virtual Target Device - Disk
aix53_rvg     Available  Virtual Target Device - Disk
aix61_rvg     Available  Virtual Target Device - Disk
rhe152        Available  Virtual Target Device - Disk
sles10        Available  Virtual Target Device - Disk
vtopt0        Defined    Virtual Target Device - File-backed Optical
vtopt1        Defined    Virtual Target Device - File-backed Optical
vtopt2        Defined    Virtual Target Device - File-backed Optical
vtopt3        Available  Virtual Target Device - Optical Media
vtscsi0       Defined    Virtual Target Device - Disk
vtscsi1       Defined    Virtual Target Device - Logical Volume
ent5          Available  Shared Ethernet Adapter
ent7          Available  Shared Ethernet Adapter

$ lsmap -all
SVSA          Physloc          Client Partition ID
-----
vhost0        U9117.MMA.101F170-V1-C21      0x00000003

VTD           aix61_rvg
Status        Available
LUN           0x8100000000000000
Backing device hdisk7
Physloc       U789D.001.DQDYKYW-P1-C2-T2-W201300A0B811A662-L1000000000000

SVSA          Physloc          Client Partition ID
-----
vhost1        U9117.MMA.101F170-V1-C22      0x00000004

```

```

VTD          aix53_rvg
Status       Available
LUN          0x8100000000000000
Backing device hdisk8
Physloc      U789D.001.DQDYKYW-P1-C2-T2-W201300A0B811A662-L2000000000000

SVSA          Physloc          Client Partition ID
-----
vhost2       U9117.MMA.101F170-V1-C23     0x00000005

VTD          IBMi61_0
Status       Available
LUN          0x8100000000000000
Backing device hdisk11
Physloc      U789D.001.DQDYKYW-P1-C2-T2-W201300A0B811A662-L5000000000000

VTD          IBMi61_1
Status       Available
LUN          0x8200000000000000
Backing device hdisk12
Physloc      U789D.001.DQDYKYW-P1-C2-T2-W201300A0B811A662-L6000000000000

SVSA          Physloc          Client Partition ID
-----
vhost3       U9117.MMA.101F170-V1-C24     0x00000006

VTD          rhe152
Status       Available
LUN          0x8100000000000000
Backing device hdisk10
Physloc      U789D.001.DQDYKYW-P1-C2-T2-W201300A0B811A662-L4000000000000

SVSA          Physloc          Client Partition ID
-----
vhost4       U9117.MMA.101F170-V1-C25     0x00000000

VTD          sles10
Status       Available
LUN          0x8100000000000000
Backing device hdisk9
Physloc      U789D.001.DQDYKYW-P1-C2-T2-W201300A0B811A662-L3000000000000

SVSA          Physloc          Client Partition ID
-----
vhost5       U9117.MMA.101F170-V1-C50     0x00000003

VTD          vtopt3
Status       Available
LUN          0x8100000000000000
Backing device cd0
Physloc      U789D.001.DQDYKYW-P4-D1

SVSA          Physloc          Client Partition ID

```



```

-----
vhost6          U9117.MMA.101F170-V1-C60          0x00000003

VTD              NO VIRTUAL TARGET DEVICE FOUND

$ lsmmap -net -all
SVEA  Physloc
-----
ent2  U9117.MMA.101F170-V1-C11-T1

SEA          ent5
Backing device  ent0
Status       Available
Physloc      U789D.001.DQDYKYW-P1-C4-T1

```

If you restore to a different logical partition where you have defined similar virtual devices from the HMC recovery step provided previously, you will find that there are no linking devices.

**Note:** The devices will always be different between machines because the machine serial number is part of the virtual device location code for virtual devices. For example:

```

$ lsdev -dev ent4 -vpd
ent4          U8204.E8A.10FE411-V1-C11-T1  Virtual I/O Ethernet Adapter
(1-lan)

Network Address.....C21E4467D40B
Displayable Message.....Virtual I/O Ethernet Adapter (1-lan)
Hardware Location Code.....U8204.E8A.10FE411-V1-C11-T1

PLATFORM SPECIFIC

Name: 1-lan
Node: 1-lan@3000000b
Device Type: network
Physical Location: U8204.E8A.10FE411-V1-C11-T1

```

This is because the backing devices are not present for the linking to occur; the physical location codes have changed, and so the mapping fails. Example 5-7 shows the same restore of the Virtual I/O Server originally running on a Power 570 onto a Power 550 that has the same virtual devices defined in the same slots.

*Example 5-7 Devices recovered if restored to a different server*

```

$ lsdev -virtual
name          status          description

```

```

ent2          Available Virtual I/O Ethernet Adapter (1-lan)
vhost0       Available Virtual SCSI Server Adapter
vsa0         Available LPAR Virtual Serial Adapter

$ lsmap -all -net
SVEA Physloc
-----
ent4  U8204.E8A.10FE411-V1-C11-T1

SEA          ent6
Backing device ent0
Status      Available
Physloc     U78A0.001.DNWGCV7-P1-C5-T1

$ lsmap -all
SVSA          Physloc          Client Partition ID
-----
vhost0       U9117.MMA.101F170-V1-C10    0x00000003

VTD          NO VIRTUAL TARGET DEVICE FOUND

```

You now need to recover the user-defined virtual devices and any backing disk structure.

## Restoring with IBM Tivoli Storage Manager

You can use the IBM Tivoli Storage Manager to restore the mksysb image of the Virtual I/O Server.

**Note:** The IBM Tivoli Storage Manager can only restore the Virtual I/O Server to the system from which it was backed up.

First, you restore the mksysb image of the Virtual I/O Server using the **dsmc** command on the Tivoli Storage Manager client. Restoring the mksysb image does not restore the Virtual I/O Server. You then need to transfer the mksysb image to another system and convert the mksysb image to an installable format.

Before you start, complete the following tasks:

1. Ensure that the system to which you plan to transfer the mksysb image is running AIX.
2. Ensure that the system running AIX has a DVD-RW or CD-RW drive.
3. Ensure that AIX has the cdrecord and mkisofs RPMs downloaded and installed. To download and install the RPMs, see the AIX Toolbox for Linux Applications Web site at:

<http://www.ibm.com/systems/p/os/aix/linux>

**Note:** Interactive mode is not supported on the Virtual I/O Server. You can view session information by typing the **dsmc** command on the Virtual I/O Server command line.

To restore the Virtual I/O Server using Tivoli Storage Manager, complete the following tasks:

1. Determine which file you want to restore by running the **dsmc** command to display the files that have been backed up to the Tivoli Storage Manager server:

```
dsmc -query
```

2. Restore the mksysb image using the **dsmc** command. For example:

```
dsmc -restore sourcefilespec
```

Where *sourcefilespec* is the directory path to the location where you want to restore the mksysb image. For example, */home/padmin/mksysb\_image*.

3. Transfer the mksysb image to a server with a DVD-RW or CD-RW drive by running the following File Transfer Protocol (FTP) commands:
  - a. Run the following command to make sure that the FTP server is started on the Virtual I/O Server:

```
startnetsvc ftp
```
  - b. Open an FTP session to the server with the DVD-RW or CD-RW drive:

```
ftp server_hostname
```

where *server\_hostname* is the hostname of the server with the DVD-RW or CD-RW drive.
  - c. At the FTP prompt, change the installation directory to the directory where you want to save the mksysb image.
  - d. Set the transfer mode to binary, running the **binary** command.
  - e. Turn off interactive prompting using the **prompt** command.
  - f. Transfer the mksysb image to the server. Run the **mput mksysb\_image** command.
  - g. Close the FTP session after transferring the mksysb image by typing the **quit** command.
4. Write the mksysb image to CD or DVD using the **mkcd** or **mkdvd** commands.

Reinstall the Virtual I/O Server using the CD or DVD that you just created. For instructions, see chapter , “Restoring from DVD backup” on page 206. Or reinstall the Virtual I/O server from a NIM server. For more information, refer to , “Restoring from a file with the NIM server” on page 209.

## 5.5.4 Recovering user-defined virtual devices and disk structure

On our original Virtual I/O Server partition, we used two additional disks in a non-rootvg volume group. If these were SAN disks or physical disks that were directly mapped to client partitions, we could just restore the virtual device links. However, if we had a logical volume or storage pool structure on the disks, we need to restore this structure first. To do this, you need to use the volume group data files.

The volume group or storage pool data files should have been saved as part of the backup process earlier. These files should be located in the `/home/ios/vgbackups` directory if you performed a full backup using the `savevgstruct` command. The following command lists all of the available backups:

```
$ restorevgstruct -ls
total 104
-rw-r--r--  1 root    staff      51200 Oct 21 14:22 extra_storage.data
```

The `restorevgstruct` command restores the volume group structure onto the empty disks. In Example 5-8, there are some new blank disks and the same storage01 and datavg volume groups to restore.

### *Example 5-8 Disks and volume groups to restore*

---

```
$ lspv
NAME          PVID          VG          STATUS
hdisk0        00c1f170d7a97dec  old_rootvg
hdisk1        00c1f170e170ae72  clientvg    active
hdisk2        00c1f170e170c9cd  clientvg    active
hdisk3        00c1f170e170dac6  None
hdisk4        00c1f17093dc5a63  None
hdisk5        00c1f170e170fbb2  None
hdisk6        00c1f170de94e6ed  rootvg      active
hdisk7        00c1f170e327afa7  None
hdisk8        00c1f170e3716441  None
hdisk9        none          None
hdisk10       none          None
hdisk11       none          None
hdisk12       none          None
hdisk13       none          None
hdisk14       none          None
hdisk15       00c1f17020d9bee9  None
```

```
$ restorevgstruct -vg extra_storage hdisk15
hdisk15
extra_storage
testlv
```

```
Will create the Volume Group:  extra_storage
Target Disks:  Allocation Policy:
```

Shrink Filesystems: no  
Preserve Physical Partitions for each Logical Volume: no

---

After you restore all of the logical volume structures, the only remaining step is to restore the virtual devices linking the physical backing device to the virtual. To restore these, use the **lsmmap** outputs recorded from the backup steps in 5.4.4, “Backing up user-defined virtual devices” on page 199, or build documentation. As previously noted, it is important to use the slot numbers and backing devices when restoring these links.

The restoration of the Shared Ethernet Adapters will need the linking of the correct virtual Ethernet adapter to the correct physical adapter. Usually, the physical adapters are placed into a VLAN in the network infrastructure of the organization. It is important that the correct virtual VLAN is linked to the correct physical VLAN. Any network support team or switch configuration data can help with this task.

The disaster recovery restore involves a bit more manual recreating of virtual linking devices (vtscsi and SEA) and relies on good user documentation. If there is no multipath setup on the Virtual I/O Server to preserve, another solution is a completely new installation of the Virtual I/O Server from the installation media and then restore from the build documentation.

After running the **mkvdev** commands to recreate the mappings, the Virtual I/O Server will host virtual disks and networks that can be used to rebuild the AIX, IBM i or Linux clients.

### 5.5.5 Restoring the Virtual I/O Server client operating system

After you have the Virtual I/O Server operational and all of the devices recreated, you are ready to start restoring any AIX, IBM i or Linux clients. The procedure for this should already be defined in your organization and, most likely, will be identical to that for any server using dedicated disk and network resources. The method depends on the solution employed and should be defined by you.

For AIX clients, this information is available in the IBM Systems Information Center at:

<http://publib.boulder.ibm.com/infocenter/pseries/v5r3/index.jsp?topic=com.ibm.aix.baseadm/doc/baseadmndita/backmeth.htm>

For IBM i clients information about system backup and recovery is available in the IBM Systems Information Center at:

<http://publib.boulder.ibm.com/infocenter/systems/scope/i5os/index.jsp?topic=/rzahg/rzahgbackup.htm>

## 5.6 Rebuilding the Virtual I/O Server

This section describes what to do if there are no valid backup devices or backup images. In this case, you must install a new Virtual I/O Server.

In the following discussion, we assume that the partition definitions of the Virtual I/O Server and of all clients on the HMC are still available. We describe how we rebuilt our configuration of network and SCSI configurations.

It is useful to generate a System Plan on the HMC as documentation of partition profiles, settings, slot numbers, and so on. Example 5-9 shows the command to create a System Plan for a managed system. Note that the file name must have the extension \*.sysplan.

*Example 5-9 Creating an HMC system plan from the HMC command line*

---

```
hscroot@hmc1:~> mksysplan -f p570.sysplan -m MT_B_p570_MMA_101F170
```

---

To view the System Plan, select **System Plans**, select the System Plan that you want to see and then **View System Plan**. A browser window is opened where you are prompted for the user name and password of the HMC. Figure 5-12 shows a System Plan generated from a managed system.

The screenshot displays the Hardware Management Console interface. The main content area shows the following sections:

- About:** System Plan: latest\_base.sysplan; Description: System plan created from MT\_B\_p570\_MMA\_101F170; Application: HMC; Version: V7R3.4.0.0; Date: Saturday, October 18, 2008 2:30:37 PM CDT.
- Systems:** System: MT\_B\_p570\_MMA\_101F170; Description: 9117-MMA\*101F170; Memory: 32768 MB; Active Processors: 4.0; Auto Start: no; Quantity: 1; Memory Region Size: 128; Total Processors: 4.
- Shared Processor Pools:** A table with columns ID, Name, Reserved, and Maximum. Row 0: DefaultPool, +, \*.
- Hardware:** System Unit: U789D.001.DQDYKYW; Serial Number: DQDYKYW; Name: U789D.001.DQDYKYW; Order Status: Own.
- Ethernet Port:** A table with columns Backplane, Slot, Port Number, Logical Location Code, MAC Address, Connection Speed, Duplex, Maximum Receiving Packet Size, Flow Control, HEA Enabled, HEA Physical Port, and Used by Partition / Profile. Note: This table contains no data.
- Fibre Channel Port:** A table with columns Backplane, Slot, Port Number, Worldwide Node Name, and Worldwide Port Name.
 

Backplane	Slot	Port Number	Worldwide Node Name	Worldwide Port Name
P1	C1	T1	20000000c974a474	10000000c974a474
P1	C1	T2	20000000c974a475	10000000c974a475
P1	C2	T1	20000000c95db102	10000000c95db102
P1	C2	T2	20000000c95db101	10000000c95db101
P1	C3	T1	20000000c9676bb6	10000000c9676bb6
P1	C3	T2	20000000c9676bb7	10000000c9676bb7
- Cards:** A table with columns Backplane, Slot, Bus, Device Feature or CCIN, Device Description, Device Serial #, Order Status, and Used by Partition / Profile.
 

Backplane	Slot	Bus	Device Feature or CCIN	Device Description	Device Serial #	Order Status	Used by Partition / Profile
P1	C1	516		Fibre Channel Serial Bus		Own	
P1	C2	517	5774	Fibre Channel-2 PORT		Own	
P1	C3	518	5774	Fibre Channel-2 PORT		Own	
P1	C4	513	5706	PCI 10/100/1000Mbps Ethernet UTP 2-port		Own	
P1	C5	514	5706	PCI 10/100/1000Mbps Ethernet UTP 2-port		Own	

Figure 5-12 Example of a System Plan generated from a managed system

In addition to the regular backups using the **backupios** command, we recommend documenting the configuration of the following topics using the commands provided:

- Network settings

Commands:

```
netstat -state
netstat -routinfo
netstat -routtable
lsdev -dev Device -attr
```

```
cfgnamsrv -ls
hostmap -ls
optimizenet -list
entstat -all Device
```

- ▶ All physical and logical volumes, SCSI devices

Commands:

```
lspv
lsvg
lsvg -lv VolumeGroup
```

- ▶ All physical and logical adapters

Command:

```
lsdev -type adapter
```

- ▶ The mapping between physical and logical devices and virtual devices

Commands:

```
lsmap -all
lsmap -all -net
```

- ▶ Code levels, users and security

Commands:

```
ioslevel
viosecure -firewall view
viosecure -view -nonint
```

With this information, you can reconfigure your Virtual I/O Server manually. In the following sections, we describe the commands we needed to get the necessary information and the commands that rebuilt the configuration. The important information from the command outputs is highlighted. In your environment the commands may differ from those shown as examples.

To start rebuilding the Virtual I/O Server, you must know which disks are used for the Virtual I/O Server itself and for any assigned volume groups for virtual I/O.

The **lspv** command lists that the Virtual I/O Server was installed on hdisk0. The first step is to install the new Virtual I/O Server from the installation media onto disk hdisk0.

```
$ lspv
hdisk0      00c0f6a0f8a49cd7      rootvg      active
hdisk1      00c0f6a02c775268      None
hdisk2      00c0f6a04ab4fd01      None
hdisk3      00c0f6a04ab558cd      None
hdisk4      00c0f6a0682ef9e0      None
hdisk5      00c0f6a067b0a48c      None
hdisk6      00c0f6a04ab5995b      None
```



hdisk7	00c0f6a04ab66c3e	None
hdisk8	00c0f6a04ab671fa	None
hdisk9	00c0f6a04ab66fe6	None
hdisk10	00c0f6a0a241e88d	None
hdisk11	00c0f6a04ab67146	None
hdisk12	00c0f6a04ab671fa	None
hdisk13	00c0f6a04ab672aa	None
hdisk14	00c0f6a077ed3ce5	None
hdisk15	00c0f6a077ed5a83	None

See *PowerVM Virtualization on IBM System p: Introduction and Configuration*, SG24-7940 for the installation procedure. The further rebuild of the Virtual I/O Server is done in two steps:

1. Rebuild the SCSI configuration.
2. Rebuild the network configuration.

### 5.6.1 Rebuild the SCSI configuration

The `lspv` command also shows us that there is an additional volume group located on the Virtual I/O Server (datavg):

```
$ lspv
hdisk0      00c0f6a0f8a49cd7      rootvg      active
hdisk1      00c0f6a02c775268      None
hdisk2      00c0f6a04ab4fd01      None
hdisk3      00c0f6a04ab558cd      datavg      active
hdisk4      00c0f6a0682ef9e0      None
hdisk5      00c0f6a067b0a48c      None
hdisk6      00c0f6a04ab5995b      None
hdisk7      00c0f6a04ab66c3e      None
hdisk8      00c0f6a04ab671fa      None
hdisk9      00c0f6a04ab66fe6      None
hdisk10     00c0f6a0a241e88d      None
hdisk11     00c0f6a04ab67146      None
hdisk12     00c0f6a04ab671fa      None
hdisk13     00c0f6a04ab672aa      None
hdisk14     00c0f6a077ed3ce5      None
hdisk15     00c0f6a077ed5a83      None
```

The following command imports this information into the new Virtual I/O Server system's ODM:

```
importvg -vg datavg hdisk3
```

In Example 5-10 shows the mapping between the logical and physical volumes and the virtual SCSI server adapters.

Example 5-10 *lsmmap -all* command

```
$ lsmmap -all
SVSA          Physloc          Client Partition
ID
-----
vhost0        U9117.MMA.100F6A0-V1-C15  0x00000002

VTD           vcd
Status        Available
LUN           0x8100000000000000
Backing device cd0
Physloc       U789D.001.DQDWWHY-P4-D1

SVSA          Physloc          Client Partition
ID
-----
vhost1        U9117.MMA.100F6A0-V1-C20  0x00000002

VTD           vnim_rvg
Status        Available
LUN           0x8100000000000000
Backing device hdisk12
Physloc       U789D.001.DQDWWHY-P1-C2-T1-W200400A0B8110D0F-L1200000000000

VTD           vnimvg
Status        Available
LUN           0x8200000000000000
Backing device hdisk13
Physloc       U789D.001.DQDWWHY-P1-C2-T1-W200400A0B8110D0F-L1300000000000

SVSA          Physloc          Client Partition
ID
-----
vhost2        U9117.MMA.100F6A0-V1-C25  0x00000003

VTD           vdb_rvg
Status        Available
LUN           0x8100000000000000
Backing device hdisk8
Physloc       U789D.001.DQDWWHY-P1-C2-T1-W200400A0B8110D0F-LE000000000000

SVSA          Physloc          Client Partition
ID
-----
vhost3        U9117.MMA.100F6A0-V1-C40  0x00000004
```

```

VTD                vapps_rvg
Status             Available
LUN               0x8100000000000000
Backing device    hdisk6
Physloc
U789D.001.DQDWWHY-P1-C2-T1-W200400A0B8110D0F-LC000000000000

SVSA              Physloc                Client Partition
ID
-----
vhost4           U9117.MMA.100F6A0-V1-C50             0x00000005

VTD                vlnx_rvg
Status             Available
LUN               0x8100000000000000
Backing device    hdisk10
Physloc
U789D.001.DQDWWHY-P1-C2-T1-W200400A0B8110D0F-L1000000000000

```

Virtual SCSI server adapter vhost0 (defined on slot 15 in HMC) has one Virtual Target Device vcd. It is mapping the optical device cd0 to vhost0.

Virtual SCSI server adapter vhost1 (defined on slot 20 in HMC) has two Virtual Target Devices, vnim\_rvg and vnimvg. They are mapping the physical volumes hdisk12 and hdisk13 to vhost1.

Virtual SCSI server adapter vhost2 (defined on slot 25 in HMC) has vdb\_rvg as a Virtual Target Device. It is mapping the physical volume hdisk8 to vhost2.

Virtual SCSI server adapter vhost3 (defined on slot 40 in HMC) has vapps\_rvg as a Virtual Target Device. It is mapping the physical volume hdisk6 to vhost3.

Virtual SCSI server adapter vhost4 (defined on slot 50 in HMC) has vlnx\_rvg as a Virtual Target Device. It is mapping the physical volume hdisk10 to vhost4.

The following commands are used to create our needed Virtual Target Devices:

```

mkvdev -vdev cd0 -vadapter vhost0 -dev vcd
mkvdev -vdev hdisk12 -vadapter vhost1 -dev vnim_rvg
mkvdev -vdev hdisk13 -vadapter vhost1 -dev vnimvg
mkvdev -vdev hdisk8 -vadapter vhost2 -dev vdb_rvg
mkvdev -vdev hdisk6 -vadapter vhost3 -dev vnim_rvg
mkvdev -vdev hdisk10 -vadapter vhost4 -dev vlnx_rvg

```

**Note:** The names of the Virtual Target Devices are generated automatically, except when you define a name using the -dev flag of the `mkvdev` command.

## 5.6.2 Rebuild network configuration

After successfully rebuilding the SCSI configuration, we now are going to rebuild the network configuration.

The `netstat -state` command shows us that `en4` is the only active network adapter:

```
$ netstat -state
Name Mtu Network Address ZoneID Ipkts Ierrs Opkts Oerrs Coll
en4 1500 link#2 6a.88.8d.e7.80.d 4557344 0 1862620 0 0
en4 1500 9.3.4 vios1 4557344 0 1862620 0 0
lo0 16896 link#1 4521 0 4634 0 0
lo0 16896 127 loopback 4521 0 4634 0 0
lo0 16896 ::1 0 4521 0 4634 0 0
```

With the `lsmap -all -net` command, we determine that `ent5` is defined as a Shared Ethernet Adapter mapping physical adapter `ent0` to virtual adapter `ent2`:

```
$ lsmap -all -net
SVEA Physloc
-----
ent2 U9117.MMA.101F170-V1-C11-T1

SEA ent5
Backing device ent0
Status Available
Physloc U789D.001.DQDYKYW-P1-C4-T1

SVEA Physloc
-----
ent4 U9117.MMA.101F170-V1-C13-T1

SEA NO SHARED ETHERNET ADAPTER FOUND
```

The information for the default gateway address is provided by the `netstat -routinfo` command:

```
$ netstat -routinfo
Routing tables
Destination Gateway Flags Wt Policy If Cost
Config_Cost

Route Tree for Protocol Family 2 (Internet):
default 9.3.4.1 UG 1 - en4 0 0
9.3.4.0 vios1 UHSb 1 - en4 0 0 =>
9.3.4/23 vios1 U 1 - en4 0 0
vios1 loopback UGHS 1 - lo0 0 0
9.3.5.255 vios1 UHSb 1 - en4 0 0
127/8 loopback U 1 - lo0 0 0
```

```
Route Tree for Protocol Family 24 (Internet v6):
::1          ::1          UH          1    -    1o0      0    0
```

To list the subnet mask, we use the `lsdev -dev en4 -attr` command:

```
$ lsdev -dev en4 -attr
attribute      value      description
user_settable

alias4                IPv4 Alias including Subnet Mask      True
alias6                IPv6 Alias including Prefix Length    True
arp                   on      Address Resolution Protocol (ARP)     True
authority             Authorized Users                       True
broadcast             Broadcast Address                      True
mtu                   1500   Maximum IP Packet Size for This Device True
netaddr               9.3.5.111 Internet Address                      True
netaddr6              IPv6 Internet Address                 True
netmask               255.255.254.0 Subnet Mask                       True
prefixlen             Prefix Length for IPv6 Internet Address True
remmtu                576    Maximum IP Packet Size for REMOTE Networks True
rfc1323               Enable/Disable TCP RFC 1323 Window Scaling True
security              none    Security Level                        True
state                 up      Current Interface Status              True
tcp_mssdfmt          Set TCP Maximum Segment Size          True
tcp_nodelay           Enable/Disable TCP_NODELAY Option     True
tcp_recvspace        Set Socket Buffer Space for Receiving  True
tcp_sendspace        Set Socket Buffer Space for Sending    True
```

The last information we need is the default virtual adapter and the default PVID for the Shared Ethernet Adapter. This is shown by the `lsdev -dev ent5 -attr` command:

```
$ lsdev -dev ent5 -attr
attribute      value      description
user_settable

accounting        disabled  Enable per-client accounting of network statistics      True
ctl_chan          ent3     Control Channel adapter for SEA failover                True
gvrp              no       Enable GARP VLAN Registration Protocol (GVRP)           True
ha_mode           auto     High Availability Mode                                  True
jumbo_frames      no       Enable Gigabit Ethernet Jumbo Frames                   True
large_receive     no       Enable receive TCP segment aggregation                  True
largesend         0        Enable Hardware Transmit TCP Resegmentation            True
netaddr           0        Address to ping                                         True
pvid              1        PVID to use for the SEA device                          True
pvid_adapter      ent2     Default virtual adapter to use for non-VLAN-tagged packets True
qos_mode          disabled  N/A                                                      True
real_adapter      ent0     Physical adapter associated with the SEA                 True
thread            1        Thread mode enabled (1) or disabled (0)                True
virt_adapters     ent2     List of virtual adapters associated with the SEA (comma separated) True
```

**Note:** In this example the IP of the Virtual I/O Server is not configured on the Shared Ethernet Adapter (ent5) but on another adapter (ent4), thus avoiding network disruption between the Virtual I/O Server and any other partition on the same system when the replacement of the physical card (ent0) used as the Shared Ethernet Adapter is necessary.

The following commands recreated our network configuration:

```
$ mkvdev -sea ent0 -vadapter ent2 -default ent2 -defaultid 1
$ mktcpip -hostname vios1 -inetaddr 9.3.5.111 -interface en5 -start -netmask
255.255.254.0 -gateway 9.3.4.1
```

These steps complete the basic rebuilding of the Virtual I/O Server.

## 5.7 Updating the Virtual I/O Server

Two scenarios for updating a Virtual I/O Server are described in this section. A dual Virtual I/O Server environment to perform regular service is recommended in order to provide a continuous connection of your clients to their Virtual I/O resources.

For clients using non-critical virtual resources, or when you have service windows that allow a Virtual I/O Server to be rebooted, you can use a single Virtual I/O Server scenario. For the dual Virtual I/O Server scenario if you are using SAN LUNs and MPIO or IBM i mirroring on the clients, the maintenance on the Virtual I/O Server will not cause additional work after the update on the client side.

### 5.7.1 Updating a single Virtual I/O Server environment

When applying routine service that requires a reboot in a single Virtual I/O Server environment, you need to plan downtime and shut down every client partition using virtual storage provided by this Virtual I/O Server.

**Tip:** Back up the Virtual I/O Servers and the virtual I/O clients if a current backup is not available, and document the virtual Ethernet and SCSI devices before the update.

To avoid complications during an upgrade or update, we advise that you check the environment before upgrading or updating the Virtual I/O Server. The

following list is a sample of useful commands for the virtual I/O client and Virtual I/O Server:

<b>lsvg rootvg</b>	On the Virtual I/O Server and AIX virtual I/O client, check for stale PPs and stale PV.
<b>cat /proc/mdstat</b>	On the Linux client using mirroring, check for faulty disks.
<b>multipath -ll</b>	On the Linux client using MPIO, check the paths.
<b>lsvg -pv rootvg</b>	On the Virtual I/O Server, check for missing disks.
<b>netstat -cdlistats</b>	On the Virtual I/O Server, check that the Link status is Up on all used interfaces.
<b>errpt</b>	On the AIX virtual I/O client, check for CPU, memory, disk, or Ethernet errors, and resolve them before continuing.
<b>dmesg, messages</b>	On the Linux virtual I/O client, check for CPU, memory, disk, or Ethernet errors, and resolve them before continuing.
<b>netstat -v</b>	On the virtual I/O client, check that the Link status is Up on all used interfaces.

Before starting an upgrade, take a backup of the Virtual I/O Server and the virtual I/O clients if a current backup is not available. To back up the Virtual I/O Server, use the **backupios** command.

## Running update on a single Virtual I/O Server

There are several options for downloading and installing a Virtual I/O Server update: download iso-images, packages, or install from CD.

**Note:** You can get the latest available updates for the Virtual I/O Server and check also the recent installation instructions from the following web site:

<http://www14.software.ibm.com/webapp/set2/sas/f/vios/download/home.html>

To update the Virtual I/O Server follow the next steps.

1. Shut down the virtual I/O clients connected to the Virtual I/O Server, or disable any virtual resource that is in use.
2. All Interim Fixes applied before the upgrade should be removed.
  - a. Become root on Virtual I/O Server:

```
$ oem_setup_env
```
  - b. List all Interim Fixes installed:

```
# emgr -P
```

c. Remove each Interim Fix by label:

```
# emgr -r -L <label name>
```

d. Exit the root shell:

```
# exit
```

3. If previous updates have been applied to the Virtual I/O Server, you have to commit those with this command:

```
# updateios -commit
```

This command does not provide any progress information, but you can run:

```
$ tail -f install.log
```

In another terminal window, follow the progress. If the command hangs, just interrupt it with CTRL-c and run it again until you see the following output:

```
$ updateios -commit
```

```
There are no uncommitted updates.
```

4. Apply the update with the **updateios** command. Use /dev/cd0 for CD or any directory containing the files. You can also mount a NFS directory with the mount command:

```
$ mount <name_of_remote_server>:/software/AIX/VIO-Server /mnt
```

```
$ updateios -dev /mnt -install -accept
```

5. Reboot the Virtual I/O Server when the update has finished:

```
$ shutdown -restart
```

6. Verify the new level with the **ioslevel** command.
7. Check the configuration of all disks and Ethernet adapters on the Virtual I/O Server.
8. Start the client partitions.

Verify the Virtual I/O Server environment, document the update, and create a backup of your updated Virtual I/O Server.

## 5.7.2 Updating a dual Virtual I/O Server environment

When applying an update to the Virtual I/O Server in a properly configured dual Virtual I/O Server environment, you can do so without having downtime to the virtual I/O services and without any disruption in continuous availability.

**Tip:** Back up the Virtual I/O Servers and the virtual I/O clients if a current backup is not available, and document the virtual Ethernet and SCSI device before the update. This reduces the time used in a recovery scenario.



## Checking network health

It is best practice to check the virtual Ethernet and disk devices on the Virtual I/O Server and virtual I/O client before starting the update on either of the Virtual I/O Servers. Check the physical adapters to verify connections. As shown in Example 5-11, Figure 5-13 on page 230, Example 5-12 on page 230, and Example 5-13 on page 231, all the virtual adapters are up and running.

*Example 5-11 The netstat -v comand on the virtual I/O client*

---

```
netstat -v
.
. (Lines omitted for clarity)
.
Virtual I/O Ethernet Adapter (l-lan) Specific Statistics:
-----
RQ Length: 4481
No Copy Buffers: 0
Filter MCast Mode: False
Filters: 255
  Enabled: 1  Queued: 0  Overflow: 0
LAN State: Operational
Hypervisor Send Failures: 0
  Receiver Failures: 0
  Send Errors: 0
Hypervisor Receive Failures: 0

ILLAN Attributes: 0000000000003002 [0000000000002000]

. (Lines omitted for clarity)
```

---

```

Work with TCP/IP Interface Status
                                                    System:E101F170
Type options, press Enter.
  5=Display details  8=Display associated routes  9=Start  10=End
 12=Work with configuration status  14=Display multicast groups

      Internet      Network      Line      Interface
Opt  Address      Address      Description  Status
     9.3.5.119    9.3.4.0     ETH01       Active
     127.0.0.1   127.0.0.0   *LOOPBACK   Active

Bottom
F3=Exit  F9=Command line  F11=Display line information  F12=Cancel
F13=Sort by column  F20=Work with IPv6 interfaces  F24=More keys

```

Figure 5-13 IBM i Work with TCP/IP Interface Status screen

*Example 5-12 The netstat -cdlistats command on the primary Virtual I/O Server*

---

```

$ netstat -cdlistats
.
. (Lines omitted for clarity)
.
Virtual I/O Ethernet Adapter (1-lan) Specific Statistics:
-----
RQ Length: 4481
No Copy Buffers: 0
Trunk Adapter: True
Priority: 1 Active: True
Filter MCast Mode: False
Filters: 255
  Enabled: 1 Queued: 0 Overflow: 0
LAN State: Operational
.
. (Lines omitted for clarity)

```

---

*Example 5-13 The netstat -cdlistats command on the secondary Virtual I/O Server*

---

```
$ netstat -cdlistats
.
. (Lines omitted for clarity)
.
Virtual I/O Ethernet Adapter (l-lan) Specific Statistics:
-----
RQ Length: 4481
No Copy Buffers: 0
Trunk Adapter: True
Priority: 2 Active: False
Filter MCast Mode: False
Filters: 255
  Enabled: 1 Queued: 0 Overflow: 0
LAN State: Operational
.
. (Lines omitted for clarity)
```

---

## Checking storage health

Checking the disk status depends on how the disks are shared from the Virtual I/O Server.

### Checking the storage health in the MPIO environment

If you have an MPIO setup on your virtual I/O Server clients similar to Figure 5-14, run the following commands before and after the first Virtual I/O Server update to verify the disk path status:

- lspath** On the AIX virtual I/O client, check all the paths to the disks. They should all be in the enabled state.
- multipath -ll** Check the paths on the Linux client.
- lsattr -El hdisk0** On the virtual I/O client, check the MPIO heartbeat for hdisk0, that the attribute hcheck\_mode is set to nonactive, and that hcheck\_interval is 60. If you run IBM SAN storage, check that reserve\_policy is no\_reserve; other storage vendors might require other values for reserve\_policy. This command should be executed on all disks on the Virtual I/O Server.

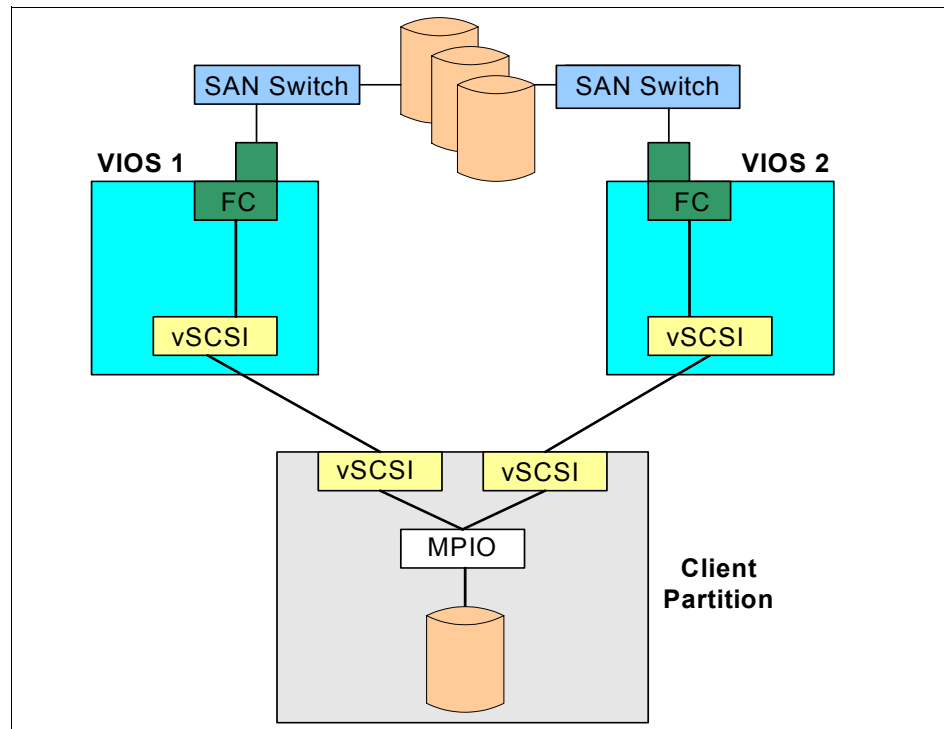


Figure 5-14 Virtual I/O client running MPIO

## Checking storage health in the mirroring environment

Figure 5-15 shows the concept of a mirrored infrastructure:

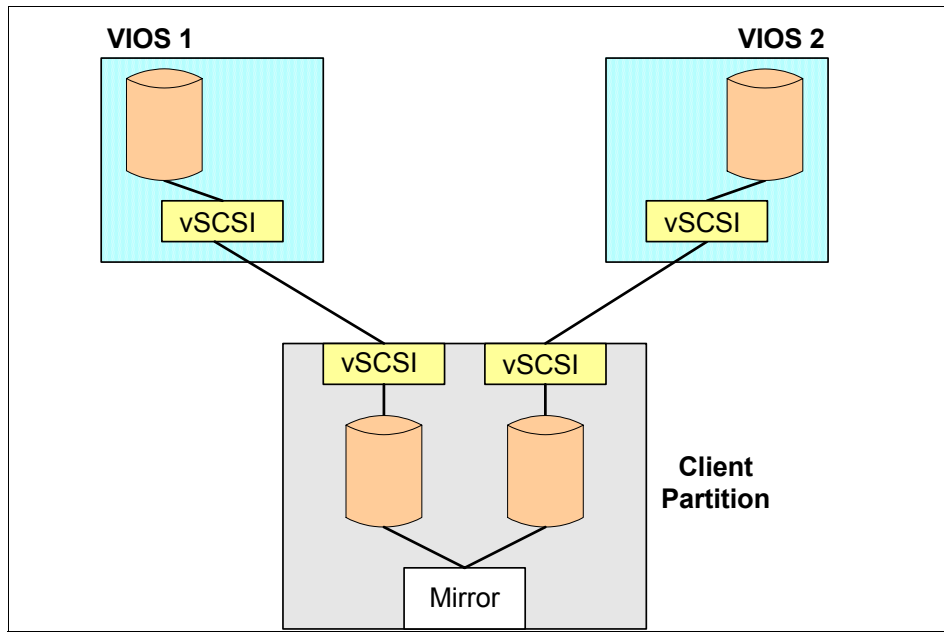


Figure 5-15 Virtual I/O client partition software mirroring

If you use mirroring on your virtual I/O clients, verify a healthy mirroring status for the disks shared from the Virtual I/O Server with the following procedures:

On the AIX virtual I/O client:

- lsvg rootvg**      Verify there are no stale PPs, and the quorum must be off.
- lsvg -p rootvg**    Verify there is no missing hdisk.

**Note:** The `fixdualvio.ksh` script in Appendix A, “Sample script for disk and NIB network checking and recovery on AIX virtual clients” on page 511 is a useful tool to do a health check.

On the IBM i virtual I/O client:

- ▶ Run **STRSST** and login to System Service Tools
- ▶ Select options **3. Work with disk units** → **1. Display disk configuration** → **1. Display disk configuration status** and verify all virtual disk units (type 6B22) are in mirrored *active* state as shown in Figure 5-16.

```

Display Disk Configuration Status

      Serial                Resource
ASP Unit Number          Type Model Name      Status
  1
    1 Y3WUTVVQMM4G      6B22 050 DD001    Active
    1 YYUUH3U9UELD      6B22 050 DD004    Active
    2 YD598QUY5XR8      6B22 050 DD003    Active
    2 YTM3C79KY4XF      6B22 050 DD002    Active

Press Enter to continue.

F3=Exit      F5=Refresh      F9=Display disk unit details
F11=Disk configuration capacity  F12=Cancel

```

Figure 5-16 IBM i Display Disk Configuration Status screen

On the Linux virtual I/O client:

**cat /proc/mdstat** Check the mirror status. See a healthy environment in Example 5-14.

Example 5-14 The mdstat command showing a healthy environment

---

```

cat /proc/mdstat
Personalities : [raid1]
md1 : active raid1 sdb3[1] sda3[0]
      1953728 blocks [2/2] [UU]
md2 : active raid1 sdb4[1] sda4[0]
      21794752 blocks [2/2] [UU]
md0 : active raid1 sdb2[1] sda2[0]
      98240 blocks [2/2] [UU]

```

---

After checking the environment and resolving any issues, back up the Virtual I/O Server and virtual I/O client if a current backup is not available.

## Step-by-step update

To update a dual Virtual I/O Server environment, do the following:

1. Find the standby Virtual I/O Server and run the **netstat** command. At the end of the output, locate the priority of the Shared Ethernet Adapter and whether it is active. In this case, the standby adapter is not active, so you can begin the upgrade of this server.

```
$ netstat -cdlistats
.
. (Lines omitted for clarity)
.
Trunk Adapter: True
Priority: 2 Active: False
Filter MCast Mode: False
Filters: 255
  Enabled: 1 Queued: 0 Overflow: 0
LAN State: Operational
.
. (Lines omitted for clarity)
```

If you have to change the active adapter, use following command to put it in backup mode manually:

```
$ chdev -attr entXX ha_mode=standby
```

2. All Interim Fixes applied before the upgrade should be removed.

- a. Become root on Virtual I/O Server:

```
$ oem_setup_env
```

- b. List all Interim Fixes installed:

```
# emgr -P
```

- c. Remove each Interim Fix by label:

```
# emgr -r -L <label name>
```

- d. Exit the root shell:

```
# exit
```

3. Apply the update from VD or a remote directory with the **updateios** command and press **y** to start the update.

```
$ updateios -dev /mnt -install -accept
. (Lines omitted for clarity)
```

```
Continue the installation [y|n]?
```

4. Reboot the standby Virtual I/O Server when the update completes:

```
$ shutdown -force -restart
```

SHUTDOWN PROGRAM  
Mon Oct 13 21:57:23 CDT 2008

Wait for 'Rebooting...' before stopping.  
Error reporting has stopped.

5. .After the reboot, verify the software level:

```
$ ioslevel  
1.5.2.1-FP-11.1
```

**Note:** At the time of writing there was no update available for Virtual I/O Server Version 2.1.

6. For an AIX MPIO environment, as shown in Figure 5-14 on page 232, run the **lspath** command on the virtual I/O client and verify that all paths are enabled. For an AIX LVM mirroring environment, as shown in Figure 5-14 on page 232, run the **varyonvg** command as shown in Example 5-15, and the volume group should begin to sync. If not, run the **syncvg -v <VGname>** command on the volume groups that used the virtual disk from the Virtual I/O Server environment to synchronize each volume group, where <VGname> is the name of the Volume Group.

For the IBM i client mirroring environment you can proceed to the next step. No manual action is required on IBM i client side as IBM i automatically resumes the suspended mirrored disk units as soon as the updated Virtual I/O Server is back operational.

**Note:** IBM i tracks changes for a suspended mirrored disk unit for a limited amount of time allowing it to resynchronize changed pages only. To our experience IBM i did not require a full mirror resynchronize when rebooting the Virtual I/O Server. But this may not be the case for any reboot taking an extended amount of time.

*Example 5-15 AIX LVM Mirror Resync*

```
# lsvg -p rootvg  
rootvg:  
PV_NAME          PV STATE          TOTAL PPs   FREE PPs   FREE  
DISTRIBUTION  
hdisk0           active           511         488  
102..94..88..102..102  
hdisk1           missing          511         488  
102..94..88..102..102  
  
# varyonvg rootvg
```



```

# lsvg -p rootvg
rootvg:
PV_NAME          PV STATE          TOTAL PPs   FREE PPs   FREE
DISTRIBUTION
hdisk0           active           511         488
102..94..88..102..102
hdisk1           active         511         488
102..94..88..102..102

# lsvg rootvg
VOLUME GROUP:    rootvg                VG IDENTIFIER:
00c478de00004c0000
00006b8b6c15e
VG STATE:        active                PP SIZE:      64 megabyte(s)
VG PERMISSION:   read/write            TOTAL PPs:    1022 (65408
megabytes)
MAX LVs:         256                  FREE PPs:     976 (62464
megabytes)
LVs:             9                    USED PPs:     46 (2944
megabytes)
OPEN LVs:        8                    QUORUM:       1
TOTAL PVs:       2                    VG DESCRIPTORS: 3
STALE PVs:    0                STALE PPs:    0
ACTIVE PVs:      2                    AUTO ON:      yes
MAX PPs per VG:  32512
MAX PPs per PV:  1016                  MAX PVs:      32
LTG size (Dynamic): 256 kilobyte(s)  AUTO SYNC:    no
HOT SPARE:       no                    BB POLICY:    relocatable
#

```

For a Linux client mirroring environment follow these steps for every md-device (md0, md1, md2):

a. Set the disk faulty (repeat the steps for all mdx devices):

```
# mdadm --manage --set-faulty /dev/md2 /dev/sda4
```

b. Remove the device:

```
# mdadm --manage --remove /dev/md2 /dev/sda2
```

c. Rescan the device (choose the corresponding path):

```
# echo 1 > /sys/class/scsi_device/0\:0\:1\:0/device/rescan
```

d. Hot-add the device to mdadm:

```
# mdadm --manage --add /dev/md2 /dev/sda4
```

e. Check the sync status; wait for it to be finished:

```
# cat /proc/mdstat
Personalities : [raid1]
md1 : active raid1 sda3[0] sdb3[1]
```

```
1953728 blocks [2/2] [UU]
```

```
md2 : active raid1 sda4[2] sdb4[1]
      21794752 blocks [2/1] [_U]
      [=>.....] recovery = 5.8% (1285600/21794752)
      finish=8.2min speed=41470K/sec
md0 : active raid1 sda2[0] sdb2[1]
      98240 blocks [2/2] [UU]
```

7. If you use Shared Ethernet Adapter Failover, shift the standby and primary connections to the Virtual I/O Server with the **chdev** command and check with the **netstat -cdlistats** command whether the state has changed, as shown in this example:

```
$ chdev -dev ent4 -attr ha_mode=standby
ent4 changed
$ netstat -cdlistats
.
. (Lines omitted for clarity)
.
Trunk Adapter: True
Priority: 1 Active: False
Filter MCast Mode: False
.
. (Lines omitted for clarity)
```

After verifying the network and storage health again on all Virtual I/O Server active client partitions proceed as outlined below to similarly update the other Virtual I/O Server as well:

8. Remove all interim fixes on the second Virtual I/O Server to be updated.
9. Apply the update to the second Virtual I/O Server which now is the standby Virtual I/O Server, using the **updateios** command.
10. Reboot the second Virtual I/O Server with the **shutdown -restart** command.
11. Check the new level with the **ioslevel** command.
12. For an AIX MPIO environment as shown in Figure 5-14 on page 232, run the **lspath** command on the virtual I/O client and verify that all paths are enabled. For an AIX LVM environment, as shown in Figure 5-15 on page 233, run the **varyonvg** command, and the volume group should begin to synchronize. If not, use the **syncvg -v <VGname>** command on the volume groups that used the virtual disk from the Virtual I/O Server environment to synchronize each volume group, where <VGname> is the name of the Volume Group.

For the IBM i client mirroring environment you can proceed to the next step. No manual action is required on IBM i client side as IBM i automatically resumes the suspended mirrored disk units as soon as the updated Virtual I/O Server is back operational.

For the Linux mirroring environment manually resynchronize the mirror again (refer to step 6 above).

13. If you use Shared Ethernet Adapter Failover reset the Virtual I/O Server role back to primary with the `chdev` command, as shown in the following example:

```
$ chdev -dev ent4 -attr ha_mode=auto
ent4 changed
$
```

14. After verifying the network and storage health again create another backup this time from both updated Virtual I/O Servers before considering the update process complete.

## 5.8 Error logging on the Virtual I/O Server

Error logging on the Virtual I/O Server uses the same error logging facility as AIX. The error logging daemon is started with the `errdaemon` command. This daemon reads error records from the `/dev/error` device and writes them to the error log in `/var/adm/ras/errlog`. Errdaemon also performs specified notifications in the notification database `/etc/objrepos/errnotify`.

The command to display binary error logs is `errlog`. See Example 5-16 for a short error listing.

*Example 5-16 errlog short listing*

---

```
$ errlog
IDENTIFIER  TIMESTAMP  T C RESOURCE_NAME  DESCRIPTION
4FC8E358   1015104608 I O hdisk8        CACHED DATA WILL BE LOST IF CONTROLLER
B6267342   1014145208 P H hdisk12       DISK OPERATION ERROR
DF63A4FE   1014145208 T S vhost2        Virtual SCSI Host Adapter detected an
B6267342   1014145208 P H hdisk12       DISK OPERATION ERROR
DF63A4FE   1014145208 T S vhost2        Virtual SCSI Host Adapter detected an
B6267342   1014145208 P H hdisk11       DISK OPERATION ERROR
B6267342   1014145208 P H hdisk11       DISK OPERATION ERROR
C972F43B   1014111208 T S vhost4        Misbehaved Virtual SCSI ClientB6267342
B6267342   1014164108 P H hdisk14       DISK OPERATION ERROR
```

---

In order to get all details listed on each event, `errlog -ls` can be used.

*Example 5-17 Detailed error listing*

---

```
$ errlog -ls |more
-----
LABEL:          SC_DISK_PCM_ERR7
IDENTIFIER:     4FC8E358
```

Date/Time: Wed Oct 15 10:46:33 CDT 2008  
Sequence Number: 576  
Machine Id: 00C1F1704C00  
Node Id: vios1  
Class: 0  
Type: INFO  
WPAR: Global  
Resource Name: hdisk8

Description  
CACHED DATA WILL BE LOST IF CONTROLLER FAILS

Probable Causes  
USER DISABLED CACHE MIRRORING FOR THIS LUN

User Causes  
CACHE MIRRORING DISABLED

Recommended Actions  
ENABLE CACHE MIRRORING

...

---

All errors are divided into classes, as shown in Table 5-3.

*Table 5-3 Error log entry classes*

Error log entry class	Description
H	Hardware error
S	Software error
O	Operator messages (logger)
U	Undetermined error class

## 5.8.1 Redirecting error logs to other servers

In some cases you may need to redirect error logs to one central instance, for example in order to be able to run automated error log analysis in one place. For the Virtual I/O Server you need to set up redirecting error logs to syslog first and then assign the remote syslog host in the syslog configuration.

In order to redirect error logs to syslog, create the file `/tmp/syslog.add` with the content shown in Example 5-18.

**Note:** You need to become root user first on the Virtual I/O server. Run the command:

```
$ oem_setup_env
```

*Example 5-18 Content of /tmp/syslog.add file*

---

```
errnotify:
en_pid = 0
en_name = "syslog"
en_persistenceflg = 1
en_method = "/usr/bin/errpt -a -l $1 |/usr/bin/fgrep -v 'ERROR_ID TIMESTAMP' |
/usr/bin/logger -t ERRDEMON -p local1.warn"
```

---

Now use the **odmadd** command to add the configuration to the ODM:

```
# odmadd /tmp/syslog.add
```

In the syslog file you can redirect all messages to any other server running **syslogd** and accepting remote logs—just add the following line to your `/etc/syslogd.conf` file:

```
*.debug @9.3.5.115
```

and restart your syslog daemon using the following command:

```
# stopsrc -s syslogd
0513-044 The syslogd Subsystem was requested to stop.
# startsrc -s syslogd
0513-059 The syslogd Subsystem has been started. Subsystem PID is 520236.
```

## 5.8.2 Troubleshooting error logs

If your error log gets corrupted for some reason, you can always move the file, and a new clean error log file will be created as shown in Example 5-19.

*Example 5-19 Create new error log file*

---

```
$ oem_setup_env
# /usr/lib/errstop
# mv /var/adm/ras/errlog /var/adm/ras/errlog.bak
# /usr/lib/errdemon
```

---

If you want to back up your error log to an alternate file and view it later, do as shown in Example 5-20.

*Example 5-20 Copy errlog and view it*

---

```
$ oem_setup_env
# cp /var/adm/ras/errlog /tmp/errlog.save
# errpt -i /tmp/errlog.save
IDENTIFIER  TIMESTAMP  T C RESOURCE_NAME  DESCRIPTION
4FC8E358    1015104608 I O hdisk8        CACHED DATA WILL BE LOST IF CONTROLLER
B6267342    1014145208 P H hdisk12        DISK OPERATION ERROR
DF63A4FE    1014145208 T S vhost2         Virtual SCSI Host Adapter detected an
B6267342    1014145208 P H hdisk12        DISK OPERATION ERROR
DF63A4FE    1014145208 T S vhost2         Virtual SCSI Host Adapter detected an
B6267342    1014145208 P H hdisk11        DISK OPERATION ERROR
B6267342    1014145208 P H hdisk11        DISK OPERATION ERROR
C972F43B    1014111208 T S vhost4         Misbehaved Virtual SCSI ClientB6267342
B6267342    1014164108 P H hdisk14        DISK OPERATION ERROR
```

---



# Dynamic operations

This chapter discusses how to set up a shared processor pool, and how to change resources dynamically, which may be useful when maintaining your virtualized environment. With this goal, the focus is on the following operations valid for AIX, IBM i and Linux operating systems:

- ▶ Addition of resources
- ▶ Movement of adapters between partitions
- ▶ Removal of resources
- ▶ Replacement of resources

## 6.1 Multiple Shared Processor Pools management

With the POWER6 systems, you can now define Multiple Shared Processor Pools (MSPP) and assign the shared partitions to any of these MSPPs. The configuration of this feature is rather simple. You only have to set the properties of a processor pool.

To set up a shared processor pool (SPP), follow the steps described below:

1. In the HMC navigation pane, open **Systems Management** and click **Servers**.
2. In the content pane, select the managed system whose shared processor pool you want to configure, click the Task button, and select **Configuration** → **Shared Processor Pool Management**. Figure 6-1 lists the available shared processor pools:

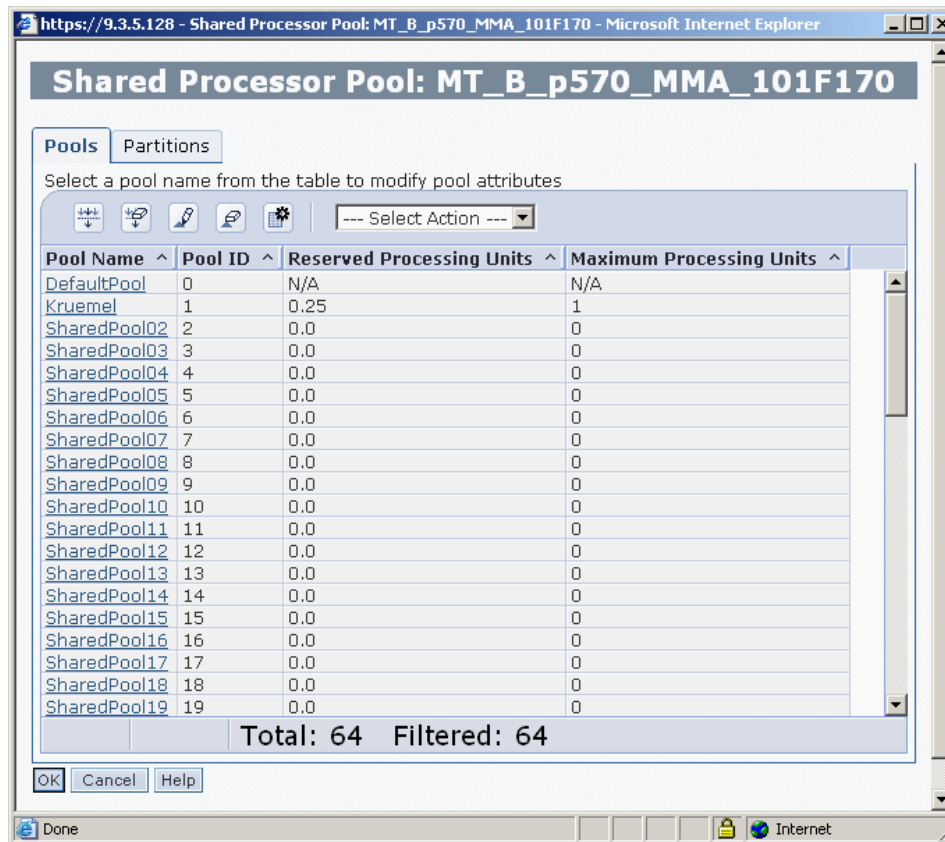


Figure 6-1 Shared Processor Pool



3. Click the name of the shared processor pool that you want to configure.
4. Enter the maximum number of processing units that you want the logical partitions in the shared processor pool to use in the Maximum processing units field. If desired, change the name of the shared processor pool in the Pool name field and enter the number of processing units that you want to reserve for uncapped logical partitions in the shared processor pool in the Reserved processing units field (Figure 6-2 on page 245). (The name of the shared processor pool must be unique on the managed system.) When you are done, click **OK**.

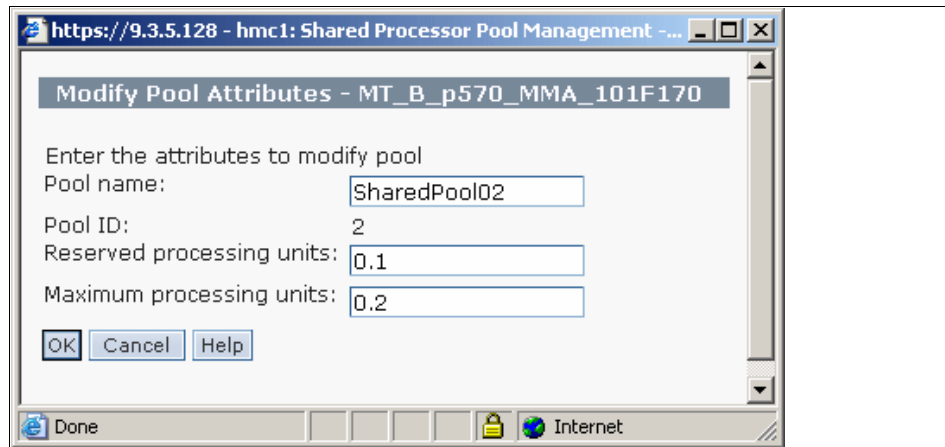


Figure 6-2 Modifying Shared Processor pool attributes

5. Repeat steps 3 and 4 for any other shared processor pools that you want to configure.
6. Click **OK**.

After this procedure (modifying the processor pool attributes) is complete, assign logical partitions to the configured shared processor pools. You can assign a logical partition to a shared processor pool at the time creating a logical partition, or you can reassign existing logical partitions from their current shared processor pools to the (new) shared processor pools that you configured using this procedure.

Click the Partitions tab and select the partition name as shown in Figure 6-3.

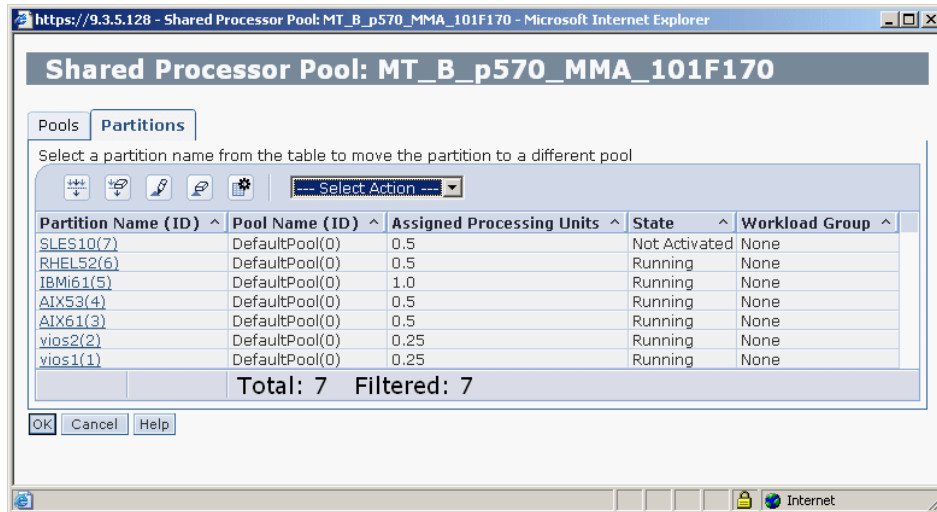


Figure 6-3 Partitions assignment to Multiple Shared Processor Pools

You then have to select to which SPP this partition should be assigned, as shown in Figure 6-4.

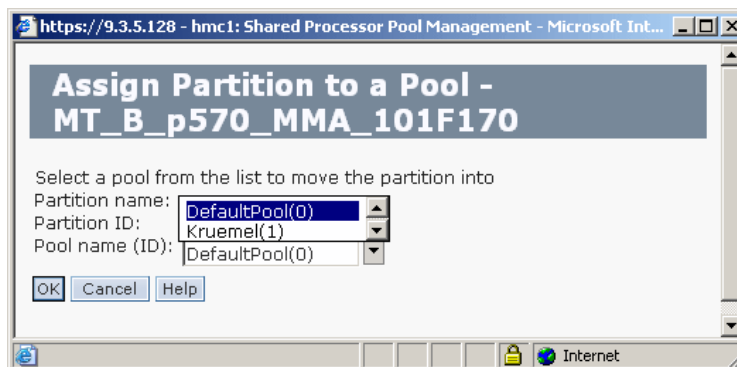


Figure 6-4 Assign a partition to a Shared Processor Pool

**Note:** The default Shared Processor Pool is the one with ID 0. This cannot be changed and it has some default configuration values that cannot be changed.

When you no longer want to use a Shared Processor Pool, you can deconfigure the shared processor pool by using this procedure to set the maximum number of processing units and reserved number of processing units to 0. Before you can

deconfigure a shared processor pool, you must reassign all logical partitions that use the shared processor pool to other shared processor pools.

### Calibrating the shared partitions' weight

You should pay attention to the values you provide for the partition weight when you define shared partition characteristics. Indeed, in case the partitions within an SPP need more processor resources, the extra resources that will be donated from the other idle partitions in the other SPPs are distributed to the partitions based on their weight. The partitions with the highest weight will gain more processor resources.

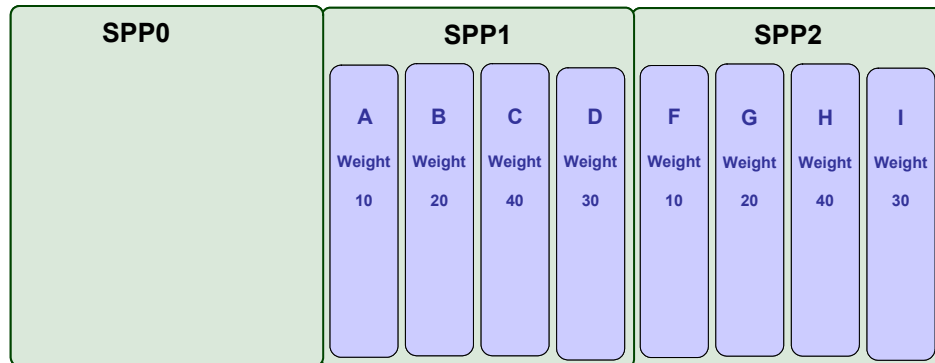


Figure 6-5 Comparing partition weights from different Shared Processor Pools

Considering the example shown in Figure 6-5, if the partitions C, D and H require extra processing resources, these extra resources will be distributed based on their weight value even though they are not all in the same SPP.

Based on the weight value shown in this example, partition D will get most of the available shared resources, partition C gets much lesser and partition H gets the least. In situations where your workload on partition H (or another partition) needs more system resources, set its weight value by taking into account the weight of the partitions in the other SPPs.

In summary, if several partitions from different SPPs compete to get additional resources, the partitions with the highest weight will be served first. You must therefore pay attention when you define a partition's weight and make sure that its value is reasonable compared to all of the other partitions in the different shared processor pools.

For more detailed information refer to *PowerVM Virtualization on IBM System p: Introduction and Configuration*, SG24-7940.

## 6.2 Dynamic LPAR operations on AIX and IBM i

In the next sections we explain how to perform dynamic LPAR operations for AIX and IBM i.

**Note:** For using IBM i 6.1 dynamic LPAR operations with *virtual* adapters make sure to have SLIC PTFs MF45568 and MF45473 applied.

**Important:** HMC communicates with partitions using RMC. So, you need to make sure RMC port has not been restricted in firewall settings. For more details refer to “Setting up the firewall” on page 144.

### 6.2.1 Adding and removing processors dynamically

The following steps explains the procedure to add or remove processors dynamically:

1. Select the logical partition where you want to initiate a dynamic LPAR operation, then select **Dynamic Logical Partitioning** → **Processor** → **Add or Remove**, as shown in Figure 6-6.

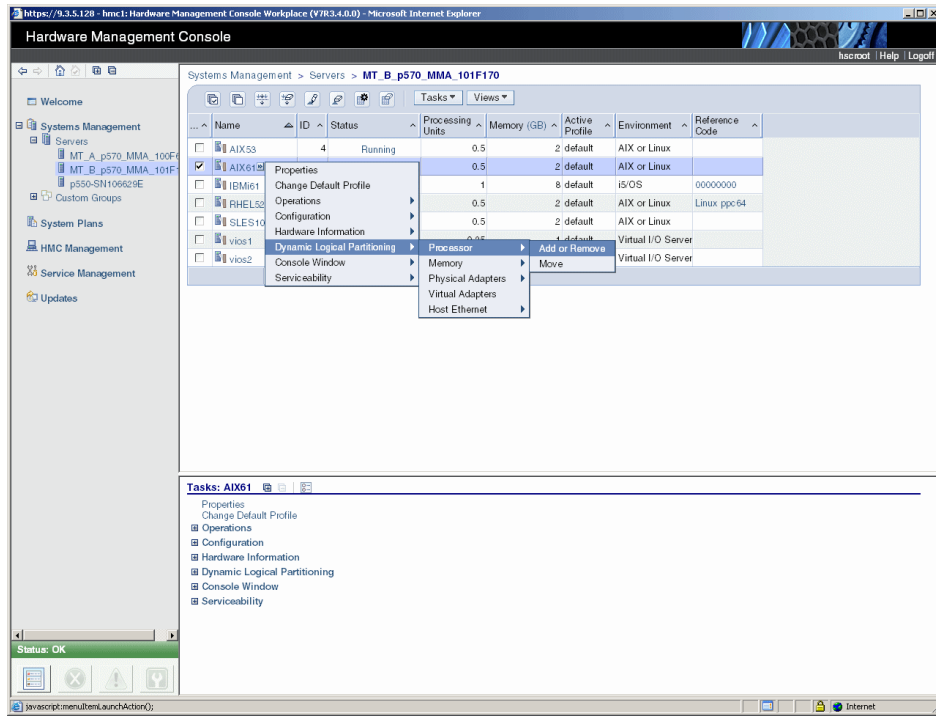


Figure 6-6 Add or remove processor operation

2. On HMC Version 7, you do not have to define a certain amount of CPU to be removed or added to the partition. You just have to inform the total amount of processor units to be assigned to the partition. You can change processing units and the virtual processors of the partition to more or less than the current value. The values for these fields have to be between the Minimum and Maximum value defined for them on the partition profile. Figure 6-7 shows a partition being set with 0.5 processing units and 1 virtual processors.

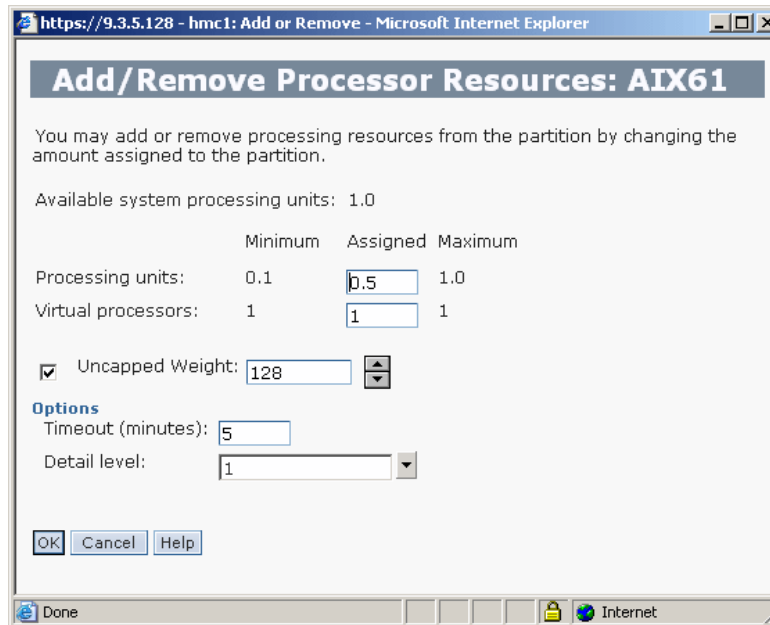


Figure 6-7 Defining the amount of CPU processing units for a partition

3. Click **OK** when done.

**Note:** In this example a partition using Micro-partition technology was given, but this process is also valid for dedicated partitions where you move dedicated processors.

From an IBM i partition the current virtual processors and processing capacity can be displayed using the **WRKSYSACT** command as shown in Figure 6-8.

```

Work with System Activity                E101F170
                                           10/17/08 14:22:16
Automatic refresh in seconds . . . . . 5
Job/Task CPU filter . . . . . . . . . . .10
Elapsed time . . . . . : 00:00:02   Average CPU util . . . . . : 16.7
Virtual Processors . . . . . : 2     Maximum CPU util . . . . . : 18.1
Overall DB CPU util . . . . . : .0     Minimum CPU util . . . . . : 15.2
                                           Current processing capacity: 1.50

Type options, press Enter.
  1=Monitor job   5=Work with job

      Job or
Opt  Task      User      Number  Thread  Pty   CPU   Total  Total  DB
      Task      User      Number  Thread  Pty   Util  Sync  Async  CPU
      QZRCRVS   QUSER    001780  00000001  20   15.8   71   110   .0
      QPADEV000C IDIMMER  001975  0000000A  1    .3    14    0    .0
      CRTPFRTA   QSYS     001825  00000001  50   .1    120   55   .0
      SMPOL001                   99   .0    0    10   .0
      QSPPF00001 QSYS     001739  00000001  15   .0    1    0    .0
      SMMIRRORMC                   0    .0    1    0    .0
                                           More...

F3=Exit  F10=Update list  F11=View 2  F12=Cancel  F19=Automatic refresh
F24=More keys
(C) COPYRIGHT IBM CORP. 1980, 2007.

```

Figure 6-8 IBM i Work with System Activity screen

## 6.2.2 Add memory dynamically

Follow these steps to dynamically add additional memory to the logical partition (see Figure 6-9):

1. Select the partition and then **Dynamic Logical Partitioning** → **Memory** → **Add or Remove**.

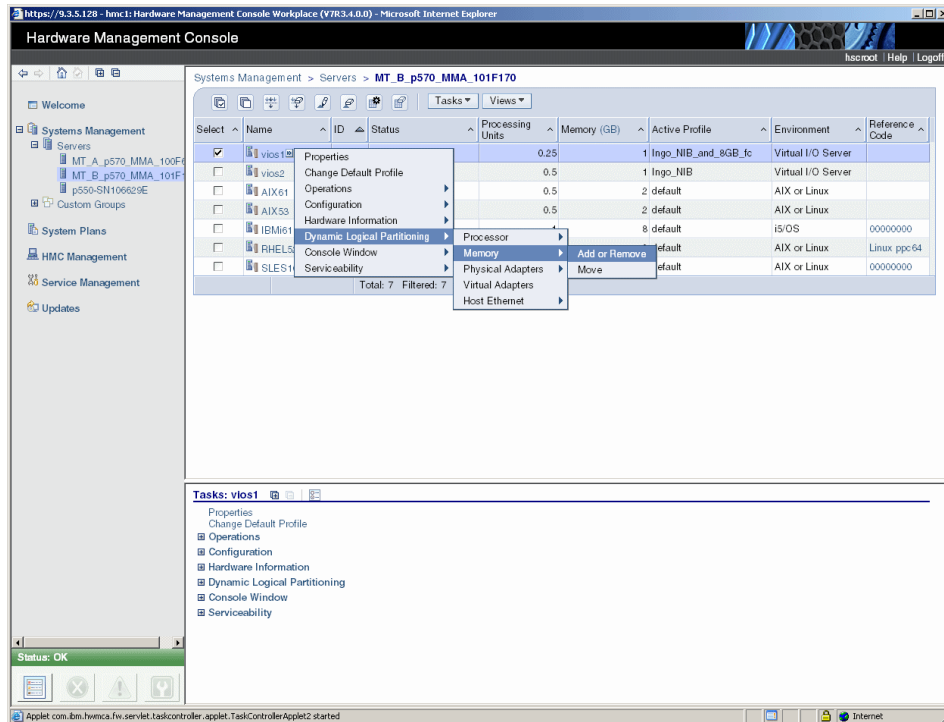


Figure 6-9 Add or remove memory operation



2. Change the total amount of memory to be assigned to the partition. Note that on HMC Version 7 you do not provide the amount of additional memory that you want to add to the partition, but the total amount of memory that you want to assign to the partition. In Figure 6-10 the total amount of memory allocated to the partition was changed to 5 GB.

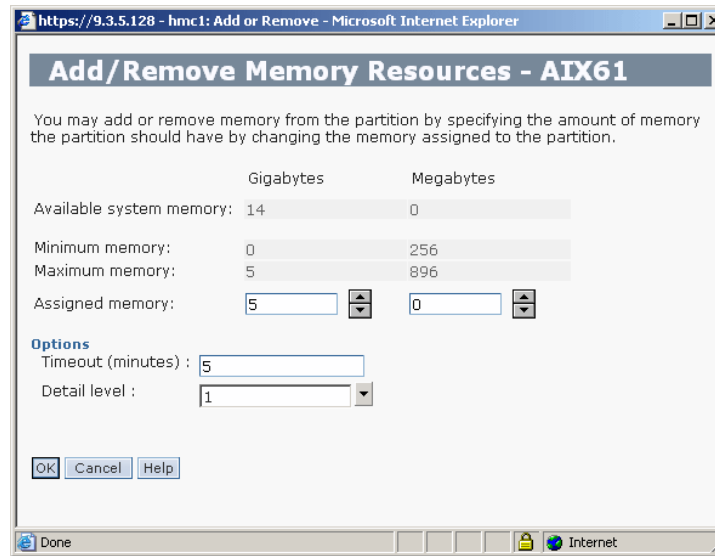


Figure 6-10 Changing the total amount of memory of the partition to 5 GB

3. Click **OK** when you are done. A status window as shown in Figure 6-11 is displayed.

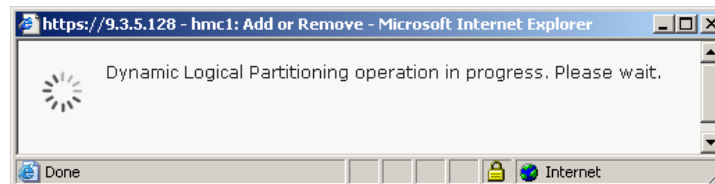


Figure 6-11 Dynamic LPAR operation in progress

**Note:** For an IBM i partition dynamically added memory is added to the *base* memory pool (system pool 2 as shown in the WRKSYSSTS or WRKSHRPOOL screen) and dynamically distributed to other memory pools when using the default automatic performance adjustment (QPFRADJ=2 system value setting).

## 6.2.3 Removing memory dynamically

The following steps describe the dynamic removal of memory from a logical partition:

1. Select the logical partition where you want to initiate a dynamic LPAR operation. The first window in any dynamic operation will be similar to Figure 6-12.

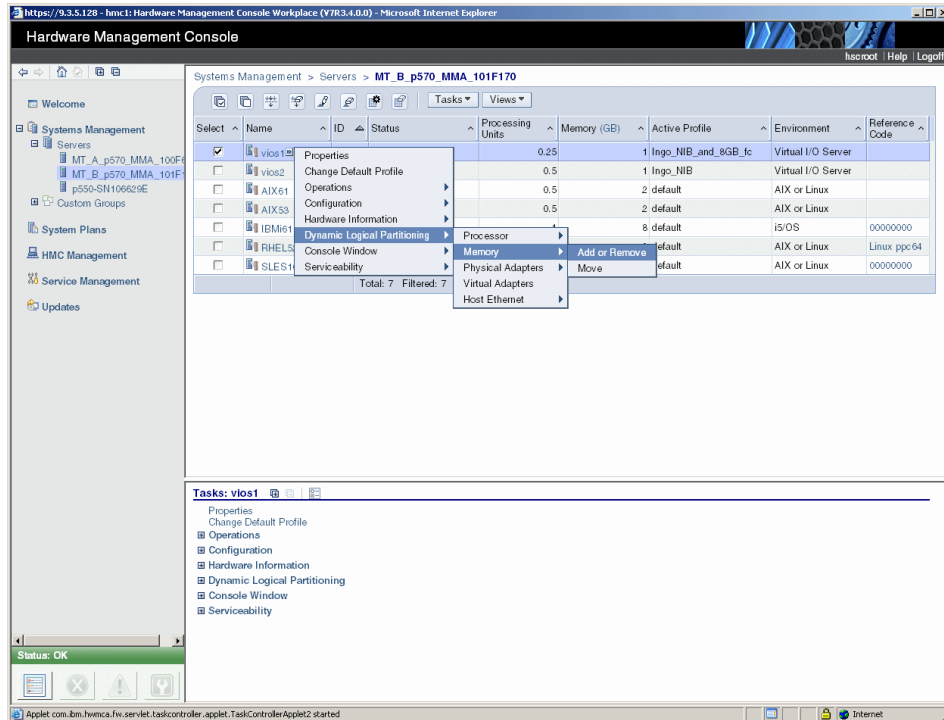


Figure 6-12 Add or remove memory operation

For our AIX partition the memory settings before the operation are:

```
# lsattr -El mem0
goodsize 5120 Amount of usable physical memory in Mbytes False
size      5120 Total amount of physical memory in Mbytes False
```

The graphical user interface to change the memory allocated to a partition is the same one used to add memory in Figure 6-10 on page 253. And the same reminder can be used here. On HMC Version 7, you do not choose the amount to remove from the partition as you did in the previous versions of HMC. Now you just change the total amount of memory to be assigned to the partition. In the command output above the partition has 5 GB and you want to remove, for example, 1 GB from it. In order to do this you just need to change the total amount of memory to 4 GB, as shown in Figure 6-13.

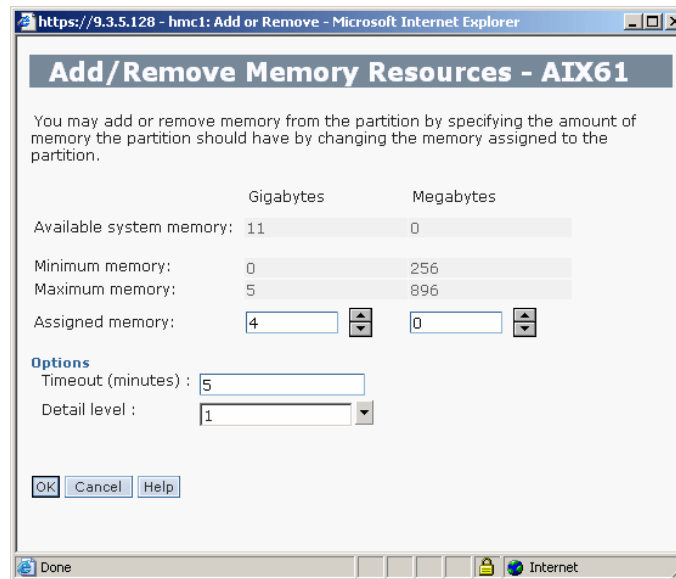


Figure 6-13 Dynamically reducing 1 GB from a partition

2. Click **OK** when done.

The following command shows the effect of the memory deletion on our AIX partition:

```
# lsattr -E1 mem0
goodsize 4096 Amount of usable physical memory in Mbytes False
size      4096 Total amount of physical memory in Mbytes False
```

**Note:** For an IBM i partition dynamically removed memory is removed from the *base* memory pool and only to the extent of leaving the minimum amount of memory required in the base pool as determined by the base storage pool minimum size (QBASPOOL system value).

## 6.2.4 Add physical adapters dynamically

The following steps show the way to add physical adapters dynamically:

1. Log in to HMC and then select the system-managed name. On the right select the partition where you want to execute a dynamic LPAR operation, as shown in Figure 6-14.

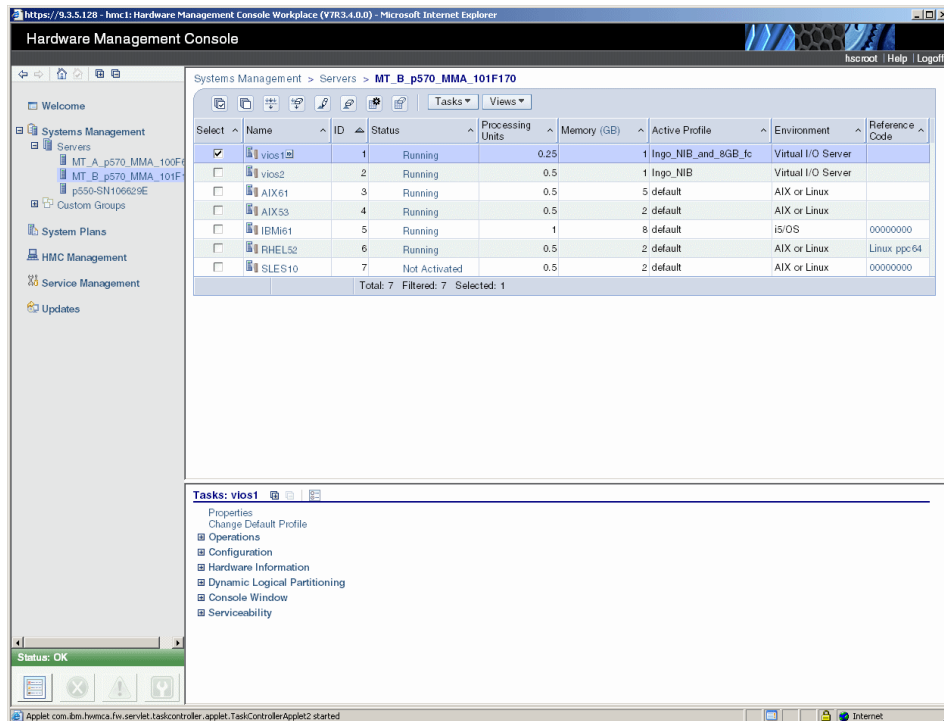


Figure 6-14 LPAR overview menu

- On the **Tasks** menu on the right side of the window choose **Dynamic Logical Partitioning** → **Physical Adapters** → **Add** as shown in Figure 6-15.

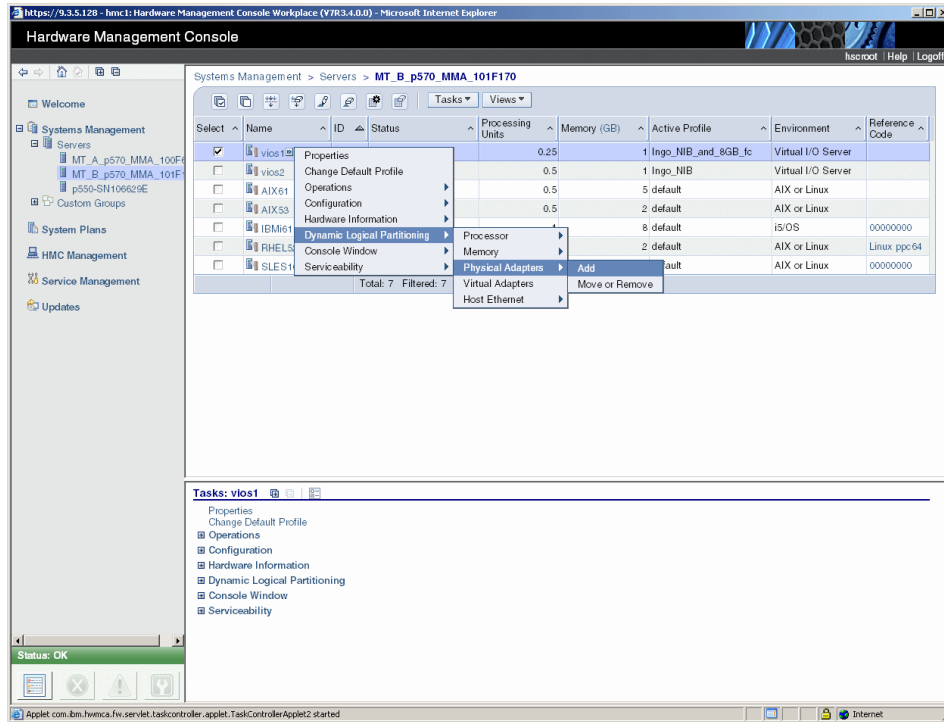


Figure 6-15 Add physical adapter operation

- The next window will look like the one in Figure 6-16. Select the physical adapter you want to add to the partition:

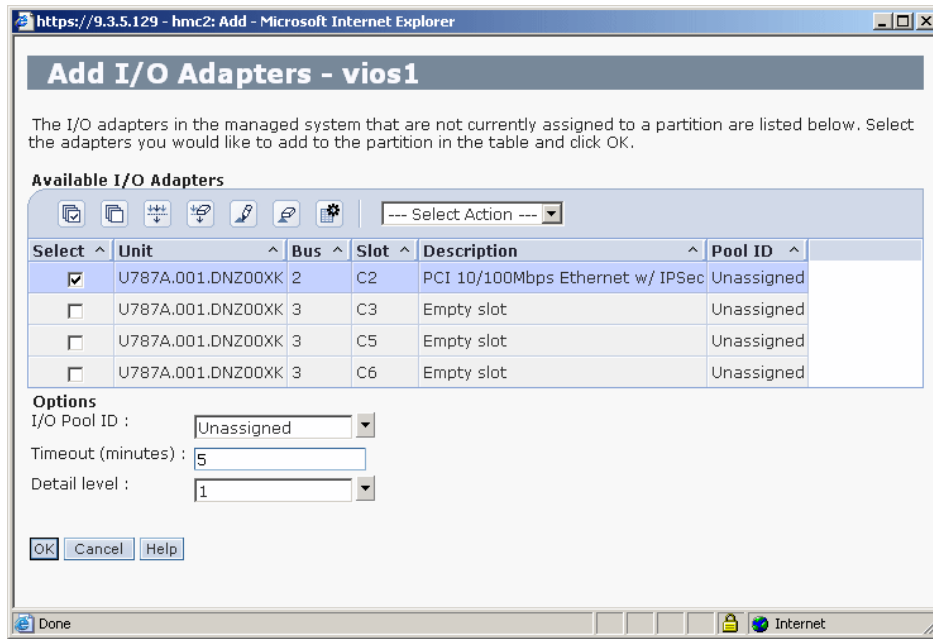


Figure 6-16 Select physical adapter to be added

- Click **OK** when done.

## 6.2.5 Move physical adapters dynamically

In order to move a physical adapter, you first have to release the adapter in the partition that currently owns it. Use the HMC to list which partition owns the adapter. In the left menu select **Systems Management** and then click the system's name. In the right menu select **Properties**. Select the **I/O** tab on the window that will appear, as shown in Figure 6-17. You can see each I/O adapter for each partition.

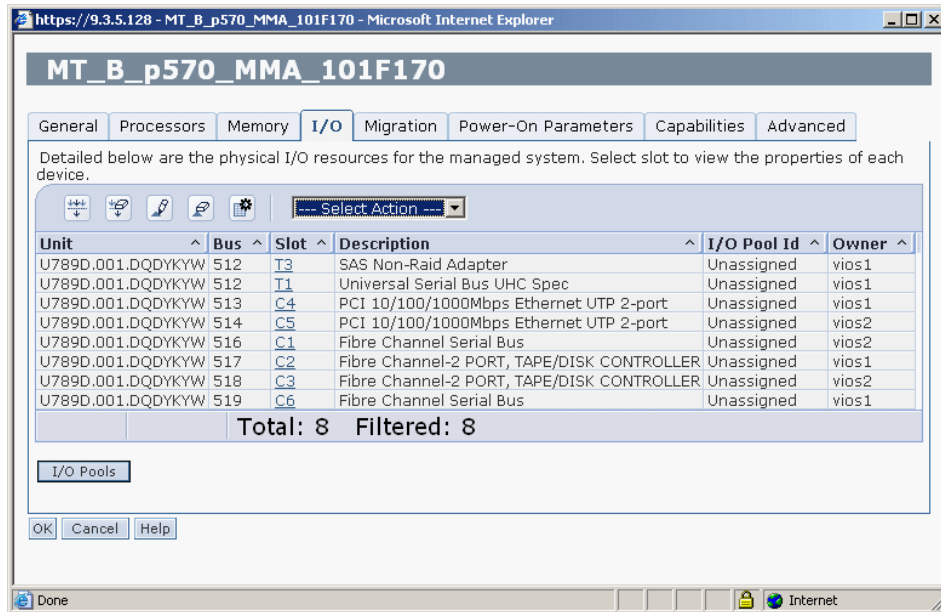


Figure 6-17 I/O adapters properties for a managed system

Usually devices, such as an optical drive, belong to the adapter to be moved and they should be removed as well.

The optical drive often needs to be moved to another partition. For an AIX partition use the `lsslot -c slot` command as padmin user to list adapters and their members. In the Virtual I/O Server you can use the `lsdev -slots` command as follows:

```
$ lsdev -slots
# Slot                Description          Device(s)
U789D.001.DQDYKYW-P1-T1 Logical I/O Slot    pci4 usbhc0 usbhc1
U789D.001.DQDYKYW-P1-T3 Logical I/O Slot    pci3 sissas0
U9117.MMA.101F170-V1-C0 Virtual I/O Slot    vsa0
U9117.MMA.101F170-V1-C2 Virtual I/O Slot    vasi0
U9117.MMA.101F170-V1-C11 Virtual I/O Slot    ent2
U9117.MMA.101F170-V1-C12 Virtual I/O Slot    ent3
U9117.MMA.101F170-V1-C13 Virtual I/O Slot    ent4
U9117.MMA.101F170-V1-C14 Virtual I/O Slot    ent6
U9117.MMA.101F170-V1-C21 Virtual I/O Slot    vhost0
U9117.MMA.101F170-V1-C22 Virtual I/O Slot    vhost1
U9117.MMA.101F170-V1-C23 Virtual I/O Slot    vhost2
U9117.MMA.101F170-V1-C24 Virtual I/O Slot    vhost3
U9117.MMA.101F170-V1-C25 Virtual I/O Slot    vhost4
U9117.MMA.101F170-V1-C50 Virtual I/O Slot    vhost5
U9117.MMA.101F170-V1-C60 Virtual I/O Slot    vhost6
```

For an AIX partition use the `rmdev -l pcin -d -R` command to remove the adapter from the configuration, i.e. releasing it to be able to move it another partition. In the Virtual I/O Server, you can use the `rmdev -dev pcin -recursive` command (n is the adapter number).

For an IBM i partition vary off any devices using the physical adapter before moving it to another partition using the `VRFCFG` command like `VRFCFG CFGOBJ(TAP02) CFGTYPE(*DEV) STATUS(*OFF)` to release the tape drive from using the physical adapter. To see which devices are attached to which adapter use the `WRKHDWRSC` command like `WRKHDWRSC *STG` for storage devices with choosing option 7=Display resource detail for an adapter resource to see its physical location (slot) information and option 9=Work with resources to list the devices attached to it.

Example 6-1 shows how to remove a Fibre Channel adapter from an AIX partition that was virtualized and does not need this adapter any more.

*Example 6-1 Removing the Fibre Channel adapter*

---

```
# lsslot -c pci
# Slot          Description          Device(s)
U789D.001.DQDYKYW-P1-C2  PCI-E capable, Rev 1 slot with 8x lanes  fcs0 fcs1
U789D.001.DQDYKYW-P1-C4  PCI-X capable, 64 bit, 266MHz slot       ent0 ent1
# rmdev -dl fcs0 -R
fcnet0 deleted
fscsi0 deleted
fcs0 deleted
# rmdev -dl fcs1 -R
fcnet1 deleted
fscsi1 deleted
fcs1 deleted
```

---

After adapter has been deleted in the virtual I/O client, the physical adapter can be moved to another partition using the HMC.



1. Select the partition that currently holds the adapter and then select **Dynamic Logical Partitioning** → **Physical Adapters** → **Move or Remove** (see Figure 6-18).

The adapter must not be set as required in the profile. To change the setting from required to desired, you have to update the profile, and shutdown and activate the LPAR again (not just reboot).

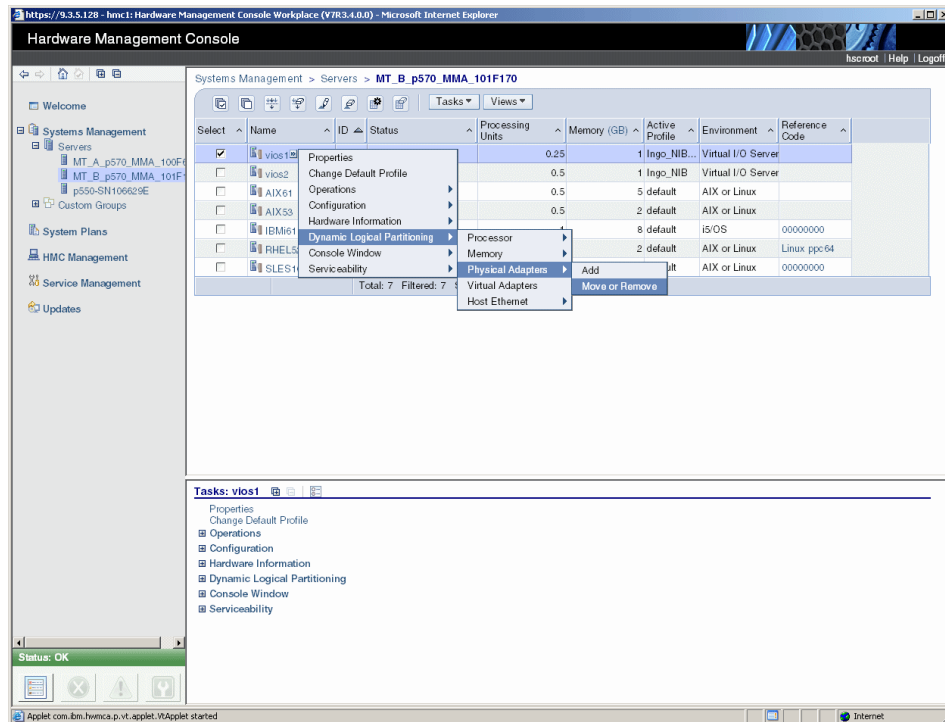


Figure 6-18 Move or remove physical adapter operation

2. Select the adapter to be moved and select the receiving partition, as shown in Figure 6-19.

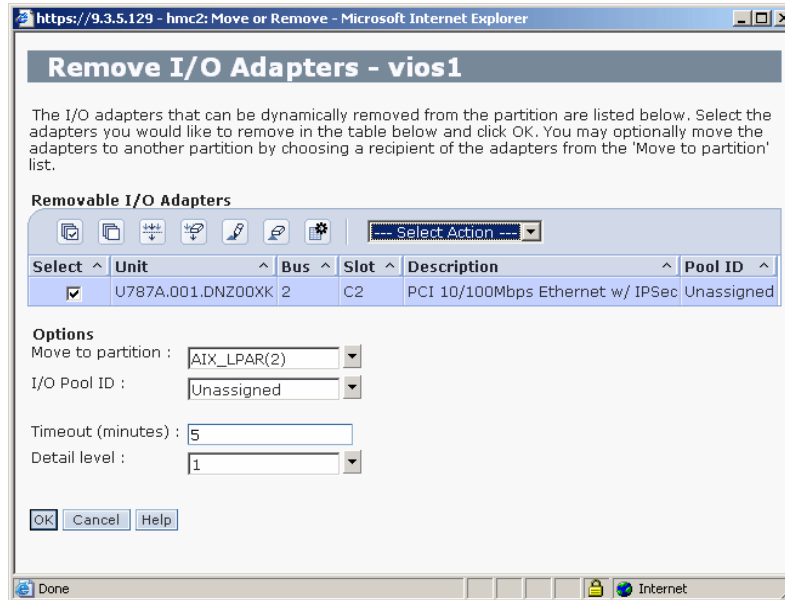


Figure 6-19 Selecting adapter in slot C2 to be moved to partition AIX\_LPAR

1. Click **OK** to execute.
2. For an AIX partition run the **cfgmgr** command (**cfgdev** in the Virtual I/O Server) in the receiving partition to make the adapter and its devices available.

An IBM i partition by default automatically discovers and configures new devices attached to it so they only need to be varied on using the **VRYCFG** command before using them.

- Remember to update the profiles of both partitions for the change to be reflected across restarts of the partitions. Alternatively, use the **Configuration** → **Save Current Configuration** option to save the changes to a new profile as shown in Figure 6-20.

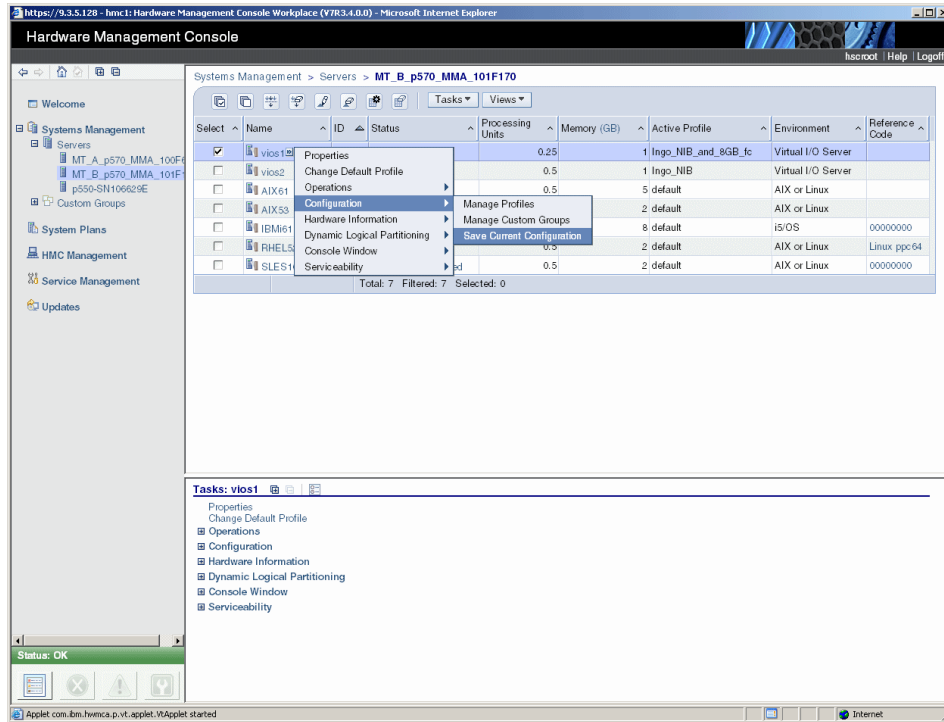


Figure 6-20 Save current configuration

## 6.2.6 Removing physical adapters dynamically

The following steps describe the procedure to remove virtual adapters from a partition dynamically:

- On the HMC select the partition to remove the adapter from and choose **Dynamic Logical Partitioning** → **Physical Adapters** → **Move or Remove** (Figure 6-21).

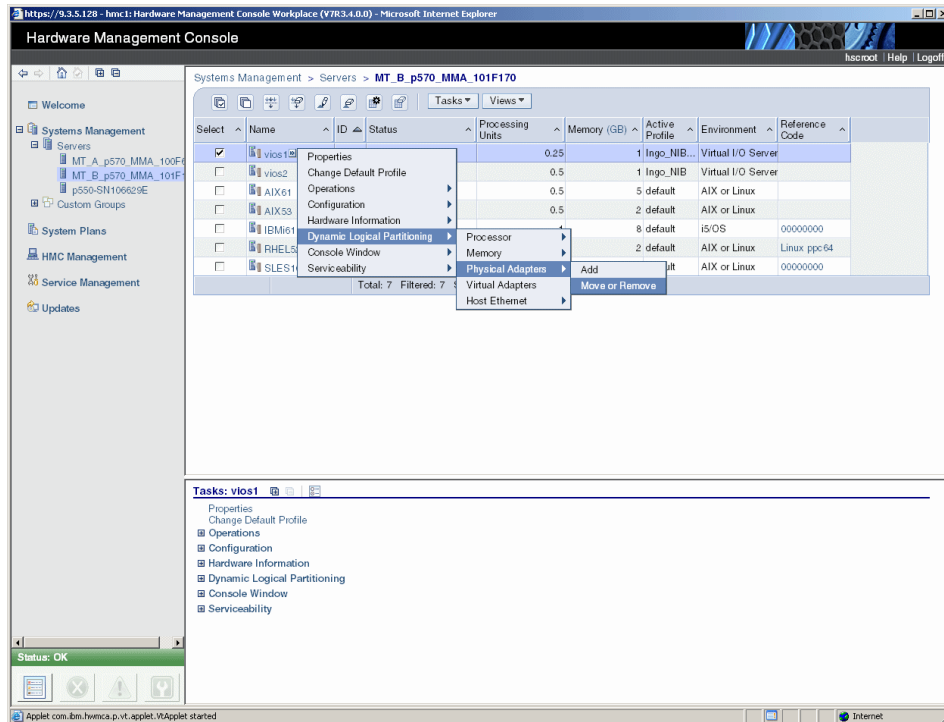


Figure 6-21 Remove physical adapter operation

2. Select the adapter you want to delete as shown in Figure 6-22.

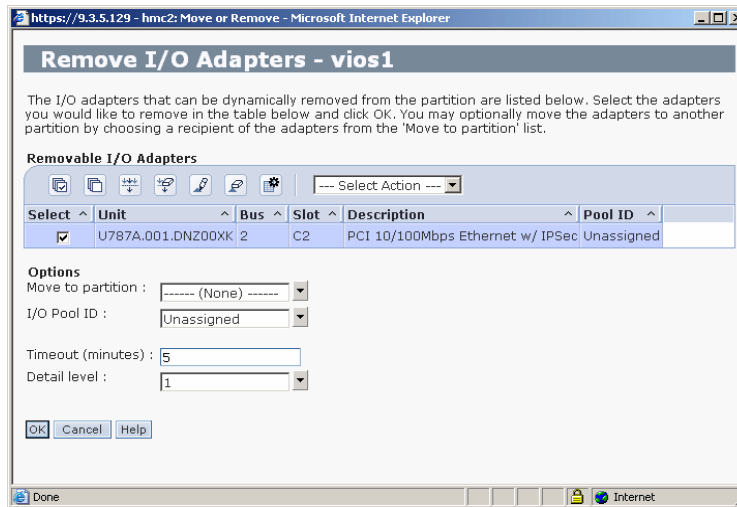


Figure 6-22 Select physical adapter to be removed

3. Click **OK** when done.

## 6.2.7 Add virtual adapters dynamically

The following steps show one way to add virtual adapters dynamically:

1. Log in to HMC and then select the system-managed name. On the right select the partition where you want to execute a dynamic LPAR operation.
2. On the **Tasks** menu on the right side of the window choose **Dynamic Logical Partitioning** → **Virtual Adapters** → **Add** as shown in Figure 6-23.

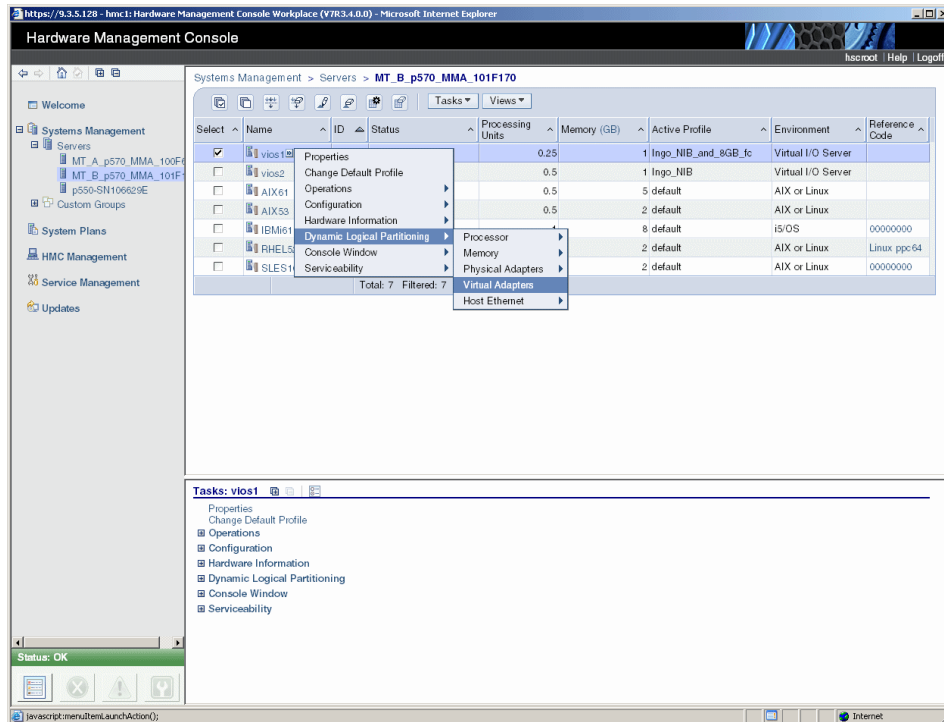


Figure 6-23 Add virtual adapter operation

- The next window will look like the one in Figure 6-24. From the Actions drop-down menu, on the upper left, choose **Create** → **SCSI Adapter**.

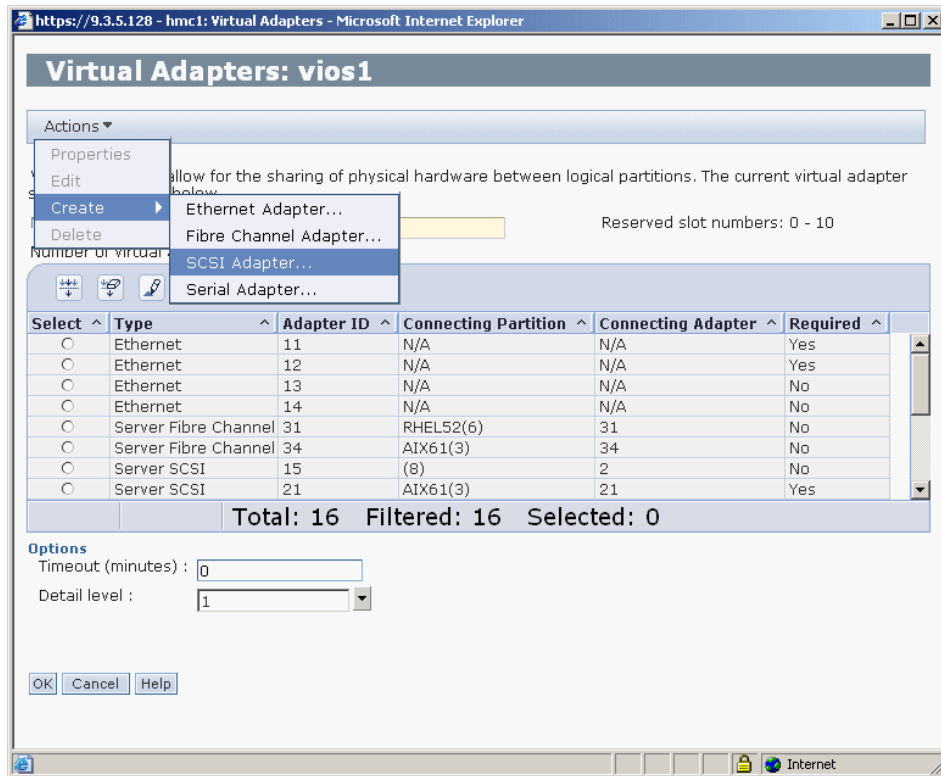


Figure 6-24 Dynamically adding virtual SCSI adapter

- Figure 6-25 shows you the window after clicking **SCSI Adapter...** Put in the slot adapter number of the new virtual SCSI being created. After that you have to select whether this new SCSI adapter can be accessed by any client partition or only by a specific one. In this case, as an example, we are only allowing the SCSI client adapter in slot 2 of the AIX61 partition to access it.

**Note:** In this example we used a different slot numbering for the client and the server virtual SCSI adapter. We recommend to create a overall numbering scheme.

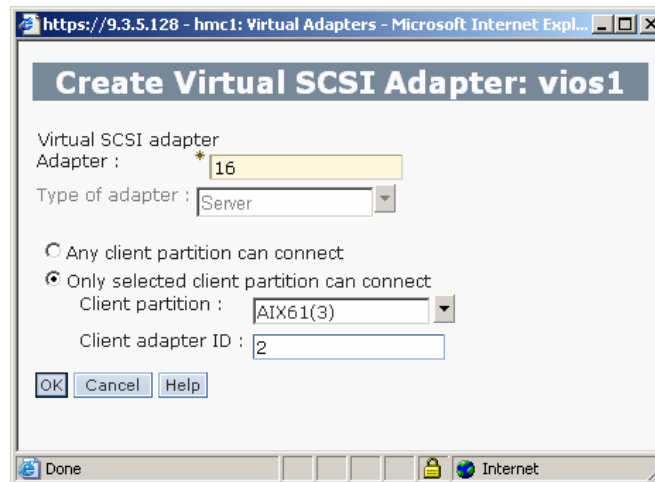


Figure 6-25 Virtual SCSI adapter properties



- The newly created adapter is listed in the adapters list as shown in Figure 6-26.

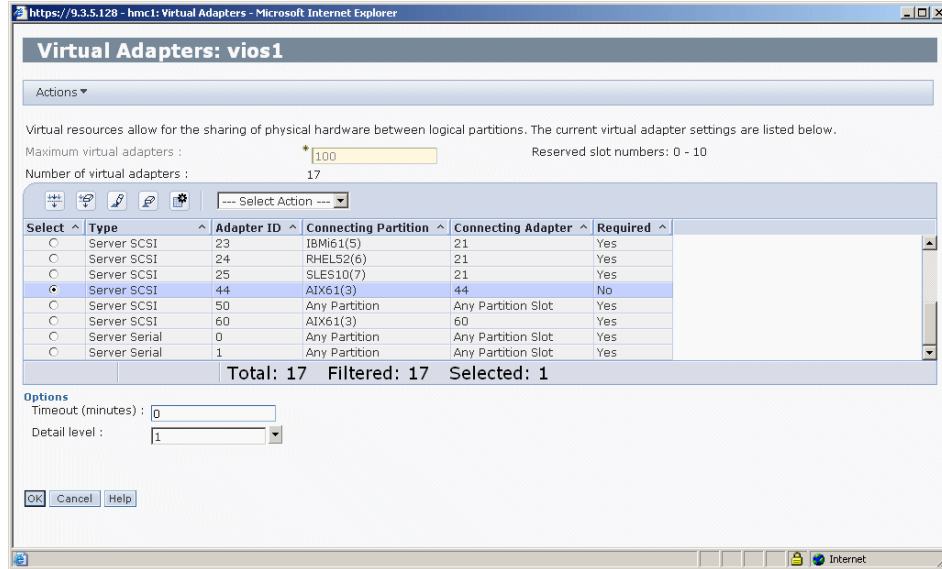


Figure 6-26 Virtual adapters for an LPAR

- Click **OK** when done.

## 6.2.8 Removing virtual adapters dynamically

The following steps describe the removal of virtual adapters from a partition dynamically:

- For AIX unconfigure respectively for IBM i vary-off any devices attached to the virtual client adapter and unconfigure the virtual adapter itself on the AIX virtual I/O client.

**Note:** First remove all associated virtual *client* adapters in the virtual I/O clients before removing a virtual *server* adapter in the Virtual I/O Server.

- On the HMC select the partition to remove the adapter from and choose **Dynamic Logical Partitioning** → **Virtual Adapters** (Figure 6-27).

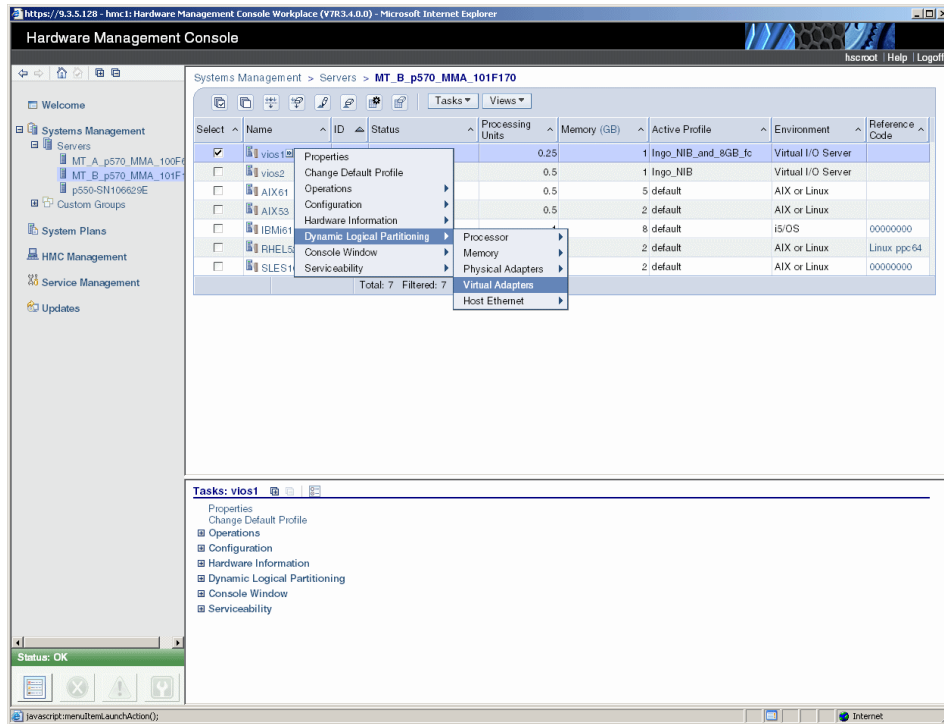


Figure 6-27 Remove virtual adapter operation

3. Select the adapter you want to delete and choose **Actions** → **Delete**. (Figure 6-28).

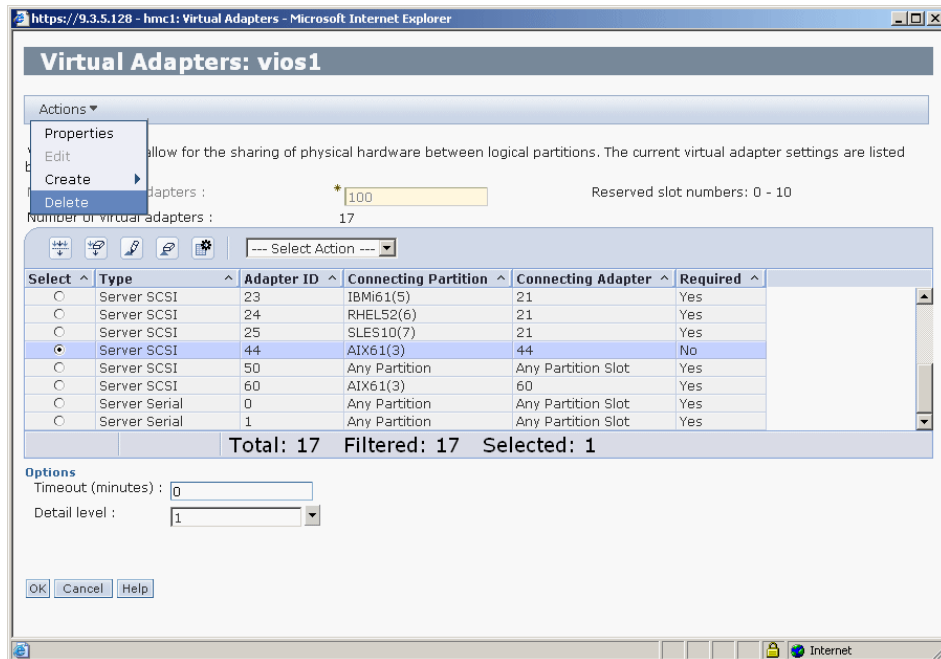


Figure 6-28 Delete virtual adapter

4. Click **OK** when done.

## 6.2.9 Removing or replacing a PCI Hot Plug adapter

PCI Hot Plug feature enables one to remove host based adapters with out shutting down the partition. Replacing an adapter could happen when, for example, you exchange 2 Gb Fibre Channel adapters for a 4 Gb Fibre Channel adapter, or another, for configuration changes or updates.

For virtual Ethernet adapters in the virtual I/O client, redundancy has to be enabled either through Shared Ethernet failover enabled on the Virtual I/O Servers, or Network Interface Backup configured if continuous network connectivity is required. If there is no redundancy for Ethernet, the replace operation can be done while the virtual I/O client is still running, but it will lose network connectivity during replacement. For virtual I/O clients that have redundant paths to their virtual disks and are not mirroring these disks, it is necessary to shut them down for the time used to replace the adapter.

On the Virtual I/O Server, in both cases there will be child devices connected to the adapter because the adapter would be in use before. Therefore, the child devices will have to be unconfigured, as well as the adapter, before the adapter can be removed or replaced. Normally there is no need to remove the child devices, for example disks and mapped disks, also known as Virtual Target Devices, in the case of a Fibre Channel adapter replacement, but they have to be unconfigured (set to the defined state) before the adapter they rely on can be replaced.

## 6.3 Dynamic LPAR operations on Linux for Power

In the next sections a detailed explanation how to run dynamic LPAR operations in Linux for Power is given.

### 6.3.1 Service and productivity tools for Linux for Power

Virtualization and hardware support in Linux for Power is realized through Open Source drivers included in the standard Linux Kernel for 64-bit POWER-based systems. However, IBM provides some additional tools for virtualization management which are useful for exploiting some advanced features and hardware diagnostics. These tools are called *Service and productivity tools for Linux for Power* and are provided at a no-cost download for all supported distributions and systems. The tools include RSCT Reliable Scalable Cluster Technology daemons used for communication with Hardware Management Console. Some packages are Open Source and included on the distribution media. However, the web site download offers the latest version. See details about each package in Table 6-1.

Table 6-1 Service &amp; Productivity tools description

<b>Service and Productivity tool file name</b>	<b>Description</b>
librtas	<p>Platform Enablement Library (base tool)</p> <p>The librtas package contains a library that allows applications to access certain functionality provided by platform firmware. This functionality is required by many of the other higher-level service and productivity tools.</p> <p>This package is Open Source and shipped by both Red Hat and Novell SUSE.</p> <p>Support for this package is provided by IBM Support Line, Novell SUSE or Red Hat, depending on the vendor providing support for your Linux OS.</p>
src	<p>SRC is a facility for managing daemons on a system. It provides a standard command interface for defining, undefining, starting, stopping, querying status and controlling trace for daemons. This package is currently IBM proprietary.</p>
rsct.core and rsct.core.utils	<p>Reliable scalable cluster technology (RSCT) core and utilities</p> <p>The RSC packages provide the Resource Monitoring and Control (RMC) functions and infrastructure needed to monitor and manage one or more Linux systems. RMC provides a flexible and extensible system for monitoring numerous aspects of a system. It also allows customized responses to detected events.</p> <p>This package is currently IBM proprietary.</p>
csm.core and csm.client	<p>Cluster Systems Management (CSM) core and client</p> <p>The CSM packages provide for the exchange of host-based authentication security keys. These tools also set up distributed RMC features on the Hardware Management Console (HMC).</p> <p>This package is currently IBM proprietary.</p>

Service and Productivity tool file name	Description
devices.chrp.base .ServiceRM	<p>Service Resource Manager (ServiceRM)</p> <p>Service Resource Manager is a Reliable, Scalable, Cluster Technology (RSCT) resource manager that creates the Serviceable Events from the output of the Error Log Analysis Tool (diagela). ServiceRM then sends these events to the Service Focal Point on the Hardware Management Console (HMC).</p> <p>This package is currently IBM proprietary.</p>
DynamicRM	<p>DynamicRM (Productivity tool)</p> <p>Dynamic Resource Manager is a Reliable, Scalable, Cluster Technology (RSCT) resource manager that allows a Hardware Management Console (HMC) to do the following:</p> <ul style="list-style-type: none"> <li>▶ Dynamically add or remove processors or I/O slots from a running partition</li> <li>▶ Concurrently update system firmware</li> <li>▶ Perform certain shutdown operations on a partition</li> </ul> <p>This package is currently IBM proprietary.</p>
lsvpd	<p>Hardware Inventory</p> <p>The lsvpd package contains the <b>lsvpd</b>, <b>lscfg</b>, and <b>lsmcode</b> commands. These commands, along with a boot-time scanning script named update-lsvpd-db, constitute a hardware inventory system. The <b>lsvpd</b> command provides Vital Product Data (VPD) about hardware components to higher-level serviceability tools. The <b>lscfg</b> command provides a more human-readable format of the VPD, as well as some system-specific information.</p> <p>This package is Open Source and shipped by both Red Hat and Novell SUSE.</p>
service1og	<p>Service Log (service tool)</p> <p>The Service Log package creates a database to store system-generated events that may require service. The package includes tools for querying the database.</p> <p>This package is Open Source and shipped by both Red Hat and Novell SUSE.</p>

Service and Productivity tool file name	Description
<b>diagela</b>	<p>The Error Log Analysis tool provides automatic analysis and notification of errors reported by the platform firmware on IBM systems. This RPM analyzes errors written to /var/log/platform. If a corrective action is required, notification is sent to the Service Focal Point on the Hardware Management Console (HMC), if so equipped, or to users subscribed for notification via the file /etc/diagela/mail_list. The Serviceable Event sent to the Service Focal Point and listed in the e-mail notification may contain a Service Request Number. This number is listed in the <i>Diagnostics Information for Multiple Bus Systems</i> manual.</p> <p>This package is currently IBM proprietary.</p>
<b>rpa-pci-hotplug</b>	<p>The rpa-pci-hotplug package contains two tools to allow PCI devices to be added, removed, or replaced while the system is in operation: <b>lsslot</b>, which lists the current status of the system's PCI slots, and <b>drslot_chrp_pci</b>, an interactive tool for performing hotplug operations.</p> <p>This package is currently IBM proprietary.</p>
<b>rpa-dlpar</b>	<p>The rpa-dlpar package contains a collection of tools allowing the addition and removal of processors and I/O slots from a running partition. These tools are invoked automatically when a dynamic reconfiguration operation is initiated from the attached Hardware Management Console (HMC).</p> <p>This package is currently IBM proprietary.</p>
IBMinvscout	<p>The Inventory Scout tool surveys one or more systems for hardware and software information. The gathered data can be used by Web services such as the Microcode Discovery Service, which generates a report indicating if installed microcode needs to be updated.</p> <p>This package is currently IBM proprietary.</p>

## IBM Installation Toolkit for Linux for Power

Alternatively to manual installation of additional packages described in , “Installing Service and Productivity tools” on page 276, the IBM Linux for Power Installation Toolkit can be used.

The IBM Installation Toolkit for Linux for Power is a bootable CD that provides access to the additional packages that you need to install in order to provide

additional capabilities of your server. It also allows you to set up an installation server to make your customized operating system installation files available for other server installations. Download the IBM Installation Toolkit for Linux for Power iso image from:

<http://www14.software.ibm.com/webapp/set2/sas/f/lopdiags/installtools/home.html>

The IBM Installation Toolkit for Linux for Power simplifies the Linux for Power installation by providing a wizard that allows you to install and configure Linux for Power machines in just a few steps. It supports DVD and network-based installs by providing an application to create and manage network repositories containing Linux and IBM value-added packages.

The IBM Installation Toolkit includes:

- ▶ The Welcome Center, the main toolkit application - A centralized user interface for system diagnostics, Linux and RAS Tools installation, microcode update and documentation.
- ▶ System Tools - An application to create and manage network repositories for Linux and IBM RAS packages.
- ▶ The POWER Advance Toolchain - A technology preview toolchain which provides decimal floating point support, Power architecture c-library optimizations, optimizations in the gcc compiler for POWER, and performance analysis tools.
- ▶ Microcode packages.
- ▶ More than 20 RAS Tools packages.
- ▶ More than 60 Linux for Power user guides and manuals.

## Installing Service and Productivity tools

Download Service and Productivity tools at:

<http://www14.software.ibm.com/webapp/set2/sas/f/lopdiags/home.html>

Then select your distribution and whether you have an HMC-connected system.

Download all the packages in one directory and run the `rpm -Uvh *rpm` command as shown in Example 6-2. Depending on your Linux distribution, version, and installation choice you will be prompted for missing dependencies. Keep your software installation source for your distribution available and accessible.

*Example 6-2 Installing Service and Productivity tools*

---

```
[root@localhost saids]# rpm -Uvh *
Preparing... ##### [100%]
```



```

1:librtas                ##### [ 7%]
2:lsvdp                  ##### [ 14%]
3:src                    ##### [ 21%]
Adding srcmstr to inittab...
4:rsct.core.utils       ##### [ 29%]
5:rsct.core              ##### [ 36%]
0513-071 The ctcas Subsystem has been added.
0513-071 The ctrmc Subsystem has been added.
0513-059 The ctrmc Subsystem has been started. Subsystem PID is 14956.
6:rpa-pci-hotplug       ##### [ 43%]
7:servicelog            ##### [ 50%]
8:csm.core               ##### [ 57%]
9:csm.client             ##### [ 64%]
0513-071 The ctrmc Subsystem has been added.
0513-059 The ctrmc Subsystem has been started. Subsystem PID is 15173.
10:devices.chrp.base.Servi##### [ 71%]
0513-071 The ctrmc Subsystem has been added.
0513-059 The ctrmc Subsystem has been started. Subsystem PID is 15356.
11:diagela               ##### [ 79%]
Starting rtas_errd (platform error handling) daemon: [ OK ]
Registration successful (id: 2)
Registration successful (id: 3)
Stopping rtas_errd (platform error handling) daemon: [ OK ]
Starting rtas_errd (platform error handling) daemon: [ OK ]
12:DynamicRM             ##### [ 86%]
0513-071 The ctrmc Subsystem has been added.
0513-059 The ctrmc Subsystem has been started. Subsystem PID is 15543.
13:IBMinvscout           ##### [ 93%]
Creating group: 'invscout' and 'system'
Creating user : 'invscout'
Assigning user 'invscout' into group 'invscout'
Creating tmp and utility directories....
Checking for /etc/invscout/ ... did not exist! I created it.
Checking for /var/adm/invscout/ ... did not exist! I created it.
Checking for /tmp/invscout/ ... did not exist! I created it.
14:rpa-dlpar             ##### [100%]

```

---

## Service and Productivity tools examples

After the packages are installed, run the `/sbin/update-lsvpd-db` command to initialize the Vital Product Data database.

Now you can list hardware with the `lscfg` command and see proper location codes, as shown in Example 6-3:

### *Example 6-3 lscfg command on Linux*

```

[root@localhost saids]# lscfg
INSTALLED RESOURCE LIST

```

The following resources are installed on the machine.  
 +/- = Added or deleted from Resource List.  
 \* = Diagnostic support not available.

Model Architecture: chrp  
 Model Implementation: Multiple Processor, PCI Bus

+ sys0		System Object
+ sysplanar0		System Planar
+ eth0	U9117.MMA.101F170-V5-C2-T1	Interpartition Logical LAN
+ eth1	U9117.MMA.101F170-V5-C3-T1	Interpartition Logical LAN
+ scsi0	U9117.MMA.101F170-V5-C21-T1	Virtual SCSI I/O Controller
+ sda	U9117.MMA.101F170-V5-C21-T1-L1-L0	Virtual SCSI Disk Drive (21400 MB)
+ scsi1	U9117.MMA.101F170-V5-C22-T1	Virtual SCSI I/O Controller
+ sdb	U9117.MMA.101F170-V5-C22-T1-L1-L0	Virtual SCSI Disk Drive (21400 MB)
+ mem0		Memory
+ proc0		Processor

---

You can use the **lsvpd** command to display vital product data (for example, the firmware level), as shown in the Example 6-4:

*Example 6-4 lsvpd command*

---

```
[root@localhost saids]# lsvpd
*VC 5.0
*TM IBM,9117-MMA
*SE IBM,02101F170
*PI IBM,02101F170
*OS Linux 2.6.18-53.el5
```

---

In order to display virtual adapters, use the **lsvio** command, as shown in Example 6-5.

*Example 6-5 Display virtual SCSI and network*

---

```
[root@linuxlpar saids]# lsvio -s
scsi0 U9117.MMA.101F170-V5-C21-T1
scsi1 U9117.MMA.101F170-V5-C22-T1
[root@linuxlpar saids]# lsvio -e
eth0 U9117.MMA.101F170-V5-C2-T1
```

eth1 U9117.MMA.101F170-V5-C3-T1

---

## Dynamic LPAR with Linux for Power

Dynamic logical partitioning is needed to change the physical or virtual resources assigned to the partition without reboot or disruption. If you create or assign a new virtual adapter, a dynamic LPAR operation is needed to make the operating system aware of this change. On Linux-based systems, existing hot plug and udev mechanisms are utilized for dynamic LPAR operations, so all the changes occur dynamically, and there is no need to run any configuration manager.

Dynamic LPAR requires a working IP connection to the Hardware Management Console (port 657) and the following additional packages installed on the Linux system, as described in , “Installing Service and Productivity tools” on page 276:

```
librtas, src, rsct.core and rsct.core.utils, csm.core and
csm.client, powerpc-utils-papr, devices.chrp.base.ServiceRM,
DynamicRM, rpa-pci-hotplug, rpa-dlpar
```

If you encounter any dynamic LPAR problems, try to ping your HMC first. If the ping is successful, try to list the rmc connection as shown in Example 6-6.

### *Example 6-6 List the management server*

---

```
[root@linuxlpar ~]# lsrsrc IBM.ManagementServer
Resource Persistent Attributes for IBM.ManagementServer
resource 1:
    Name           = "9.3.5.128"
    Hostname       = "9.3.5.128"
    ManagerType    = "HMC"
    LocalHostname  = "9.3.5.115"
    ClusterTM      = "9078-160"
    ClusterSNum    = ""
    ActivePeerDomain = ""
    NodeNameList   = {"linuxlpar"}
```

---

## Addition of processors dynamically

Once the tools are installed, depending on the available shared system resources, users can use HMC to add (virtual) processors, memory to the desired partition (adding processing units does not require dynamic LPAR) as shown in Figure 6-29.

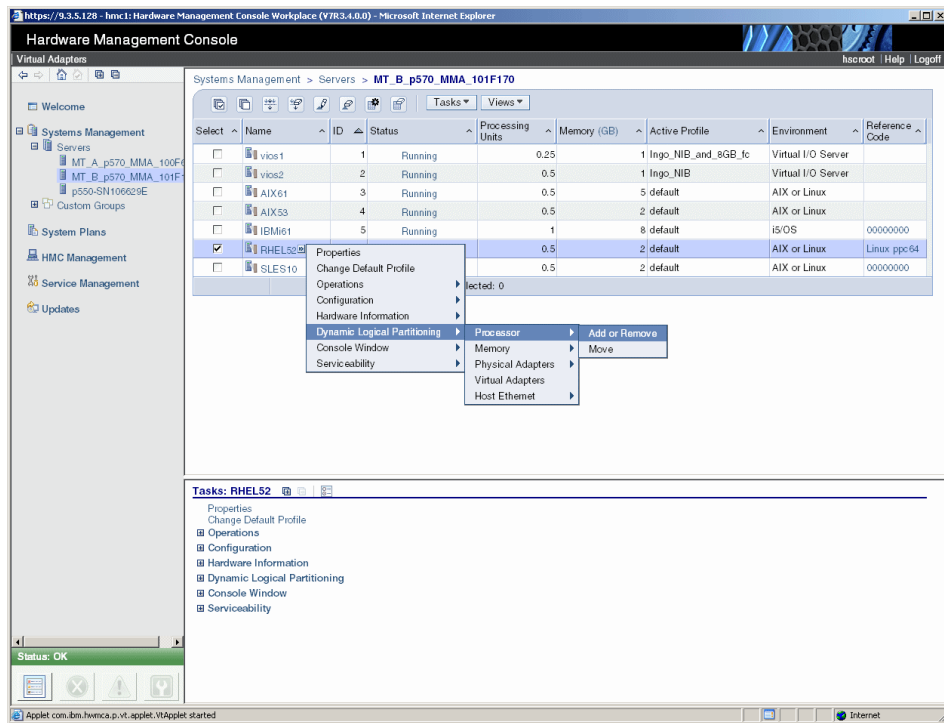


Figure 6-29 Add processor to a Linux partition

In the panel shown in Figure 6-30 on page 281, you can increase the number of processors.

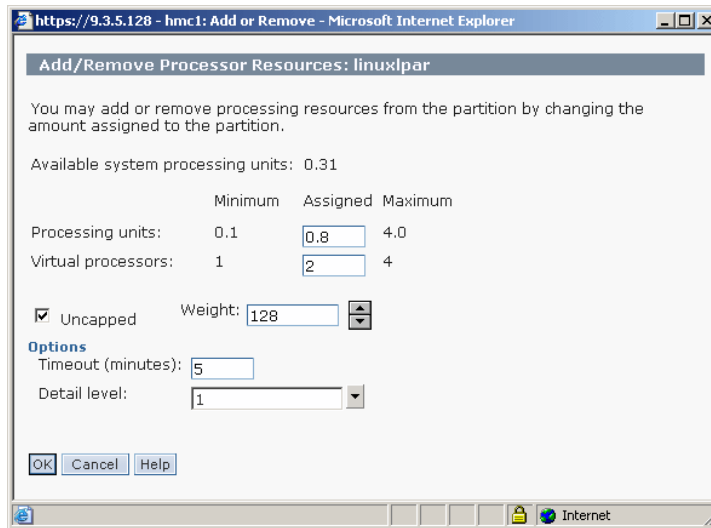


Figure 6-30 Increasing the number of virtual processors

You will be able to receive the following messages on the client if you run the **tail -f /var/log/messages** command as shown in Example 6-7.

*Example 6-7 Linux finds new processors*

---

```
Dec  2 11:26:08 linuxlpar : drmgr: /usr/sbin/drslot_chrp_cpu -c cpu -a -q 60 -p
ent_capacity -w 5 -d 1
Dec  2 11:26:08 linuxlpar : drmgr: /usr/sbin/drslot_chrp_cpu -c cpu -a -q 1 -w
5 -d 1
Dec  2 11:26:08 linuxlpar kernel: Processor 2 found.
Dec  2 11:26:09 linuxlpar kernel: Processor 3 found.
```

---

Besides the messages in the log directory file, users can monitor the changes by executing **cat /proc/ppc64/lparcfg**. The Example 6-8 on page 281 below shows that the partition had 0.5 CPU as its entitled capacity.

*Example 6-8 lparcfg command before adding CPU dynamically*

---

```
lparcfg 1.7
serial_number=IBM,02101F170
system_type=IBM,9117-MMA
partition_id=7
R4=0x32
```

```

R5=0x0
R6=0x80070000
R7=0x800000040004
BoundThrds=1
CapInc=1
DisWheRotPer=5120000
MinEntCap=10
MinEntCapPerVP=10
MinMem=128
MinProcs=1
partition_max_entitled_capacity=100
system_potential_processors=16
DesEntCap=50
DesMem=2048
DesProcs=1
DesVarCapWt=128
DedDonMode=0

partition_entitled_capacity=50
group=32775
system_active_processors=4
pool=0
pool_capacity=400
pool_idle_time=0
pool_num_procs=0
unallocated_capacity_weight=0
capacity_weight=128
capped=0
unallocated_capacity=0
purr=19321795696
partition_active_processors=1
partition_potential_processors=2
shared_processor_mode=1

```

---

The user of this partition added 0.1CPU dynamically. Two lpar configuration attributes that reflect the changes associated with addition/deletion of CPU(s) are `partition_entitled_capacity` and `partition_potential_processor`. The change in the values of **lparcfg** attributes, after addition of 0.1 CPU, is shown in Example 6-9 on page 282

---

*Example 6-9 lparcfg command after addition of 0.1 CPU dynamically*

---

```
lparcfg 1.7
```

```
serial_number=IBM,02101F170
system_type=IBM,9117-MMA
partition_id=7
R4=0x3c
R5=0x0
R6=0x80070000
R7=0x800000040004
BoundThrds=1
CapInc=1
DisWheRotPer=5120000
MinEntCap=10
MinEntCapPerVP=10
MinMem=128
MinProcs=1
partition_max_entitled_capacity=100
system_potential_processors=16
DesEntCap=60
DesMem=2048
DesProcs=2
DesVarCapWt=128
DedDonMode=0
```

```
partition_entitled_capacity=60
group=32775
system_active_processors=4
pool=0
pool_capacity=400
pool_idle_time=0
pool_num_procs=0
unallocated_capacity_weight=0
capacity_weight=128
capped=0
unallocated_capacity=0
purr=19666496864
partition_active_processors=2
partition_potential_processors=2
shared_processor_mode=1
```

---

## Removal of processor(s) dynamically

If you repeat the same steps and decrease the number of processors and run the **dmesg** command, the messages will be as shown in Example 6-10.

### *Example 6-10 Ready to die message*

---

```
IRQ 18 affinity broken off cpu 0
```

```
IRQ 21 affinity broken off cpu 0
cpu 0 (hwid 0) Ready to die...
cpu 1 (hwid 1) Ready to die...
```

---

**Note:** The DLPAR changes made to the partition attributes i.e., CPU, memory are not saved to the current active profile. Hence, users may save the current partition configuration by selecting partition **Name** → **Configuration** → **Save Current Configuration** and then during the next reboot use the saved partition profile as the default profile. An alternative method is to effect the same changes in the default profile.

## Add memory dynamically

**Note:** Users must ensure installation of powerpc-utils-papr rpm. Refer to the productivity download site for the release information.

At the time of writing this book, dynamic addition of memory alone is supported. Adding memory is supported by Red Hat Enterprise (RHEL5.0 or later) and Novell SUSE (SLES10 or later) Linux distributions.

Before adding or removing user can get the partition's current memory information by executing `cat /proc/meminfo` as shown in Example 6-11.

*Example 6-11 Display of total memory in the partition before adding memory*

---

```
[root@VIOCRHEL52 ~]# cat /proc/meminfo | head -3
MemTotal:      2057728 kB
MemFree:       1534720 kB
Buffers:       119232 kB
[root@VIOCRHEL52 ~]#
```

---

From the Hardware Management Console of the partition navigate to **System Management** → your **Server** → **Dynamic Logical Partitioning** → **Memory** → **Add or Remove**. The entire navigation process is described in the following Figure 6-31.



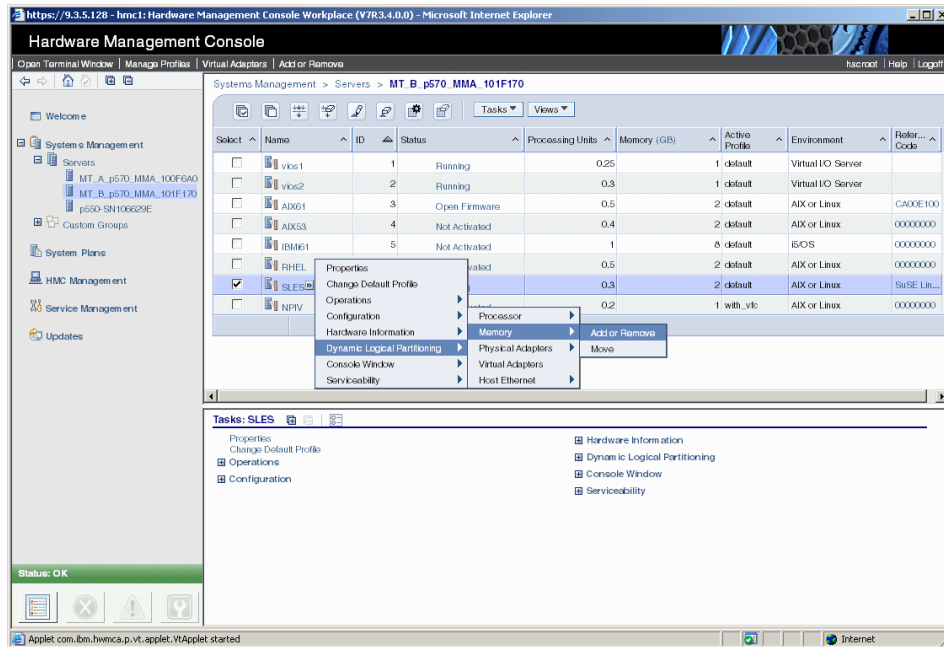


Figure 6-31 DLPAR add or remove memory

The selection of the Add or Remove option results in the display of the following screen (Figure 6-32)

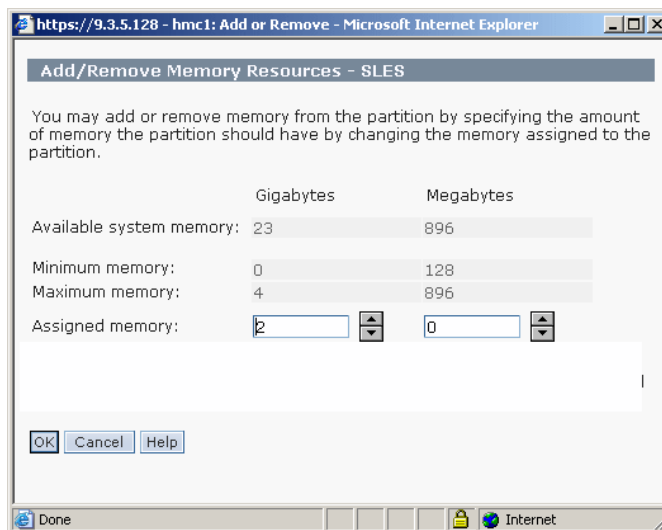


Figure 6-32 DLPAR adding 2 GB memory

Enter the desired memory for the partition in the Assigned Memory box by increasing the value. (In the current example it is increased by 1GB) Click OK

After the action is completed you can verify the addition or removal of memory by executing the command `cat /proc/meminfo` (Example 6-12).

*Example 6-12 Total memory in the partition after adding 1GB dynamically*

---

```
[root@VIOCRHEL52 ~]# cat /proc/meminfo | head -3
MemTotal:      3106304 kB
MemFree:       2678528 kB
Buffers:       65024 kB
[root@VIOCRHEL52 ~]#
```

---

**Note:** The DLPAR changes made to the partition attributes i.e., CPU, memory are not saved to the current active profile. Hence, users may save the current partition configuration by selecting partition **Name** → **Configuration** → **Save Current Configuration** and then during the next reboot use the saved partition profile as the default profile. An alternative method is to effect the same changes in the default profile.

## Managing virtual SCSI changes in Linux

If a new virtual SCSI adapter is added to a Linux partition and dynamic LPAR is functional, this adapter and any attached disks are immediately ready for use.

Sometimes you may need to add a virtual target device to an existing virtual SCSI adapter. In this case the operation is not a dynamic LPAR operation. The adapter itself does not change, just an additional new disk is attached to the same adapter. You need to issue a scan command to recognize this new disk and run the `dmesg` command to see the result, as shown in Example 6-13.

*Example 6-13 Rescanning a SCSI host adapter*

---

```
# echo "- - -" > /sys/class/scsi_host/host0/scan
# dmesg
# SCSI device sdb: 585937500 512-byte hdwr sectors (300000 MB)
sdb: Write Protect is off
sdb: Mode Sense: 2f 00 00 08
sdb: cache data unavailable
sdb: assuming drive cache: write through
SCSI device sdb: 585937500 512-byte hdwr sectors (300000 MB)
sdb: Write Protect is off
sdb: Mode Sense: 2f 00 00 08
sdb: cache data unavailable
sdb: assuming drive cache: write through
sdb: sdb1 sdb2
sd 0:0:2:0: Attached scsi disk sdb
```

```
sd 0:0:2:0: Attached scsi generic sgl type 0
```

---

The added disk is recognized and ready to use as /dev/sdb.

If you are using software mirroring on the Linux client and one of the adapters was set *faulty* due to Virtual I/O Server maintenance, you may need to rescan the disk after both Virtual I/O Servers are available again. Use the following command to issue a disk rescan:

```
echo 1 > /sys/bus/scsi/drivers/sd/<SCSI-ID>/block/device/rescan
```

More examples and detailed information about SCSI scanning is provided in the IBM Linux for Power Wiki page at:

<http://www-941.ibm.com/collaboration/wiki/display/LinuxP/SCSI+-+Hot+add+%2C+remove%2C+rescan+of+SCSI+devices>

## 6.4 Dynamic LPAR operations on the Virtual I/O Server

This section discusses two maintenance tasks for a Virtual I/O Server partition:

- ▶ Ethernet adapter replacement on the Virtual I/O Server.
- ▶ Replacing a Fibre Channel adapter on the Virtual I/O Server

If you want to change any processor, memory, or I/O configuration, follow the steps described in 6.2, “Dynamic LPAR operations on AIX and IBM i” on page 248 as they are similar to ones required for the Virtual I/O Server.

### 6.4.1 Ethernet adapter replacement on the Virtual I/O Server

You can do the replace and remove functions with the `diagmenu` command. Follow the next steps:

1. Enter **diagmenu** and press Enter.
2. Read the Diagnostics Operating Instructions and press Enter to continue.
3. Select **Task Selection (Diagnostics, Advanced Diagnostics, Service Aids, etc.)** and press Enter
4. Go down and select Hot Plug Task and press Enter.
5. Select **PCI Hot Plug Manager** and press Enter
6. Select **Replace/Remove a PCI Hot Plug Adapter** and press Enter.
7. Select the correct adapter you want to replace and press Enter.

8. Select replace in the Operation field and press Enter.
9. Before a replace operation is being operated the adapter can be identified by a blinking LED at the adapter card. You will see the following message:

The visual indicator for the specified PCI slot has been set to the identify state. Press Enter to continue or enter x to exit.

If there are still devices connected to the adapter and a replace or remove operation is performed on that device, there will be an error message in diagmenu:

The visual indicator for the specified PCI slot has been set to the identify state. Press Enter to continue or enter x to exit.

The specified slot contains device(s) that are currently configured. Unconfigure the following device(s) and try again.

```
pci5
ent0
ent1
ent2
ent3
```

These messages mean that devices dependent on this adapter have to be unconfigured first. We will now replace a single Physical Ethernet adapter that is part of a Shared Ethernet Adapter. Here are the steps to do it:

1. Use the **diagmenu** command to unconfigure the Shared Ethernet Adapter and then select **Task Selection** → **Hot Plug Task** → **PCI Hot Plug Manager** → **Unconfigure a Device**.

You should get output similar to the following:

Device Name

Move cursor to desired item and press Enter. Use arrow keys to scroll.

[MORE...16]

ent7	Defined		Standard Ethernet Network Interface
ent0	Available	05-20	4-Port 10/100/1000 Base-TX PCI-X Adapt
ent1	Available	05-21	4-Port 10/100/1000 Base-TX PCI-X Adapt
ent2	Available	05-30	4-Port 10/100/1000 Base-TX PCI-X Adapt
ent3	Available	05-31	4-Port 10/100/1000 Base-TX PCI-X Adapt
ent4	Available		Virtual I/O Ethernet Adapter (1-lan)
ent5	Available		Virtual I/O Ethernet Adapter (1-lan)
<b>ent6</b>	<b>Available</b>		<b>Shared Ethernet Adapter</b>
et0	Defined	05-20	IEEE 802.3 Ethernet Network Interface
et1	Defined	05-21	IEEE 802.3 Ethernet Network Interface

[MORE...90]

Select the Shared Ethernet Adapter (in this example, ent6), and in the following dialogue choose to keep the information about the database: Type or select values in entry fields.  
Press Enter AFTER making all desired changes.

```

                                     [Entry Fields]
* Device Name                        [ent6]
+
  Unconfigure any Child Devices      no
+
  KEEP definition in database        yes
+

```

Press Enter to accept the changes. The system will show that the adapter is now defined:

```
ent6 Defined
```

2. Perform the same operation on the physical adapter (in this example ent0, ent1, ent2, ent3 and pci5) with the difference that now “Unconfigure any Child devices” has to be set to yes.
3. Run **diagmenu**, select **Task Selection** → **Hot Plug Task** → **PCI Hot Plug Manager** → **Replace/Remove a PCI Hot Plug adapter**, and select the physical adapter. You can see a output screen similar to the following:

```
Command: running      stdout: yes      stderr: no
```

Before command completion, additional instructions may appear below.

```
The visual indicator for the specified PCI slot has
been set to the identify state. Press Enter to continue
or enter x to exit.
```

Press Enter as directed and the next message will appear:

```
The visual indicator for the specified PCI slot has
been set to the action state. Replace the PCI card
in the identified slot and press Enter to continue.
Enter x to exit. Exiting now leaves the PCI slot
in the removed state.
```

4. Locate the blinking adapter, replace it, and press Enter. The window will show the message Replace Operation Complete.
5. Run **diagmenu**, select **Task Selection** → **Hot Plug Task** → **PCI Hot Plug Manager** → **Configure a Defined Device** and select the physical Ethernet adapter ent0 that was replaced.
6. Repeat the Configure operation for the Shared Ethernet Adapter.

This method changes if the physical Ethernet adapter is part of a Network Interface Backup Configuration or an IEE 802.3ad link aggregation.

## 6.4.2 Replacing a Fibre Channel adapter on the Virtual I/O Server

For Virtual I/O Servers it is recommended that you have at least two Fibre Channel adapters attached for redundant access to FC-attached disks. This allows for concurrent maintenance, since the multipathing driver of the attached storage subsystem is supposed to handle any outage of a single Fibre Channel adapter. We show the procedure to hot-plug a Fibre Channel adapter connected to a IBM DS4000 series storage device. Depending on the storage subsystem used and the multipathing driver installed, your results may be different.

If there are disks mapped to the virtual SCSI adapters, these devices have to be unconfigured first since there has been no automatic configuration method used to define them.

1. Use the **diagmenu** command to unconfigure devices dependent on the Fibre Channel adapter. Run **diagmenu**, select **Task Selection** → **Hot Plug Task** → **PCI Hot Plug Manager** → **Unconfigure a device**.

Select the disk (or disks in question), and set its state to Defined, as shown in the following:

Unconfigure a Device

Device Name

Move cursor to desired item and press Enter. Use arrow keys to scroll.

[MORE...43]

hdisk6	Available	04-08-02	3542	(200) Disk Array Device
hdisk9	Defined	09-08-00-4,0	16 Bit LVD SCSI Disk Drive	
inet0	Available		Internet Network Extension	
iscsi0	Available		iSCSI Protocol Device	
lg_dumplv	Defined		Logical volume	
lo0	Available		Loopback Network Interface	
loglv00	Defined		Logical volume	
<b>lpar1_rootvg</b>	<b>Available</b>		<b>Virtual Target Device - Disk</b>	
lpar2_rootvg	Available		Virtual Target Device - Disk	
lvdd	Available		LVM Device Driver	

[MORE...34]

2. After that has been done for every mapped disk (Virtual Target Device), set the state of the Fibre Channel Adapter also to Defined:

Unconfigure a Device

Device Name

?

Move cursor to desired item and press Enter. Use arrow keys to scroll.

```
[MORE...16]
et1      Defined    05-09      IEEE 802.3 Ethernet Network Inter
et2      Defined
et3      Defined
et4      Defined
fcnet0   Defined    04-08-01   Fibre Channel Network Protocol De
fcnet1   Defined    06-08-01   Fibre Channel Network Protocol De
fcs0    Available 04-08    FC Adapter
fcs1     Available 06-08      FC Adapter?
fscsi0   Available 04-08-02   FC SCSI I/O Controller Protocol D
fscsi1   Available 06-08-02   FC SCSI I/O Controller Protocol D?
[MORE...61]
```

Be sure to set Unconfigure any Child Devices to Yes, as this will unconfigure the fcnet0 and fscsi0 devices as well as the RDAC driver device dac0:

Type or select values in entry fields.  
Press Enter AFTER making all desired changes.

```
[Entry Fields]
* Device Name          [fcs0]
  Unconfigure any Child Devices    yes
  KEEP definition in database      yes
```

Following is the output of that command, showing the other devices unconfigured:

COMMAND STATUS

```
Command: OK          stdout: yes          stderr: no
```

Before command completion, additional instructions may appear below.

```
fcnet0 Defined
dac0 Defined
fscsi0 Defined
fcs0 Defined
```

3. Run **diagmenu**, select **Task Selection** → **Hot Plug Task** → **PCI Hot Plug Manager** → **Replace/Remove a PCI Hot Plug Adapter**.
4. Select the adapter to be replaced. Set the operation to replace, then press Enter. You will be presented with the following dialogue:

COMMAND STATUS

```
Command: running     stdout: yes          stderr: no
```

Before command completion, additional instructions may appear below.

The visual indicator for the specified PCI slot has been set to the identify state. Press Enter to continue or enter x to exit.

5. Press Enter as directed and the next message will appear.

The visual indicator for the specified PCI slot has been set to the action state. Replace the PCI card in the identified slot and press Enter to continue. Enter x to exit. Exiting now leaves the PCI slot in the removed state.

6. Locate the blinking adapter, replace it, and press Enter. The system will show the message Replace Operation Complete.
7. Select **diagmenu**, select **Task Selection** → **Hot Plug Task** → **PCI Hot Plug Manager** → **Install/Configure Devices Added After IPL**.
8. Press Enter. This calls the **cfgdev** command internally and puts all previously unconfigured devices back to Available.
9. If an Fibre Channel adapter is replaced, the settings such as zoning on the Fibre Channel switch and the definition of the WWPN of the replaced adapter to the storage subsystem have to be done before the replaced adapter can access the disks on the storage subsystem. For IBM DS4000 storage subsystems, we recommend switching the LUN mappings back to their original controllers, as they may have been distributed to balance I/O load.





# PowerVM Live Partition Mobility

PowerVM Live Partition Mobility allows for the movement of an active and non-activated partition from one POWER6 technology-based server to another POWER6-based server with no application downtime, resulting in better system utilization, improved application availability, and energy savings. With PowerVM Live Partition Mobility, planned application downtime due to regular server maintenance can be a thing of the past. PowerVM Live Partition Mobility requires systems with a POWER6 processor running AIX or Linux operating systems and PowerVM Enterprise Edition.

## 7.1 What's new in PowerVM Live Partition Mobility

The following new features now support PowerVM Live Partition Mobility:

- ▶ Processor compatibility modes enable you to move logical partitions between servers with different processor types without upgrading the operating environments installed in the logical partitions.
- ▶ You can move a logical partition that is configured to access storage over a fibre channel network that supports N\_Port ID Virtualization (NPIV) using virtual fibre channel adapters.
- ▶ PowerHA (or High Availability Cluster Multi-Processing) is aware of Partition Mobility. You can move a mobile partition that is running PowerHA to another server without restarting PowerHA.
- ▶ You can move a logical partition from a server that is managed by an Hardware Management Console (HMC) to a server that is managed by a different HMC.

This subject is covered in the Redbooks publication *IBM System p Live Partition Mobility*, SG24-7460 available at:

<http://publib-b.boulder.ibm.com/abstracts/sg247460.html>

The goal of this chapter is to provide a short checklist to prepare your systems for PowerVM Live Partition Mobility.

## 7.2 PowerVM Live Partition Mobility requirements

To prepare for PowerVM Live Partition Mobility, check the following requisites before running a partition migration.

### 7.2.1 HMC requirements

PowerVM Live Partition Mobility can include one or more HMCs as follows:

- ▶ Both the source and destination servers are managed by the same HMC (or redundant HMC pair). In this case, the HMC must be at version 7 release 3.2, or later.
- ▶ The source server is managed by one HMC and the destination server is managed by a different HMC. In this case, both the source HMC and the destination HMC must meet the following requirements:

- The source HMC and the destination HMC must be connected to the same network so that they can communicate with each other.
- The source HMC and the destination HMC must be at version 7, release 3.4.

Use the **lshmc** command to display the HMC version

```
hscroot@hmc1:~> lshmc -V
"version= Version: 7
  Release: 3.4.0
  Service Pack: 0
HMC Build level 20080929.1
", "base_version=V7R3.4.0
```

- Secure shell have to be set up correctly between the 2 HMC.

Run the following command from the source server HMC to configure the ssh authentication to the destination server HMC, 9.3.5.180 is the IP address of the destination HMC:

```
hscroot@hmc1:~> mkauthkeys -u hscroot --ip 9.3.5.180 --g
Enter the password for user hscroot on the remote host 9.3.5.180:
```

Run the following command from the source server HMC to verify the ssh authentication to the destination server HMC:

```
hscroot@hmc1:~> mkauthkeys -u hscroot --ip 9.3.5.180 --test
```

## 7.2.2 Common system requirements

The following are the common system requirements:

- Both source and destination systems are POWER6-based systems.
- The PowerVM Enterprise Edition license code must be installed on both systems.
- Systems have a Virtual I/O Server installed with Version 1.5.1.1 or later. You can check this by running the **ioslevel** command on the Virtual I/O Server.

```
$ ioslevel
2.1.0.0
```

- Systems must have a firmware level of 01Ex320 or later, where x is a S for BladeCenter, a L for Low End servers, a M for Midrange servers, or a H for High End servers. You can check this on the HMC by running the **lslic** command:

```
hscroot@hmc1:~> lslic -m MT_A_p570_MMA_100F6A0 -t sys -F
perm_ecnumber_primary
01EM320
```

If you need to upgrade, look at the following Web page:

[http://publib.boulder.ibm.com/infocenter/systems/scope/hw/index.jsp?topic=/ipha5/fix\\_serv\\_firm\\_kick.htm](http://publib.boulder.ibm.com/infocenter/systems/scope/hw/index.jsp?topic=/ipha5/fix_serv_firm_kick.htm)

- Systems must have the same logical memory block size. This can be checked using the Advanced System Management Interface (ASMI).
- At least one of the source and one of the destination Virtual I/O Servers are set as mover service partition. Check this in their partition properties on the HMC.

**Note:** Setting a Virtual I/O Server as mover service partition automatically creates a Virtual Asynchronous Services Interface (VASI) adapter on this Virtual I/O Server.

- Both Virtual I/O Servers should have their clocks synchronized. See the Time reference in the Setting tab in the Virtual I/O Server properties on the HMC.

### 7.2.3 Source system requirements

Additionally, check that the following requisites are met on your source system:

- The source system uses an external shared access storage system listed in the supported list.  
<http://techsupport.services.ibm.com/server/vios/documentation/datasheet.html>
- The source system uses virtual network adapters defined as bridged shared Ethernet adapters on the Virtual I/O Server, check this by running the **lsmap** command on the Virtual I/O Server.

```
$ lsmap -net -all
SVEA   Physloc
-----
ent2   U9117.MMA.101F170-V1-C11-T1

SEA                    ent5
Backing device        ent0
Status                Available
Physloc                U789D.001.DQDYKYW-P1-C4-T1
```

### 7.2.4 Destination system requirements

Check that the following requisites are met on your destination system:

- The memory available on the destination system is greater or equal to the memory of the source system.
- If the mobile partition uses dedicated processors, the destination system must have at least this number of available processors.
- If the mobile partition is assigned to a Shared-Processor Pool, the destination system must have enough spare entitlement to allocate it to the mobile partition in the destination Shared-Processor Pool.
- The destination server must have at least one Virtual I/O Server that has access to all the subnets used by the mobile partition.

## 7.2.5 Migrating partition requirements

Check that the migrating partition is ready:

- The partition is not designated as a redundant error path reporting partition. Check this in the partition properties on the HMC. Changing this setting currently requires a reboot of the partition.
- The partition is not part of an LPAR workload group. A partition can be dynamically removed from a group. Check this in the properties of the partition of the HMC.
- The partition has a unique name. A partition cannot be migrated if any partition exists with the same name on the destination server.
- The additional virtual adapter slots for this partition (slot ID higher or equal to 2) do not appear as *required* in the partition profile. Check this in the properties of the partition of the HMC.

## 7.2.6 Active and inactive migrations

You can do an active partition migration if the following requisites are met. If this is not the case, you can still run an inactive partition migration:

- The partition is in the Running state.
- The partition does not have any dedicated adapters.
- The partition does not use *huge* pages. Check this in the advanced properties of the partition on the HMC.
- The partition does not use the Barrier Synchronization Register. Check that the “number of BSR arrays” is set to zero in the memory properties of the partition on the HMC. Changing this setting currently requires a reboot of the partition.
- The operating system must be at one of the following levels:
  - AIX 5.3 TL 7 or later

- Red Hat Enterprise Linux Version 5 (RHEL5) Update 1 or later
- SUSE Linux Enterprise Services 10 (SLES 10) Service Pack 1 or later

## 7.3 Managing a live partition migration

In addition to these requirements, it is also recommended to have a high speed Ethernet link between the systems involved in the partition migration. A minimum of 1 Gbps link is recommended.

There are no architected maximum distances between systems for PowerVM Live Partition Mobility. The maximum distance is dictated by the network and storage configuration used by the systems. Standard long-range network and storage performance considerations apply.

### 7.3.1 The migration validation

The migration validation process verifies that the migration of a given partition on one server to another specified server meets all the compatibility requirements and therefore has a good chance of succeeding.

Because the validation is also integrated with the migration operation wizard, you can also use the *Migrate* operation. In the event of any problems being detected, these are reported the same way as for the validation operations.

### 7.3.2 Validation and migration

The migration operation uses a wizard to get the required information. This wizard is accessed from the HMC.

1. Select the partition to migrate and using the popup menu, select **Operations** → **Mobility** → **Validate**
2. In the Migration Validation wizard, fill the **remote HMC** and **Remote User** fields and click **Refresh Destination System** to have the list of available systems on the remote HMC.
3. Select the destination system
4. In the Profile Name, enter a profile name that differs from the names currently created for the mobile partition. This profile will be overwritten with the current partition configuration, click **Validate** as shown in Figure 7-1.

https://9.3.5.128 - hmc1: Validate - Microsoft Internet Explorer

**Partition Migration Validation - MT\_B\_p570\_MMA\_101F170 - AIX\_LPM**

Fill in the following information to set up a migration of the partition to a different managed system. Click Validate to ensure that all requirements are met for this migration. You cannot migrate until the migration set up has been verified.

Source system : MT\_B\_p570\_MMA\_101F170  
 Migrating partition: AIX\_LPM  
 Remote HMC:   
 Remote User:   
 Destination system:  Refresh Destination System  
 Destination profile name:   
 Destination shared processor pool:   
 Source mover service partition:   
 Destination mover service partition:   
 Wait time (in min):   
 Virtual Storage assignments :

Select	Source Slot ID	Slot Type	Destination VIOS

View VLAN Settings... Validate Migrate Cancel Help

Done Internet

Figure 7-1 Partition Migration Validation

5. After the migration validation is successful, you can choose to migrate the partition, click **Migrate** as shown in Figure 7-2.

https://9.3.5.128 - hmc1: Validate - Microsoft Internet Explorer

**Partition Migration Validation - MT\_B\_p570\_MMA\_101F170 - AIX\_LPM**

Fill in the following information to set up a migration of the partition to a different managed system. Click Validate to ensure that all requirements are met for this migration. You cannot migrate until the migration set up has been verified.

Source system : MT\_B\_p570\_MMA\_101F170  
 Migrating partition: AIX\_LPM  
 Remote HMC: 9.3.5.180  
 Remote User: hscroot  
 Destination system: Server-8204-E8A-SN10FE401 Refresh Destination System  
 Destination profile name: mobility  
 Destination shared processor pool: DefaultPool (0)  
 Source mover service partition: vios1 MSP Pairing...  
 Destination mover service partition: vios01  
 Wait time (in min): 5

Virtual Storage assignments :

Select	Source Slot ID	Slot Type	Destination VIOS
<input checked="" type="checkbox"/>	3	SCSI	vios01

View VLAN Settings... Validate Migrate Cancel Help

Done Internet

Figure 7-2 Partition Migration



### 7.3.3 How to fix missing requirements

In the validation summary window, you can check the mandatory requirements missing detected by selecting the **Errors** report. There is additional information related to these errors by selecting the **Detailed information** report. It often helps to determine the precise origin of the missing requirements, as shown in Figure 7-3.

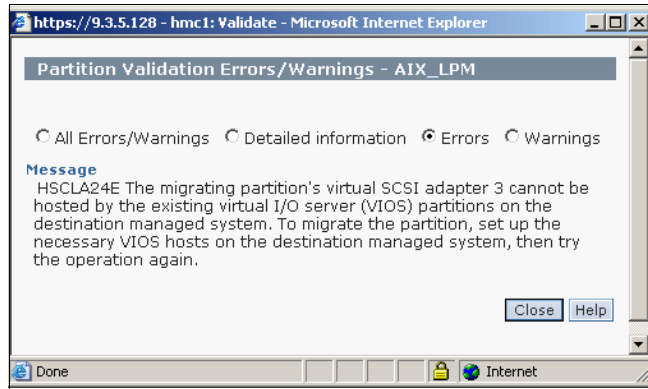


Figure 7-3 Partition migration validation detailed information

The most common missing requirements can be treated as shown in Table 7-1.

Table 7-1 Missing requirements for PowerVM Live Partition Mobility

Partition validation error message reported	Correction
The HMC was unable to find a valid mover service partition	<ul style="list-style-type: none"> <li>▶ Check that there is a Virtual I/O Server on the source and the destination system has “Mover service partition” checked in its general properties.</li> <li>▶ Check that both Virtual I/O Servers can communicate with each other through the network.</li> </ul>
Can not get physical device location - vcd is backed by optical	<ul style="list-style-type: none"> <li>▶ Partitions probably have access to the CD/DVD drive through a virtual device. You then have to temporarily remove the mapping on the virtual CD/DVD drive on the Virtual I/O Server with <b>rmdev</b>.</li> </ul>

Partition validation error message reported	Correction
<p>The migrating partition's virtual SCSI adapter <i>xx</i> cannot be hosted by the existing Virtual I/O Server (VIOS) partitions on the destination managed system.</p>	<p>Check the Detailed information tab.</p> <ul style="list-style-type: none"> <li>▶ If there is a message that mentions "Missing Begin Tag reserve policy mismatched", check that the moving storage disks reserve policy is set to <code>no_reserve</code> on the source and destination disks on the Virtual I/O Servers with a command similar to:  <code>echo "lsattr -El hdiskxxx"   oem_setup_env</code>            You can fix it with the command:  <code>chdev -dev hdiskxxx -attr reserve_policy=no_reserve</code></li> <li>▶ If not, check in your SAN zoning that the destination Virtual I/O Server can access the same LUNs as the source. For IBM storage, you can check the LUN you have access to by running the following command on the Virtual I/O Servers:  <code>echo "fget_config -Av"   oem_setup_env</code>            For other vendors, contact your representative.</li> </ul>

For further information on setting up a system for PowerVM Live Partition<sup>1</sup> Mobility, refer to *IBM System p Live Partition Mobility*, SG24-7460 at:

<http://publib-b.boulder.ibm.com/abstracts/sg247460.html>

## 7.4 Differences with Live Application Mobility

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

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

Live Application Mobility is a feature of AIX Version 6 and will function on all systems that support AIX Version 6, while PowerVM Live Partition Mobility is a PowerVM feature that works for AIX and Linux operating systems that operate on POWER6-based System p servers starting from AIX 5.3 that are configured with the PowerVM Enterprise Edition feature.

<sup>1</sup>

The differences between Live Application Mobility and PowerVM Live Partition Mobility are shown in Table 7-2.

*Table 7-2 PowerVM Live Partition Mobility versus Live Application Mobility*

<b>PowerVM Live Partition Mobility</b>	<b>Live Application Mobility</b>
Requires SAN storage	Uses NFS-mounted file systems to access data
Requires POWER6-based systems	Requires POWER4™, POWER5, or POWER6 based systems
Requires PowerVM Enterprise license	Requires the WPAR migration manager and the Application Mobility license
Can move any supported OS	Can move the applications running in a WPAR on an AIX 6 system only
Move the entire OS	Does not move the OS
Any application may run on the system	Restrictions apply to applications that can run in a WPAR
The resource allocations move with the migrated partition	The administrator may have to adapt the resources allocated to the source and destination partitions





# System Planning Tool

This section describes how the PC-based System Planning Tool, SPT can be used to create a configuration to be deployed on a system. When deploying the partition profiles, assigned resources are generated on the HMC or in IVM. The Virtual I/O Server operating system can be installed during the deployment process. In the scenario on page 306 the Virtual I/O Server, AIX, IBM i and Linux operating systems are all installed using DVD or NIM.

SPT can be downloaded for free from the System Planning Tool Web site. The generated system plan can be viewed from the SPT on a PC or directly on an HMC. Once you have saved your changes, the configuration can be deployed to the HMC or IVM. Detailed information about the SPT can be found at:

<http://www-304.ibm.com/jct01004c/systems/support/tools/systemplanningtool>

**Note:** At the time of writing SPT (Version 3.08.296) does not support the JS12 and JS22 blades.

**Note:** At the time of writing SPT (Version 3.08.296) does not support NPIV.

Further information about how to create and deploy a system plan can be found in *PowerVM Virtualization on IBM System p: Introduction and Configuration*, SG24-7940.

## 8.1 Sample scenario

This scenario shows the system configuration used for this book.

Figure 8-1 and Figure 8-2 show the basic layout of partitions and the slot numbering of virtual adapters.

An additional virtual SCSI server adapter with slot number 60 is added for the virtual tape and one client SCSI adapter with slot number 60 is added to the aix61 and aix53 partitions (not shown in Figure 8-1).

**Note:** At the time of writing virtual tape is not supported on IBM i and Linux partitions.

**Tip:** It is not required to have the same slot numbering for the server and client adapters. It just makes it easier to keep track.

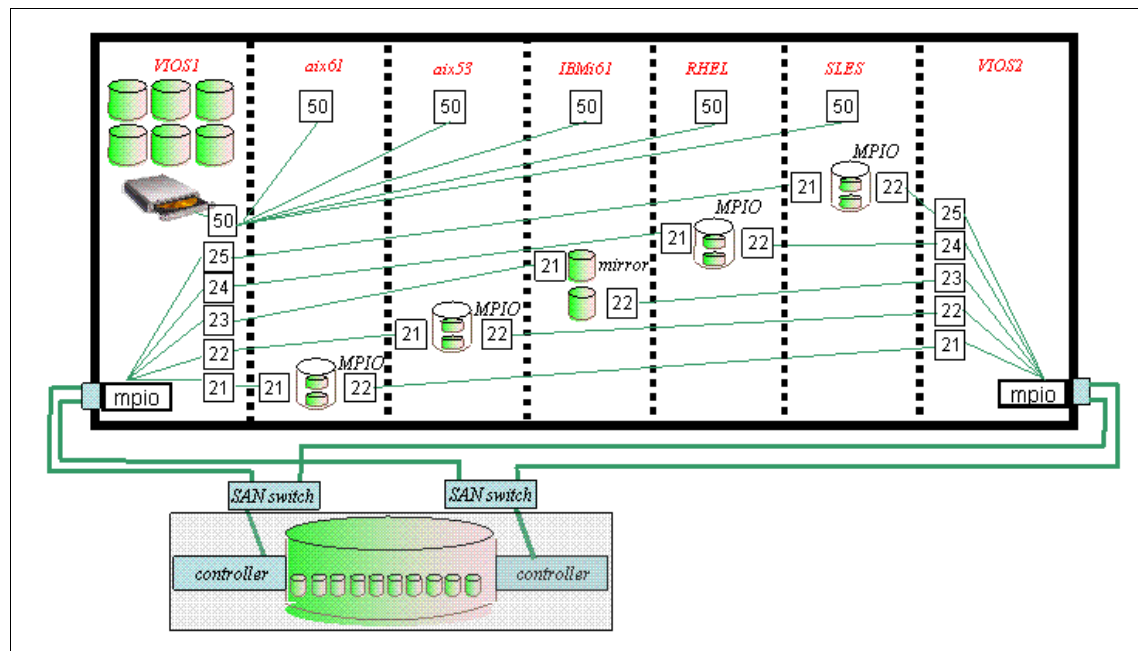


Figure 8-1 The partition and slot numbering plan of virtual storage adapters

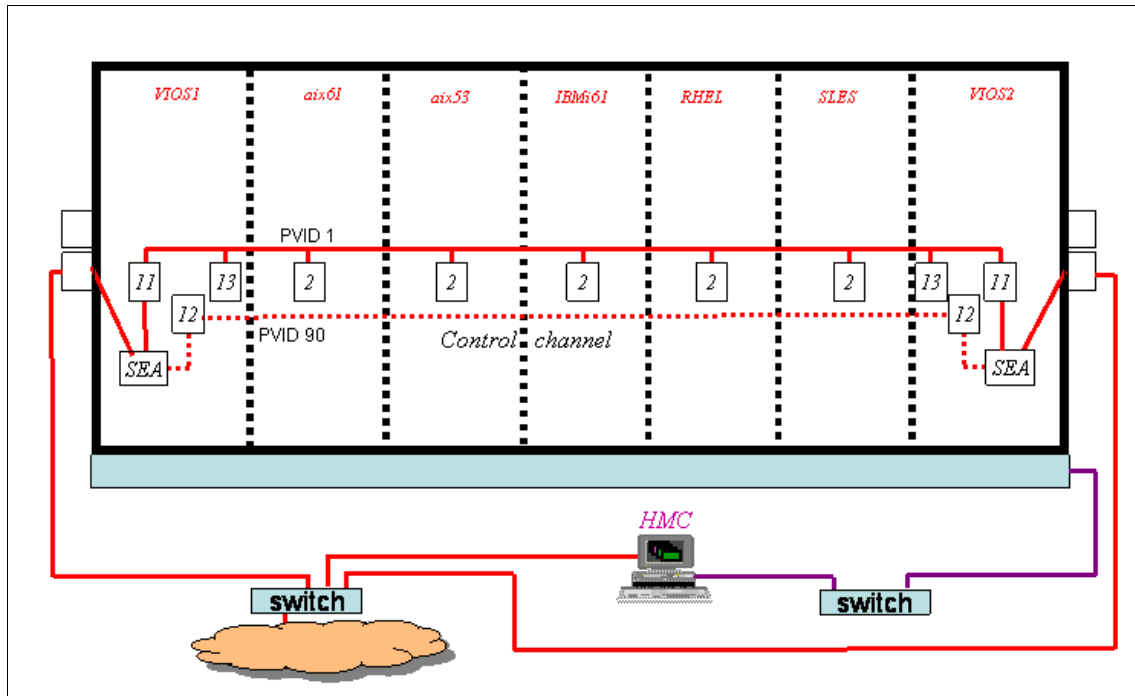


Figure 8-2 The partition and slot numbering plan for virtual Ethernet adapters

### 8.1.1 Preparation recommendation

When deploying a System Plan on the HMC the configuration is validated against the installed system (adapter and disk features and their slot numbers, amount of physical and CoD memory, number/type of physical and CoD CPU and more). A way of simplifying the matching of the SPT System Plan and the physical system is:

- ▶ Create a System Plan of the physical system on the HMC or IVM.
- ▶ Export the System Plan to SPT on your PC and convert it to SPT format.

**Tip:** If the System Plan cannot be converted, use the System Plan to manually create the compatible configuration

- ▶ Use this System Plan as a template and customize it to meet your requirements.
- ▶ Import the completed SPT System Plan to the HMC or IVM and deploy it.

## 8.1.2 Planning the configuration with SPT

Figure 8-3 shows the Partition properties where you can add or modify partition profiles. Notice the Processor and Memory tabs for setting these properties.

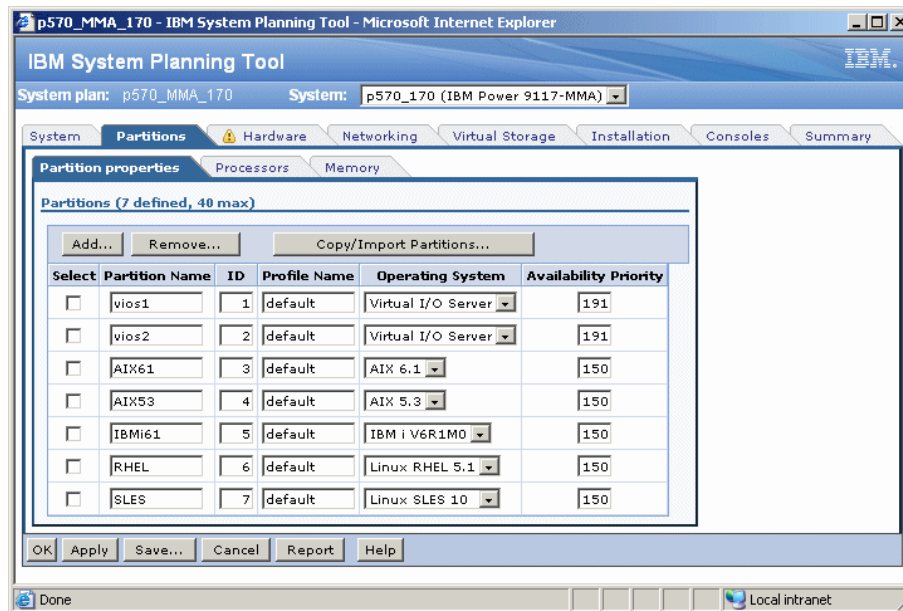


Figure 8-3 The SPT Partition properties window

A very useful feature in SPT is the ability to edit virtual adapter slot numbers and check consistency for server-client SCSI slot numbers. It is also recommended to increase the maximum number of virtual slots for the partitions.

Figure 8-4 shows the SCSI connection window. Click on the **Edit Virtual Slots** to open the window.



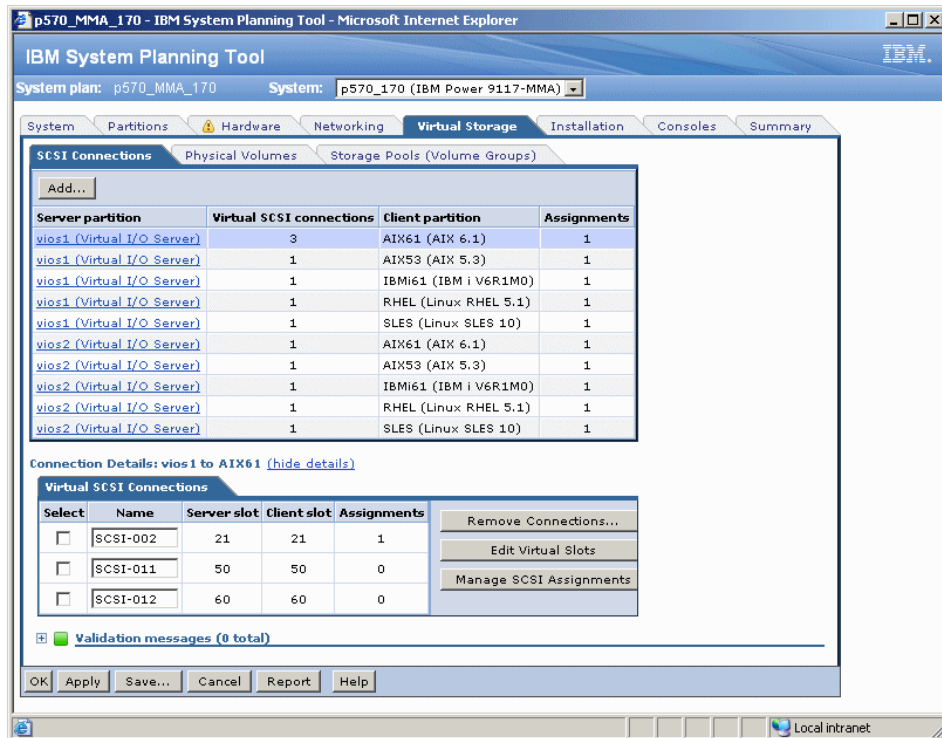


Figure 8-4 The SPT SCSI connections window

Figure 8-5 shows the **Edit Virtual Slots** window. Here the two Virtual I/O Servers are listed on the left side and the client partitions on the right. The maximum number of virtual adapters is increased to 100. In the SCSI area you can check that the server adapters from the Virtual I/O Server matches the client partition aix61.

All values in fields can be edited to suit your numbering convention.

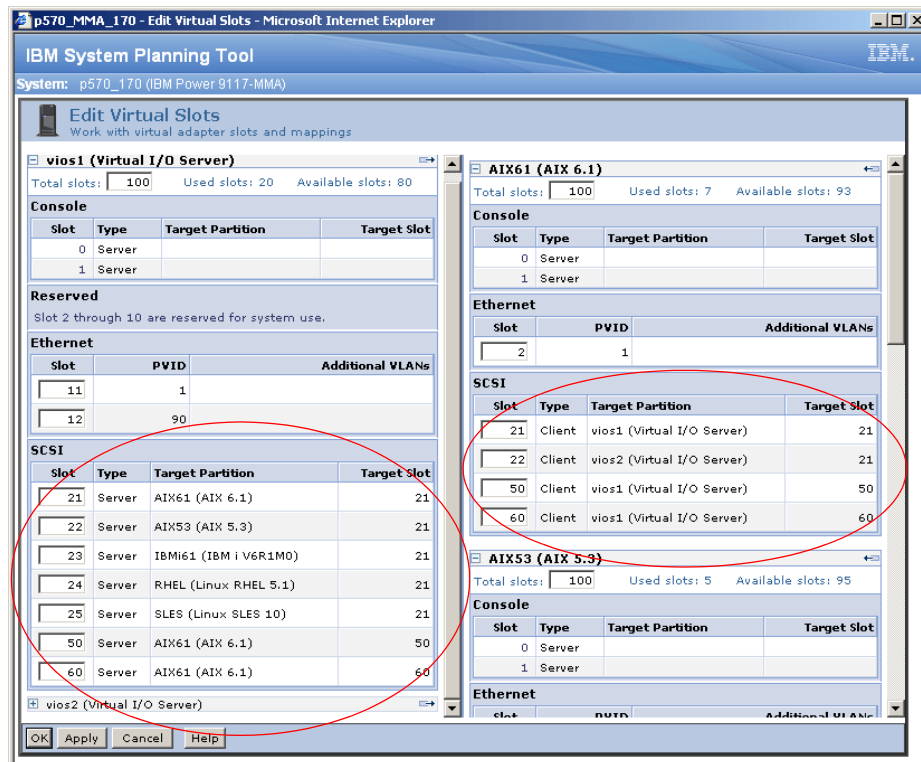


Figure 8-5 The SPT Edit Virtual Slots window

When all elements of the System Plan are done, it can be imported to the HMC to be deployed. Figure 8-6 shows the HMC with the imported SPT System Plan ready to be deployed.

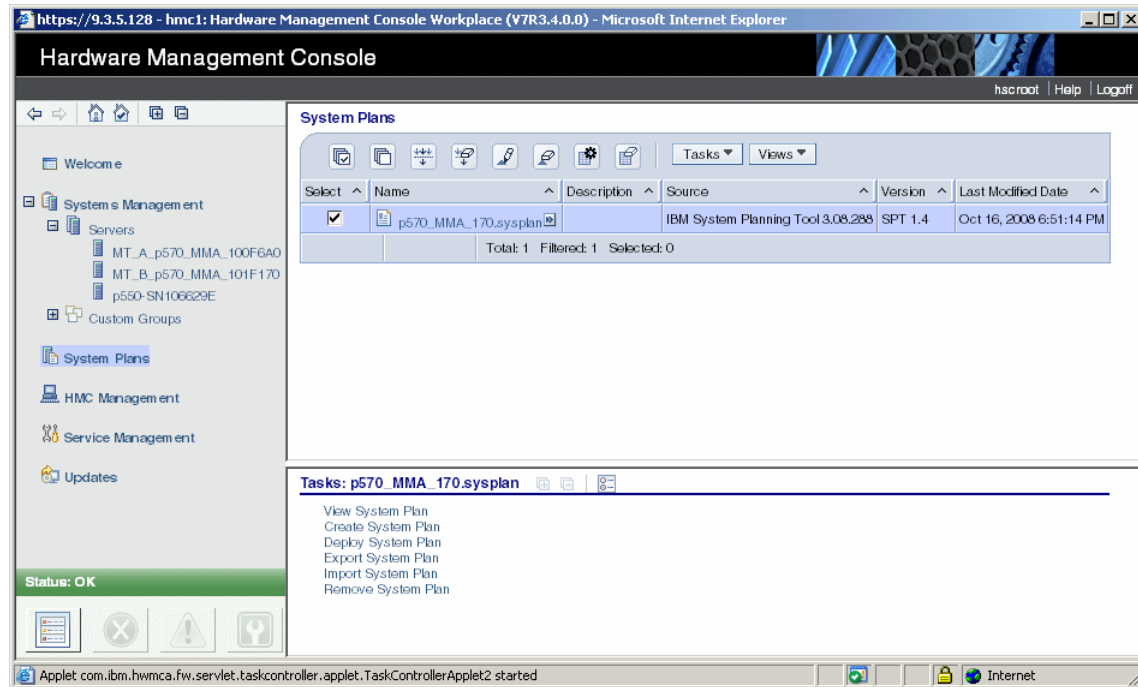


Figure 8-6 The SPT System Plan ready to be deployed

Click on **Deploy System Plan** to start the Deploy System Plan Wizard and follow the directions. Figure 8-7 shows the first menu screen where the System Plan and target managed system are selected.

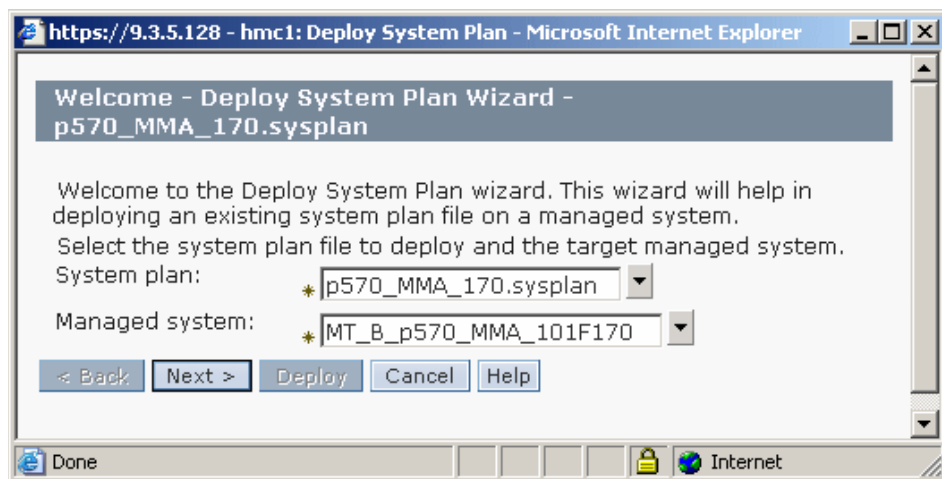


Figure 8-7 Deploy System Plan Wizard

The next step is validation. The selected System Plan is validated against the target Managed System. If the validation Status reports Failed, the cause is usually found in the Validation Messages. The most common is mismatch between SPT configuration and the Managed System configuration. The validation screen is shown in Figure 8-8.

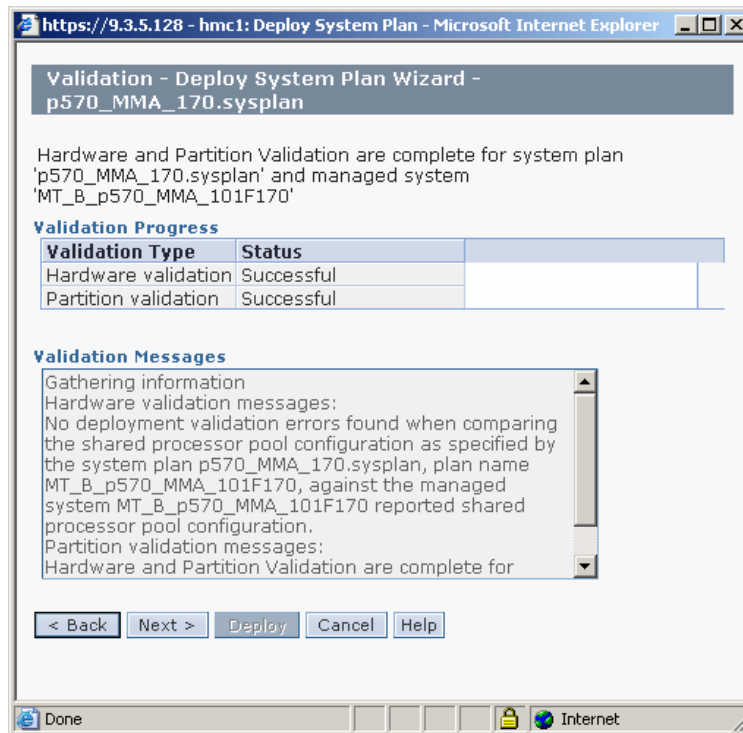


Figure 8-8 The System Plan validation screen

Next the partition profiles to be deployed are selected as shown in Figure 8-9. In this case all partition profiles should be deployed.

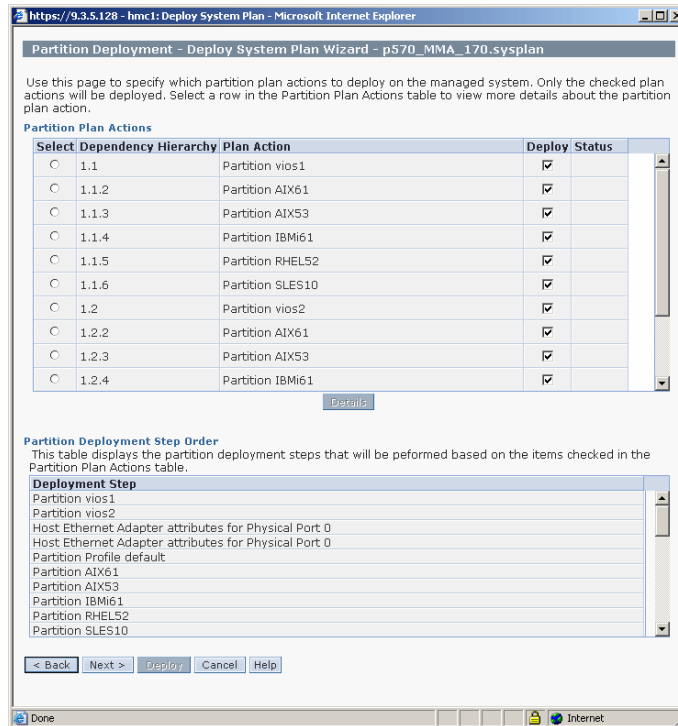


Figure 8-9 The Partition Deployment menu

In the next menu you have the option to install the Virtual I/O Server operating environment. Uncheck the Deploy box if the operating system is not to be installed as shown in Figure 8-10. The box is checked by default.

**Note:** The HMC must be prepared with the correct resources for operating system installation. The Virtual I/O Server operating system can be loaded onto the HMC from DVD or NIM using the `OS_insta11` command and defined as a resource using the `defsysplanres` command (or using the HMC graphical user interface, **HMC Management** → **Manage Install resources**).

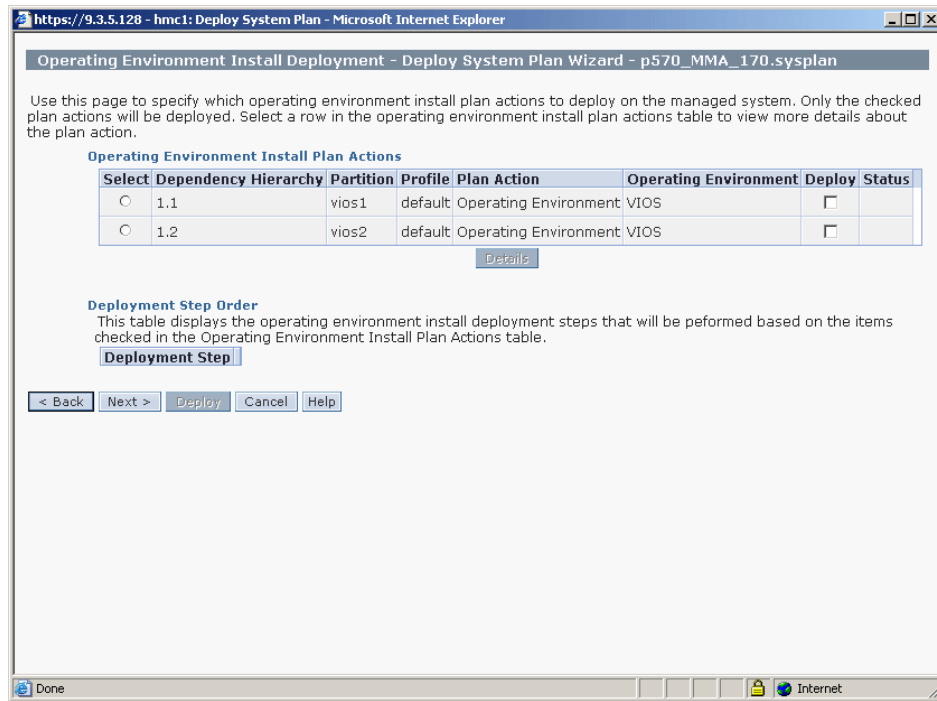


Figure 8-10 The Operating Environment Install Deployment menu

When the preparation steps have been completed, click **Deploy** to start the deployment. The deployment progress is logged as shown in Figure 8-11.

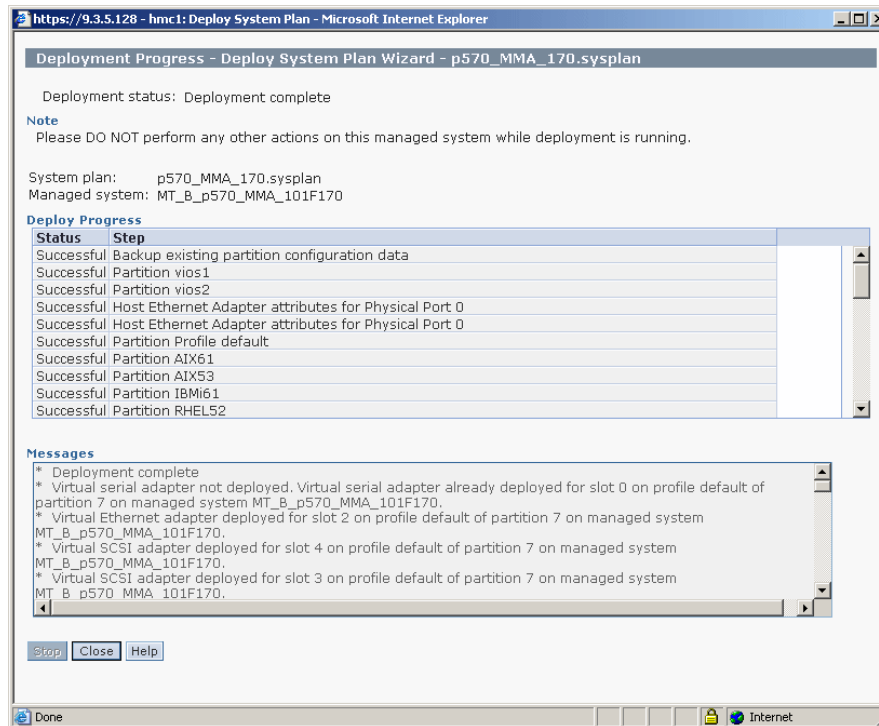


Figure 8-11 The Deployment Progress screen.

Figure 8-12 shows the partition profiles when deployed on the HMC. These partitions are now ready for installation of the operating systems.

**Tip:** At the time of writing SPT (Version 3.08.296), SPT does not support SCSI server adapters set to “Any can connect” like the adapters used for virtual optical and virtual tape devices. To reserve the slots for these devices you can create the SCSI client-server connections to one partition in SPT. After the profiles have been deployed you can change the server adapter to “Any client partition can connect” and add client adapters to the remaining client partitions.

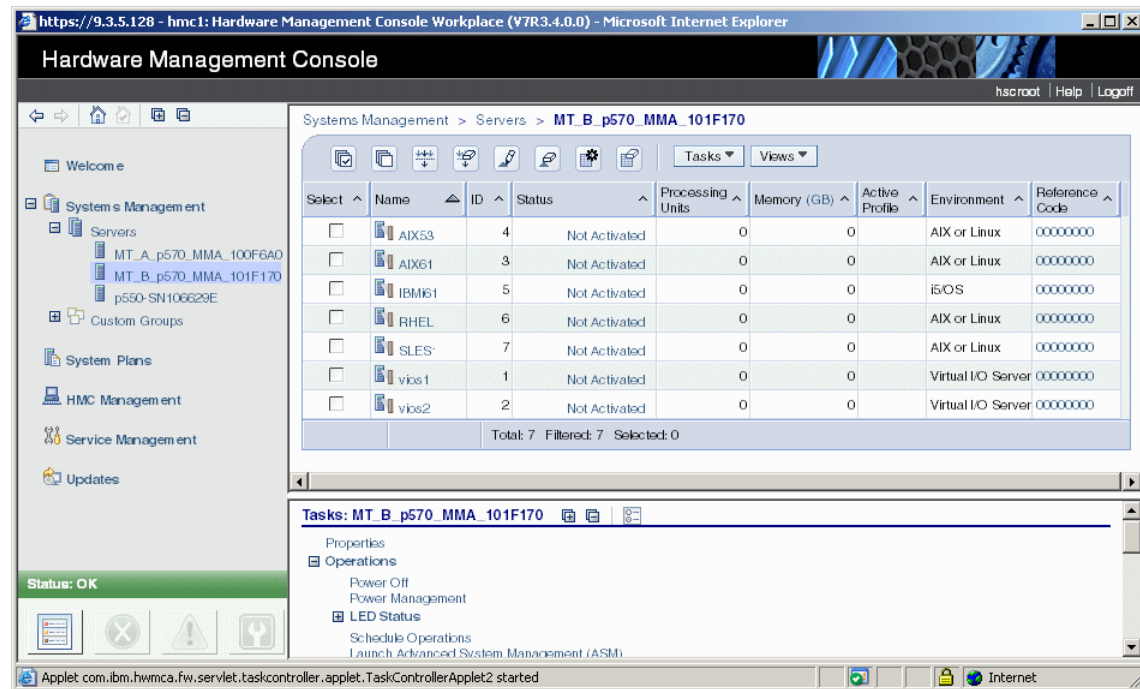



Figure 8-12 Partition profiles deployed on the HMC

All profiles are created with physical and virtual adapter assignments. In this scenario the operating system for the Virtual I/O Servers or any of the client partitions was not installed in the deployment process. After the System Plan has been deployed and the configuration has been customized, a System Plan should be created from the HMC.

The HMC generated System plan is an excellent documentation of the installed system. This System Plan can also be used as backup of the managed system configuration.

**Important:** If the first page of the System Plan is marked:

 This system plan is not valid for deployment.

it can not be used to restore the configuration.



**Note:** In case a System Plan cannot be generated using the HMC graphical user interface, the HMC command may be used:

```
HMC restricted shell> mksysplan -m <managed system> -f\  
<filename>.sysplan --noprobe
```

### 8.1.3 Initial setup checklist

This section shows a high level listing of common steps for an initial setup of a new system using SPT. You should customize the list to fit your environment.

1. Make a System Plan from the HMC of the new system.

Delete the pre installed partition if the new system comes with such a partition.

This System Plan is a baseline for configuring the new system. It will have the adapters slot assignment, CPU and memory configurations.

2. Export the System Plan from the HMC into SPT.

In SPT the file must be converted to SPT format.

3. Complete the configuration as much as possible in SPT.

- Add one Virtual I/O Server partition if using virtual I/O.
- Add one more Virtual I/O Server for a dual configuration, if required.
  - Dual Virtual/I/O Server provides higher serviceability.
- Add the client partition profiles.
- Assign CPU and memory resources to all partitions.
- Create the required configurations for storage and network in SPT.
- Add virtual storage as local disks or SAN disks.

**Note:** At the time of writing SPT (Version 3.08.296) does not support NPIV configurations.

- Configure SCSI connections for MPIO or mirroring if using a dual Virtual I/O Server configuration.
- Configure virtual networks and SEA for attachment to external networks.
- For dual Virtual I/O Server configuration, configure SEA failover or Network Interface Backup (NIB) as appropriate for virtual network redundancy.
- Assign virtual IVE network ports if an IVE adapter is installed.

- Create a virtual server adapter for virtual DVD and for virtual tape if a tape drive is installed.
- Apply your slot numbering structure according to your plan.
- 4. Import the SPT System Plan into the HMC and deploy it to have the profiles generated. Alternatively profiles can be generated directly on the HMC.
- 5. If using SAN disks, create and map them to the host or host group of the Fibre Channel adapters.
  - If using Dual Virtual I/O Servers the `reserve_policy` must be changed from `single_path` to `no_reserve`.
  - SAN disks must be mapped to all Fibre Channel adapters that will be target in Partition Mobility.
- 6. Install the first Virtual I/O Server from DVD or NIM.
  - Upgrade the Virtual I/O Server if updates are available.
  - Mirror the rootvg disk.
  - Create or install SSH keys.

The SSH subsystem is installed in the Virtual I/O Server by default.
  - Configure time protocol services.
  - Add users.
  - Set security level and firewall settings if required.
- 7. Configure an internal network connected to the external network by configuring a Shared Ethernet Adapter (SEA).
  - Consider adding a separate virtual adapter to the Virtual I/O Server to carry the IP address instead of assigning it to the SEA.
- 8. Create a backup of the Virtual I/O Server to local disk using **backupios** command
- 9. Map disks to the client partitions with the **mkvdev** command.
  - Map local disks or local partitions.
  - Map SAN disks.
  - Map SAN NPIV disks.
- 10. Map the DVD drive to a virtual DVD for the client partitions using the **mkvdev** command.
- 11. If available, map the tape drive to a virtual tape drive for the client partitions using the **mkvdev** command.
- 12. Add a client partition to be NIM server to install the AIX and Linux partitions. If a NIM server is already available, skip this step.

- Boot a client partition to SMS and install AIX from the virtual DVD.
  - Configure NIM on the client partition.
  - Let the NIM resources reside in a separate volume group. The rootvg volume group should be kept as compact as possible.
13. Copy the base mksysb image to the NIM server and create the required NIM resources.
  14. If using dual Virtual I/O Servers, do a NIM install of the second Virtual I/O Server from the base backup of the first Virtual I/O Servers. If a single Virtual I/O Server is used, jump to step number 20.
  15. Configure the second Virtual I/O Server.
  16. Map disks from the second Virtual I/O Server to the client partitions.
  17. Configure SEA failover for network redundancy on the first Virtual I/O Server.
  18. Configure SEA failover for network redundancy on the second Virtual I/O Server.
  19. Test that SEA failover is operating correctly.
  20. Install the operating system on the client partitions using NIM or the virtual DVD.
    - Configure NIB if this is used for network redundancy.
    - If using MPIO, change the hcheck\_interval parameter with the **chdev** command to have the state of paths updated automatically.
    - Test that NIB failover is configured correctly in client partitions if NIB is used for network redundancy.
    - Test mirroring if this is used for disk redundancy.
  21. Create a system backup of both Virtual I/O Servers using the **backupios** command.
  22. Document the Virtual I/O Server environment.
    - List virtual SCSI, NPIV and network mappings with the **lsmap** command.
    - List network definitions
    - List security settings.
    - List user definitions.
  23. Create a system backup of all client partitions.
  24. Create a System Plan of the installed configuration from the HMC as documentation and backup.
  25. Save the profiles on the HMC. Click on the Managed System and go to **Configuration** → **Manage Partition Data** → **Backup**. Enter the name of your

profile backup. This backup is a record of all partition profiles on the Managed System.

26. Back up HMC information to DVD, to a remote system, or to a remote site. Go to **HMC Management** → **Back up HMC Data**. In the menu, select target for the backup and follow instructions. The backup contains all HMC settings like user, network, security and profile data.
27. Start collecting performance data. It is valuable to collect long term performance data to have a baseline of performance history.



# Automated management

This chapter provides information on how to automate management functions for Virtual I/O Servers on an HMC. These operations are discussed:

- ▶ Automating remote operations
- ▶ Remotely powering a Power System on and off
- ▶ Remotely starting and stopping logical partitions
- ▶ Scheduling jobs on a Virtual I/O Server

## 9.1 Automating remote operations

in an environment with several partitions it is sometimes convenient to be able to start and stop the partitions from the HMC command line.

You can make the startup of the Virtual I/O Servers easier by placing them in a system profile and starting this System Profile. These system profiles contain logical partitions and an associated partition profile to use.

The menu to create a System Profile is shown in Figure 9-1. To access the menu click on a Managed System, open the Configuration menu and select Manage System Profiles. In this menu you can select New to add a System Profile. Give the System Profile a name and select partitions to be included in the profile.

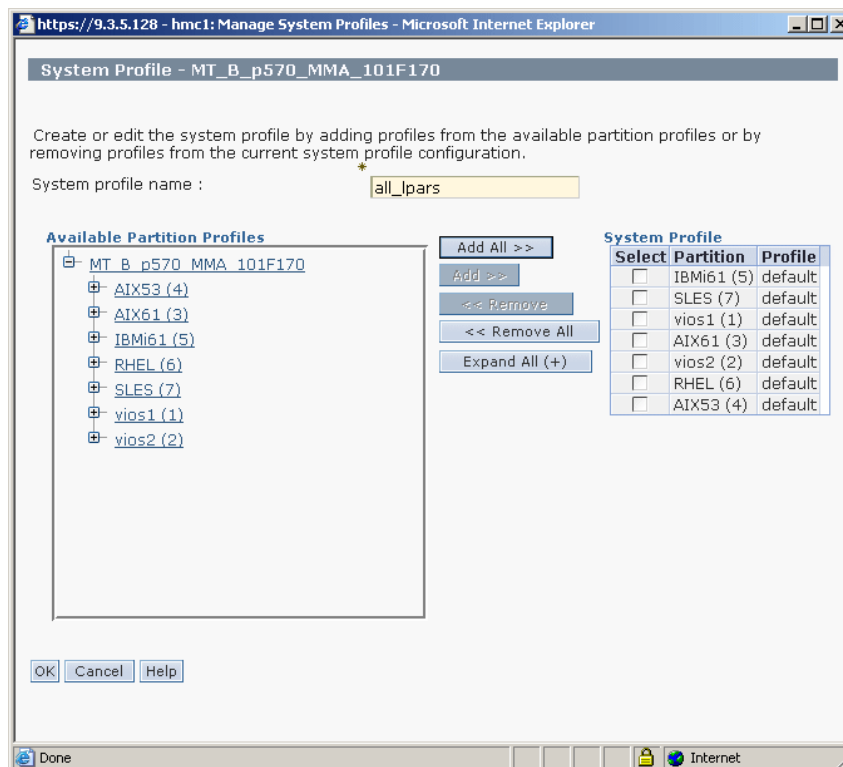


Figure 9-1 Creating a System Profile on the HMC

**Note:** When a Virtual I/O Server is part of a system profile, the system will automatically start this partition first.

For more information about system profiles, see the IBM Systems Hardware Information Center at:

<http://publib.boulder.ibm.com/infocenter/eserver/v1r3s/topic/iphbl/iphblmanage.sprofile.htm>

The `ssh` command must be used to run HMC commands remotely. In addition remote command execution must be enabled on the HMC. It is found in the HMC Management panel as Remote Command Execution. Tick off the box to enable as shown in Figure 9-2.

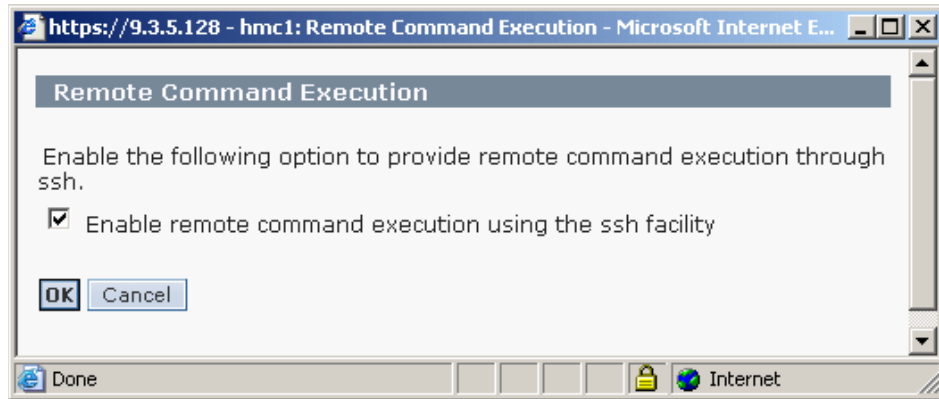


Figure 9-2 The HMC Remote Command Execution menu

From a central control console running SSH, you can issue remote commands to the HMC to perform all of the operations needed to power on a system, start a System Profile for the Virtual I/O Servers and all of the AIX, IBM i or Linux virtual clients. To be able to run HMC commands from a management console, perform an SSH key exchange from the management console onto the HMC.

The procedure for this is in the IBM Systems Hardware Information Center at:

<http://publib.boulder.ibm.com/infocenter/eserver/v1r3s/topic/iphai/settingupsecurescriptexecutionsbetweensshclientsandthehmc.htm>

The HMC commands provided in the following sections are examples of how to automate the start and were accurate at the time of writing. Check that the commands and syntax have not changed for your environment.

## 9.1.1 Remotely powering a Power Systems server on and off

Use the **chsysstate** command on the HMC to power the system on or off. To power on a system to partition standby, run the following command, where the managed system name is the name of the server as shown on the HMC:

```
chsysstate -m <managed system name> -o onstandby -r sys
```

**Tip:** The `lssyscfg -r sys -F name` command can be used to list the managed systems.

**Tip:** To monitor the status of the server startup, use the `lsrefcode` command and check the LED status codes. An example of this command is:

```
lsrefcode -r sys -m <managed system name> -F refcode
```

To power the system off immediately, run the following command:

```
chsysstate -m sys1 -r sys -o off --immed
```

## 9.1.2 Remotely starting and stopping logical partitions

Run the following command to activate all partitions in the System Profile named `all_lpars`:

```
chsysstate -m <managed system name> -o on -r sysprof -n all_lpars
```

**Tip:** Use the `lsrefcode` command to monitor the state of the partitions being started, for example:

```
lsrefcode -r lpar -m <managed system name> -F lpar_name,refcode
```

**Note:** When there are Virtual I/O Servers in the System Profile, these will automatically be activated before client partitions. If client partitions appear to be started first, they will wait for the Virtual I/O Servers to be started.

Run the following command to shut down a partition immediately:

```
chsysstate -m <managed system> -r lpar -o shutdown -n <lpar name> --immed
```

**Tip:** Find more information about the **chsysstate** command with `man chsysstate` command on the HMC.



## 9.2 Scheduling jobs on the Virtual I/O Server

Starting with Virtual I/O Server Version 1.3, the **crontab** command is available to enable you to submit, edit, list, or remove cron jobs. A cron job is a command run by the cron daemon at regularly scheduled intervals, such as system tasks, nightly security checks, analysis reports, and backups.

With the Virtual I/O Server, a cron job can be submitted by specifying the **crontab** command with the **-e** flag. The **crontab** command invokes an editing session that enables you to modify the padmin users' crontab file and create entries for each cron job in this file.

**Note:** When scheduling jobs, use the padmin user's crontab file. You cannot create or edit other users' crontab files.

When you finish creating entries and exit the file, the **crontab** command copies it into the `/var/spool/cron/crontabs` directory and places it in the padmin file.

The following syntax is available to the **crontab** command:

```
crontab [-e padmin | -l padmin | -r padmin | -v padmin]
```

- e padmin** Edits a copy of the padmin's crontab file. When editing is complete, the file is copied into the crontab directory as the padmin's crontab file.
- l padmin** Lists padmin's crontab file.
- r padmin** Removes the padmin's crontab file from the crontab directory.
- v padmin** Lists the status of the padmin's cron jobs.





# High-level management

This chapter describes the following IBM high-level management tools:

- ▶ IBM Systems Director
- ▶ Cluster Systems Management

## 10.1 IBM Systems Director

IBM Systems Director 6.1 is a platform management foundation that streamlines the way physical and virtual systems are managed across a multi-system environment. Leveraging industry standards, IBM Systems Director supports multiple operating systems and virtualization technologies across IBM and non-IBM platforms. IBM Systems Director 6.1 is an easy to use, point and click, simplified management solution.

Through a single user interface, IBM Systems Director provides consistent views for visualizing managed systems, determining how these systems relate to one another while identifying their individual status, thus helping to correlate technical resources with business needs.

IBM Systems Director's Web and command-line interfaces provide a consistent interface focused on these common tasks:

- ▶ Discovering, navigating and visualizing systems on the network with the detailed inventory and relationships to the other network resources
- ▶ Notifying users of problems that occur on systems and the ability to navigate to the source of the problem
- ▶ Notifying users when systems need updates, and distributing and installing updates on a schedule
- ▶ Analyzing real-time data for systems, and setting critical thresholds that notify the administrator of emerging problems
- ▶ Configuring settings of a single system, and creating a configuration plan that can apply those settings to multiple systems
- ▶ Updating installed plug-ins to add new features and function to the base capabilities
- ▶ Managing the life cycle of virtual resources

### Plug-ins included with Systems Director

Base plug-ins provided with Systems Director deliver core capabilities to manage the full life cycle of IBM Server, storage, network and virtualization systems. The base plug-ins include:

<b>Discovery Manager</b>	Discovers virtual and physical systems and related resources.
<b>Status Manager</b>	Provides health status, alerts and monitors of system resources.
<b>Update Manager</b>	Notifies, downloads and installs updates for systems.
<b>Automation Manager</b>	Performs actions based on system events.

- Configuration Manager** Configures one or more systems resource settings.
- Virtualization Manager** Creates, edits, relocates and deletes virtual resources.
- Remote Access Manager** Provides a remote console, a command line and file transfer features to target systems.

### **Additional Plug-ins for Systems Director**

Systems Director allows you to extend the base platform with additional plug-ins, separately installed.

#### **Active Energy Manager**

Energy and thermal monitoring, reporting, capping and control, energy thresholds

#### **Tivoli Provisioning Manager for OS Deployment**

Automated deployment of Windows Server and Vista, Linux, Solaris and MacOS images

#### **BladeCenter Open Fabric Manager**

Management of I/O and network interconnects for up to 100 BladeCenter chassis

#### **Virtual image manager**

Customization and deployment of VMware ESX, Xen, and PowerVM virtual images

There are four main components of a Systems Director 6.1 deployment:

- ▶ Systems Director Server acts as the central management server. It may be deployed on AIX, Windows Server 2003, or Red Hat or SUSE Linux and may be easily configured for high availability (HA) operations. Apache Derby (for small environments) or IBM DB2, Oracle, or SQL Server may be employed for management databases.
- ▶ Management Console provides a common browser-based interface for administrators to the full set of Systems Director tools and functions for all supported platforms.
- ▶ Common and Platform Agents reside on managed servers running supported operating systems and hypervisors. Common Agent enables use of the full suite of Systems Director 6.1 services. Platform Agent implements a subset of these. It provides a small footprint option for servers that perform limited functions or are equipped with less powerful processors.
- ▶ Agentless support is provided for x86 servers that run older versions of Windows or Linux operating systems, along with other devices that conform to Distributed Component Object Model (DCOM), Secure Shell (SSH) or Simple Network Management Protocol (SNMP) specifications.

A single Systems Director 6.1 server can manage thousands of physical and/or logical servers equipped as Common or Platform Agents. There is no limit to the number of SNMP devices that can be managed.

The overall Systems Director 6.1 environment is summarized in Figure 10-1

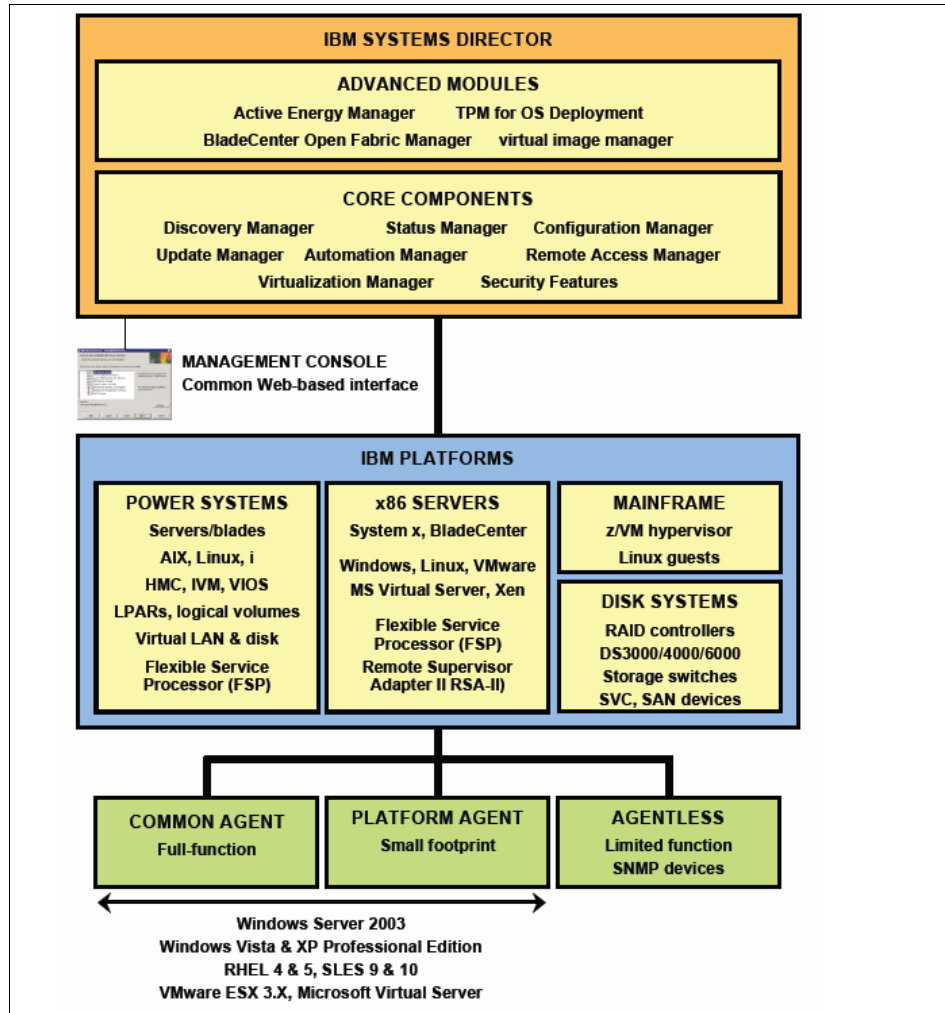


Figure 10-1 IBM Systems Director Environment

### Choosing the management level for managed systems

IBM Systems Director provides different levels of management for managed systems and managed objects. For each managed system, you need to choose

the management level that provides the management functionality you need for that managed system.

IBM Systems Director has three management levels:

- ▶ Agentless Managed: systems without any IBM Systems Director software installed.
- ▶ Platform Agent Managed: systems with Platform Agent installed.
- ▶ Common Agent Managed: systems with Common Agent installed.

These three management levels have different requirements and provide differing levels of management functionality in IBM Systems Director.

IBM Director is provided at no additional charge for use on IBM systems. You can purchase additional IBM Director Server licenses for installation on non-IBM Servers. In order to extend IBM Director capabilities, several extensions can be optionally purchased and integrated.

### 10.1.1 IBM Director installation on AIX

IBM Director can be installed on any AIX or Linux on Power Systems partitions or on Windows Server 2003.

1. Download the installation package from the IBM Systems Director Downloads Web Site at:

<http://www.ibm.com/systems/management/director/downloads/>

2. unzip and extract the contents of the installation package, type the following command:

```
# gzip -cd install_package | tar -xvf -
```

where `install_package` is the file name of the downloaded installation package.

3. Change to the directory in which the installation script is located.

```
# cd /install_files/
```

where `install_files` is the path to the extracted installation files.

4. Run the `./server/dirinstall.server` command as shown in Example 10-1.

#### *Example 10-1 Installing IBM Director*

```
# ./server/dirinstall.server
```

```
*****
```

You must configure the agent manager prior to starting the server.

To configure the agent manager, run

```
/opt/ibm/director/bin/configAgtMgr.sh
```





Starting IBM Director...

The starting process may take a while. Please use `smstatus` to check if the server is active.

7. Check the status of the server:

```
# smstatus -r
```

```
Active
```

## 10.1.2 Log on to IBM Systems Director

After installing IBM Systems Director Server, log on using a Web browser, discover managed systems, and request access to them. Complete the following steps:

1. Point your browser to the following URL, **HOSTNAME** is the IP address of your IBM Director server:  
`https://HOSTNAME:8422/ibm/console`
2. Type the user ID and password that correspond to an authorized IBM Systems Director administrator, you can use root user, click log in.

Figure 10-2 shows the IBM Director login screen.

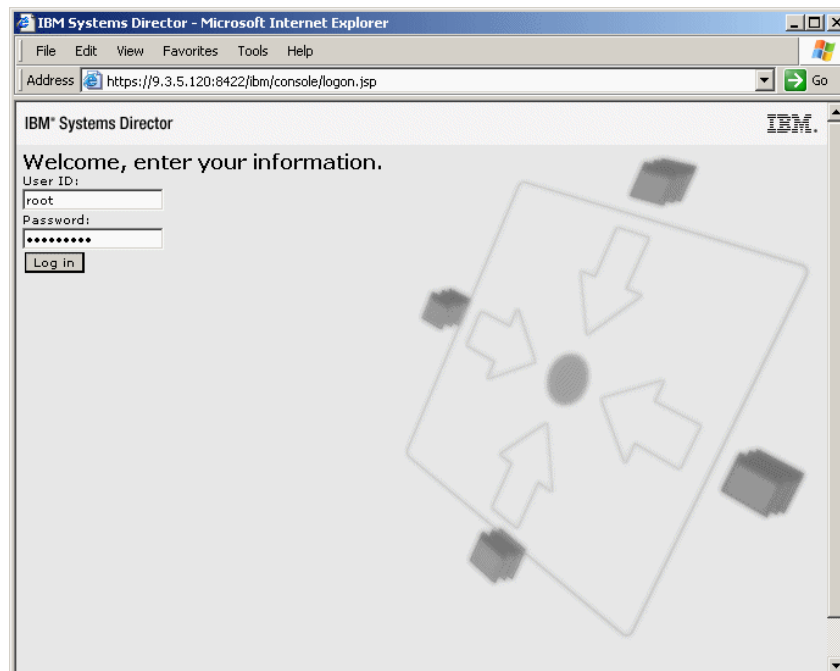


Figure 10-2 IBM Director login screen

3. You are now logged on the IBM Director Server

Figure 10-3 shows the Welcome to IBM Systems Director screen

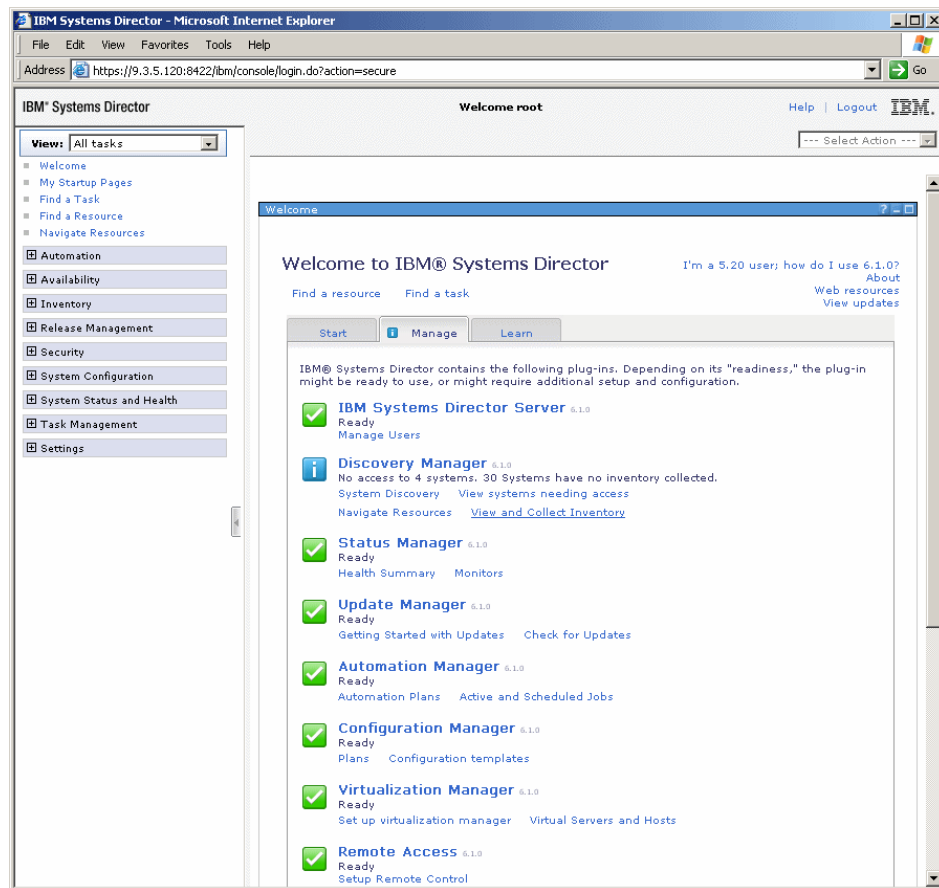


Figure 10-3 Welcome to IBM Systems Director screen

### 10.1.3 Preparing managed systems

You need to configure AIX systems, Virtual I/O Server, and HMC before you can discover and manage them with IBM Systems Director Server. Typically, managed systems are first discovered using the discovery task in IBM Systems Director Web interface. Then, Platform Agent or Common Agent is installed on the managed systems.

## Hardware Management Console

Before discovering Hardware Management Console (HMC) devices, you need to open the Open Pegasus and SLP ports.

In the HMC Navigation Area pane, expand **HMC Management**, Click **HMC Configuration** → **Customize Network settings** → **LAN Adapter**, select the LAN Adapter that is connected to your LAN and click **Details** → **Firewall**, select **Open Pegasus** and click on **Allow Incoming**, select **SLP** and click on **Allow Incoming**,

Figure 10-4 shows the HMC LAN Adapter Details screen.

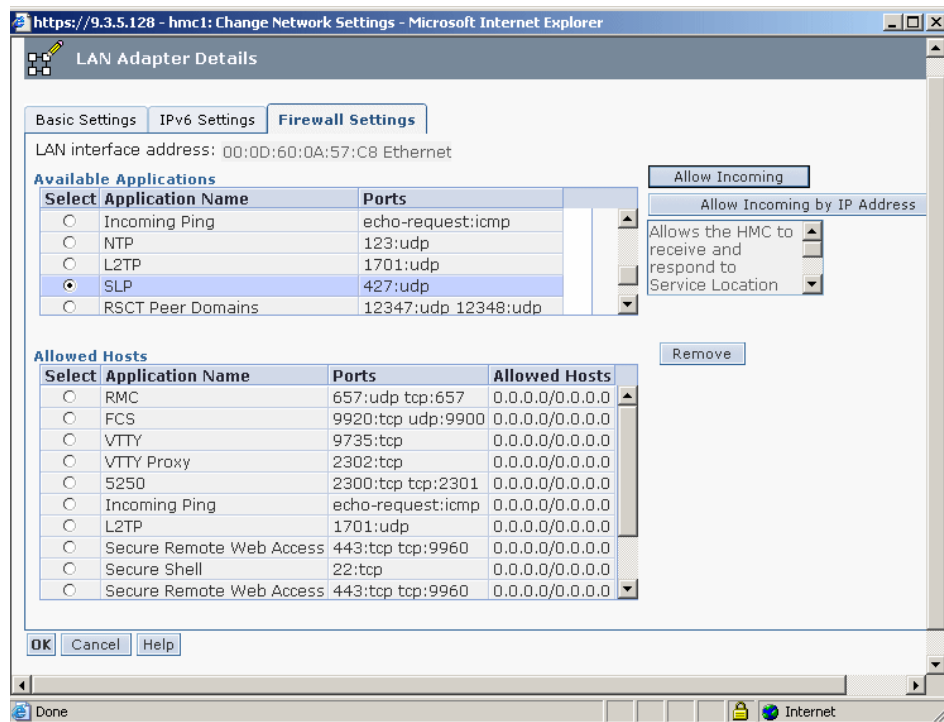


Figure 10-4 HMC LAN Adapter Details

## Virtual I/O Server

Before discovering Virtual I/O Server devices, you need to start the IBM Director agent.

- To list all the attributes associated with the agent configuration:

```
$ cfgsvc -ls DIRECTOR_agent
RESTART_ON_REBOOT
```

## 2. Configure the agent:

```
$ cfgsvc DIRECTOR_agent -attr Restart_On_Reboot=TRUE
```

The Restart\_On\_Reboot attribute set to TRUE specifies to restart the IBM Director agent when the Virtual I/O Server is being rebooted.

## 3. Check the agent configuration

```
$ lssvc DIRECTOR_agent  
RESTART_ON_REBOOT:TRUE
```

## 4. Start the agent

```
$ startsvc DIRECTOR_agent
```

This is the first time starting Director Agent.

Please waiting several minutes for the initial setup...

Starting The LWI Nonstop Profile...

The LWI Nonstop Profile succesfully started. Please refer to logs to check the LWI status.

```
ALR0280I: Enable processing for the following features were successful  
com.ibm.eserver.usmi.tools.CommonInstallEngine_1.0.0
```

```
ALR0282I: The platform needs to be restarted to complete the operation.
```

Stopping The LWI Nonstop Profile...

Waiting for The LWI Nonstop Profile to exit...

Waiting for The LWI Nonstop Profile to exit...

Waiting for The LWI Nonstop Profile to exit...

Stopped The LWI Nonstop Profile.

Starting The LWI Nonstop Profile...

The LWI Nonstop Profile succesfully started. Please refer to logs to check the LWI status.

Running IBM Systems Director Agent feature installation...

Stopping The LWI Nonstop Profile...

Waiting for The LWI Nonstop Profile to exit...

Waiting for The LWI Nonstop Profile to exit...

Waiting for The LWI Nonstop Profile to exit...

Stopped The LWI Nonstop Profile.

Starting cimserver...

Starting cimlistener...

Starting tier1slp...

Starting dirsmpd...

## **Power Systems running AIX**

Before discovering AIX devices, you need to install the IBM Director Common Agent.

1. Download the installation package from the IBM Systems Director Downloads Web Site at:

<http://www.ibm.com/systems/management/director/downloads/>

2. unzip and extract the contents of the installation package, type the following command:

```
# gzip -cd install_package | tar -xvf -
```

where install\_package is the file name of the downloaded installation package.

3. Change to the directory in which the installation script is located.

```
# cd /install_files/
```

where install\_files is the path to the extracted installation files.

4. Run the `./server/dirinstall.agent` command as shown in Example 10-1.

*Example 10-2 Installing IBM Director Common Agent*

---

```
# ./dirinstall.agent
```

```
*****
```

```
*
```

This Program is licensed under the terms of the agreement located in the license

file in the Program's installation license folder or in the license folder on the source media.

By installing, copying, accessing, or using the Program, you agree to the terms of this agreement. If you do not agree to the terms, do not install, copy, access, or use the Program.

```
*****
```

```
Attempting to install openssl.base openssl.license
```

```
+-----+
```

```
Pre-installation Verification...
```

```
+-----+
```

---

5. Start the agent

```
# /opt/ibm/director/agent/bin/lwistart.sh
```

6. Check that the agent is running

```
# /opt/ibm/director/agent/bin/lwistatus.sh
```

```
ACTIVE
```

## 10.1.4 Discover managed systems

Discovery is the process by which IBM Systems Director Server identifies and establishes connections with network-level resources that IBM Systems Director can manage. You can use system discovery or advanced system discovery to identify resources within your environment, collect data about the resources, and establish connections with the resource.

Use System discovery task to discover systems at a specific network address or range of addresses. After a resource has been discovered, it becomes a system that can be managed by IBM Systems Director.

In the IBM Systems Director Web interface navigation area, expand **Inventory**, click **System Discovery**, enter the IP address of the system and Click **Discover** as shown in Figure 10-5.

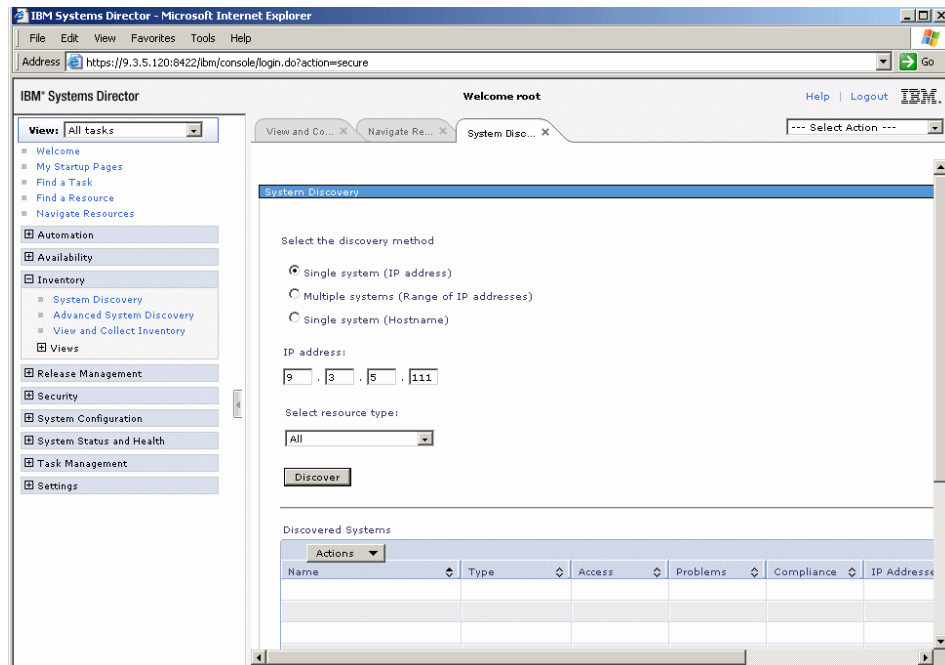


Figure 10-5 IBM Director System Discovery

When discovery is completed, system is displayed in the **Discovered Systems** table as shown in Figure 10-6.

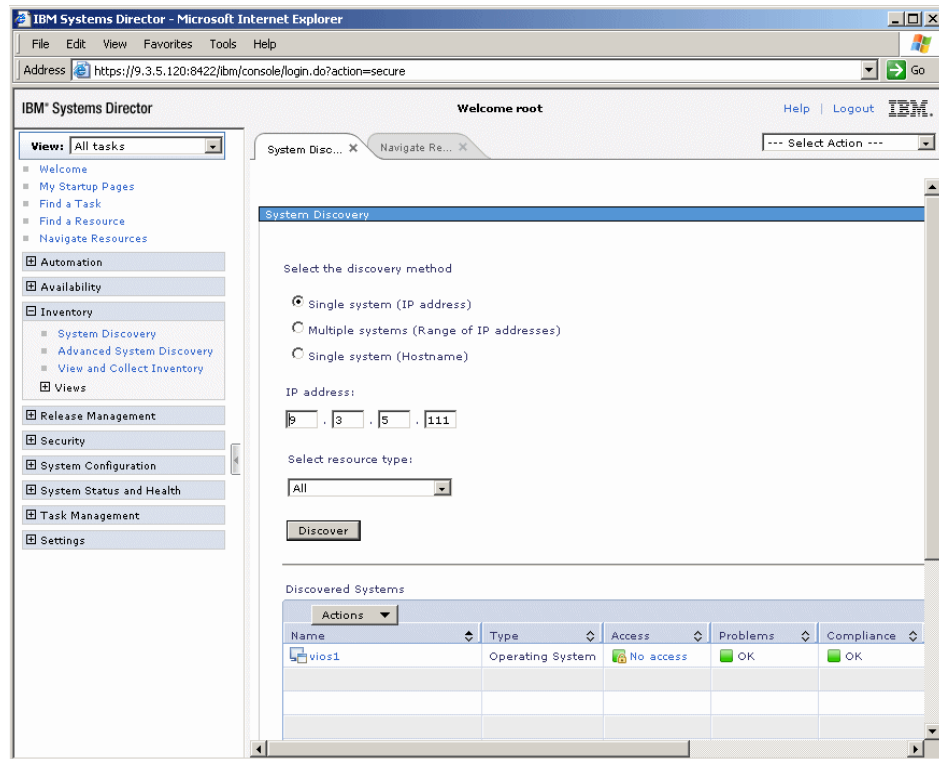


Figure 10-6 IBM Director Discovered Systems table

To manage a resource within an environment, that resource must first be discovered and after access have to be granted.

Click **No Access**, enter the credentials for the system and click **Request Access** as shown in Figure 10-7.

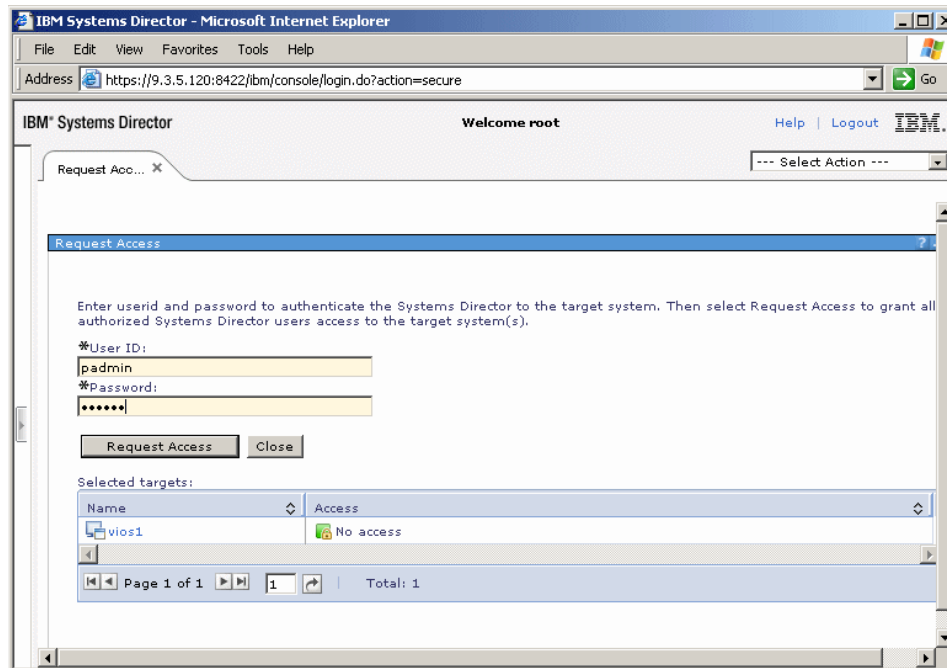


Figure 10-7 IBM Director Request Access

After access is granted, the Access status is **OK** as shown in Figure 10-8.



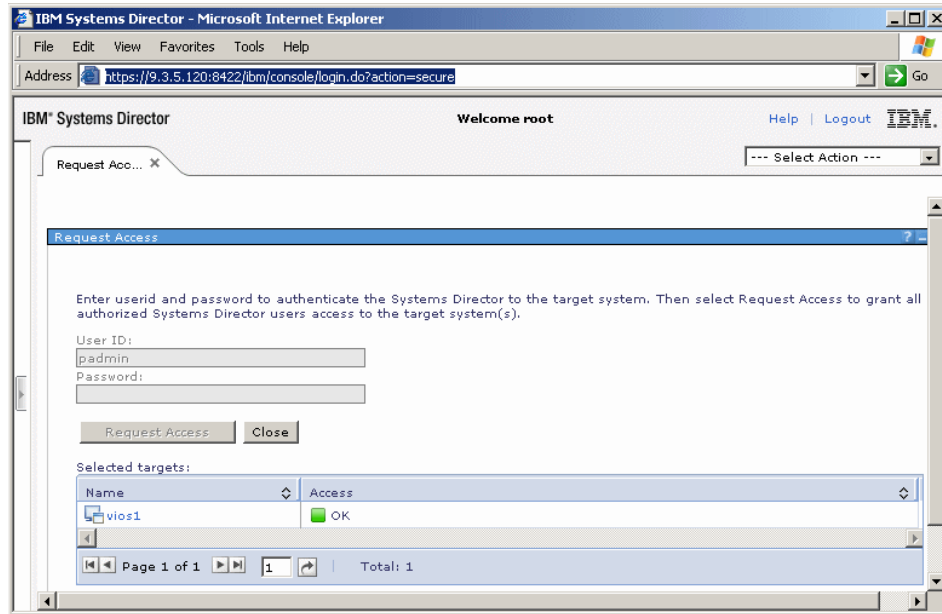


Figure 10-8 IBM Director Access granted

### 10.1.5 Collect inventory data

Inventory collection is the process by which IBM Systems Director Server establishes connections with systems that have already been discovered and collects data about the hardware and software that is currently installed on those resources.

In the IBM Systems Director Web interface navigation area, expand **Inventory**, click **View and Collect Inventory**, in the Target Systems list, select the system for which you want to collect inventory data and click **Collect Inventory** as shown in Figure 10-9.

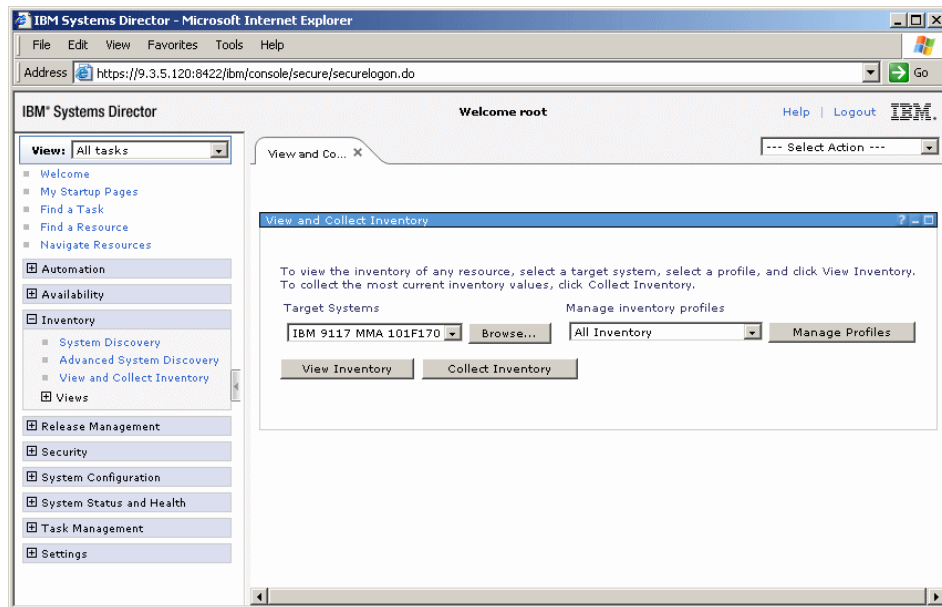


Figure 10-9 IBM Director View and Collect Inventory

Use the Schedule tab to set the inventory collection task to run immediately by checking the **Run Now** box as shown in Figure 10-10.

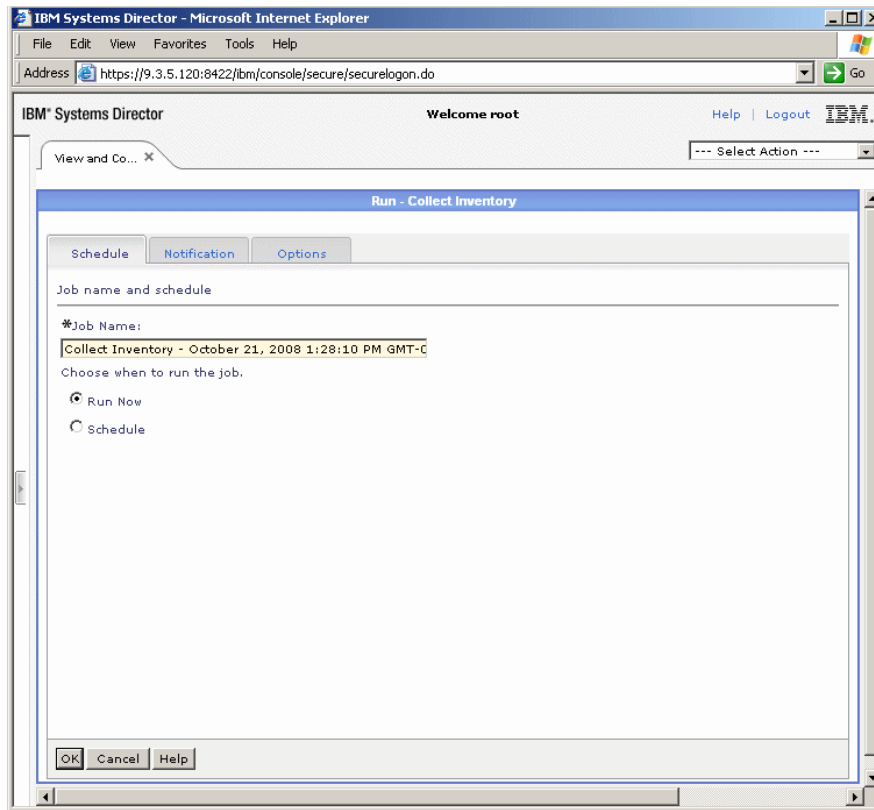


Figure 10-10 IBM Director Run Collect Inventory

Click **OK**, an inventory collection job is created and a message is displayed with buttons and information about the job as shown in Figure 10-11.

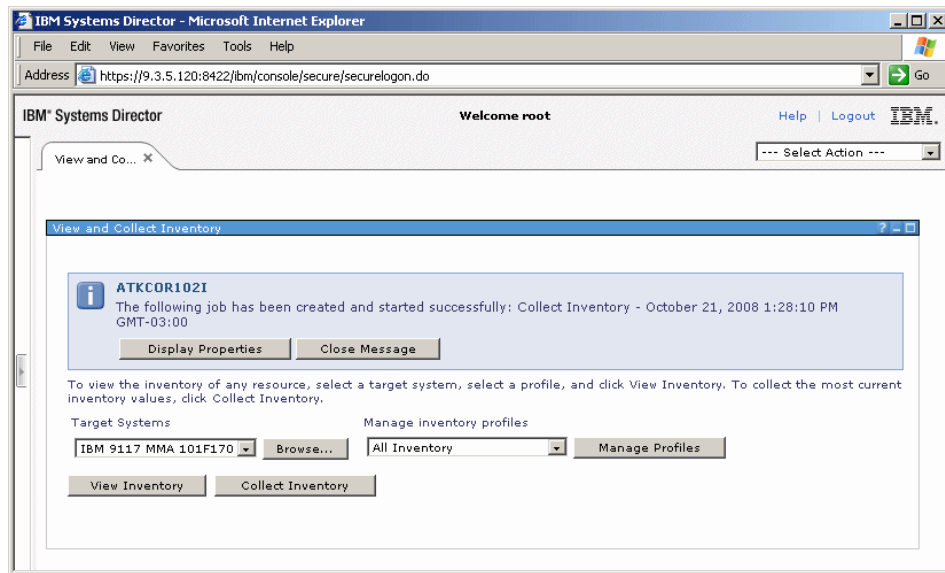


Figure 10-11 IBM Director Collection job

## 10.1.6 View Managed resources

Use Navigate Resources when you want to view the status of managed resources.

In the IBM Systems Director Web interface navigation pane, click **Navigate Resources** as shown in Figure 10-12.

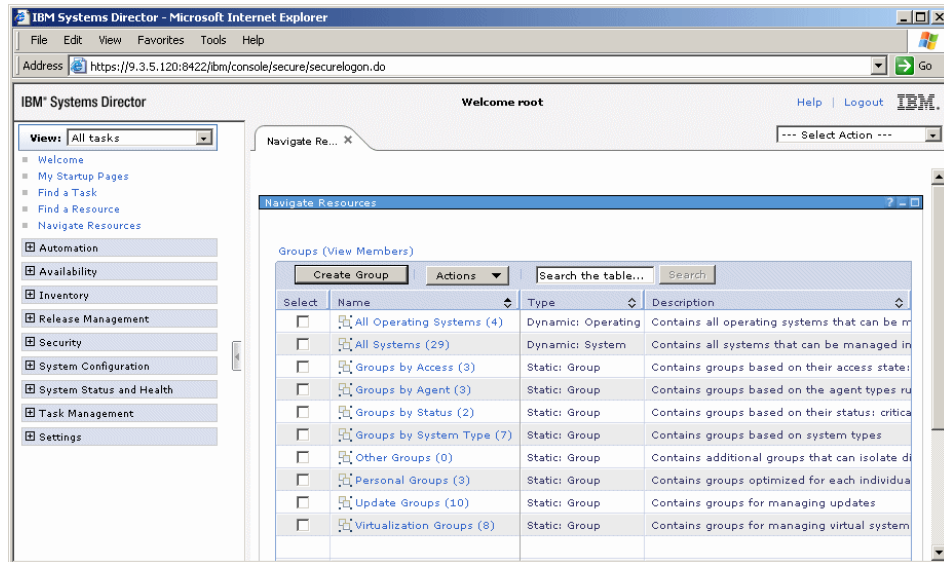


Figure 10-12 IBM Director Navigate Resources

In the Name Groups column, click **All Systems** as shown in Figure 10-11.

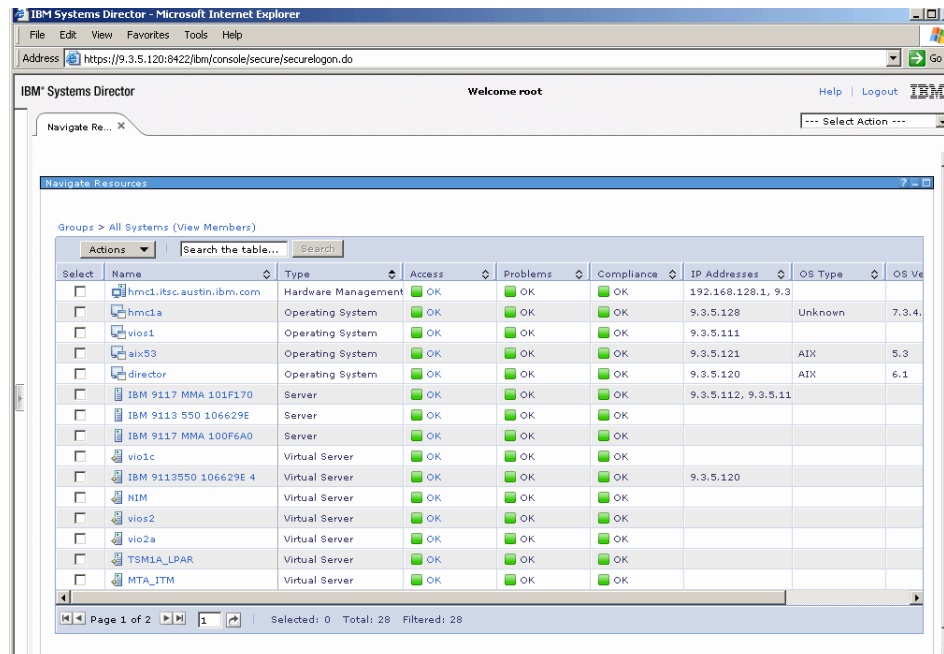


Figure 10-13 IBM Director All Systems (View Members)

## 10.1.7 Power Systems Management summary

You can view a summary of the resources managed by IBM Power Systems and their status. You can also access common management tasks for managing your Power Systems resources. Note that information on this page is refreshed automatically when there are any changes.

To view the Power Systems Management summary, In the IBM Systems Director navigation area, click **Welcome**, on the Welcome page, click **Manage**, and click the **Power Systems Management** section heading. The Power Systems Management summary is displayed as shown in Figure 10-14.

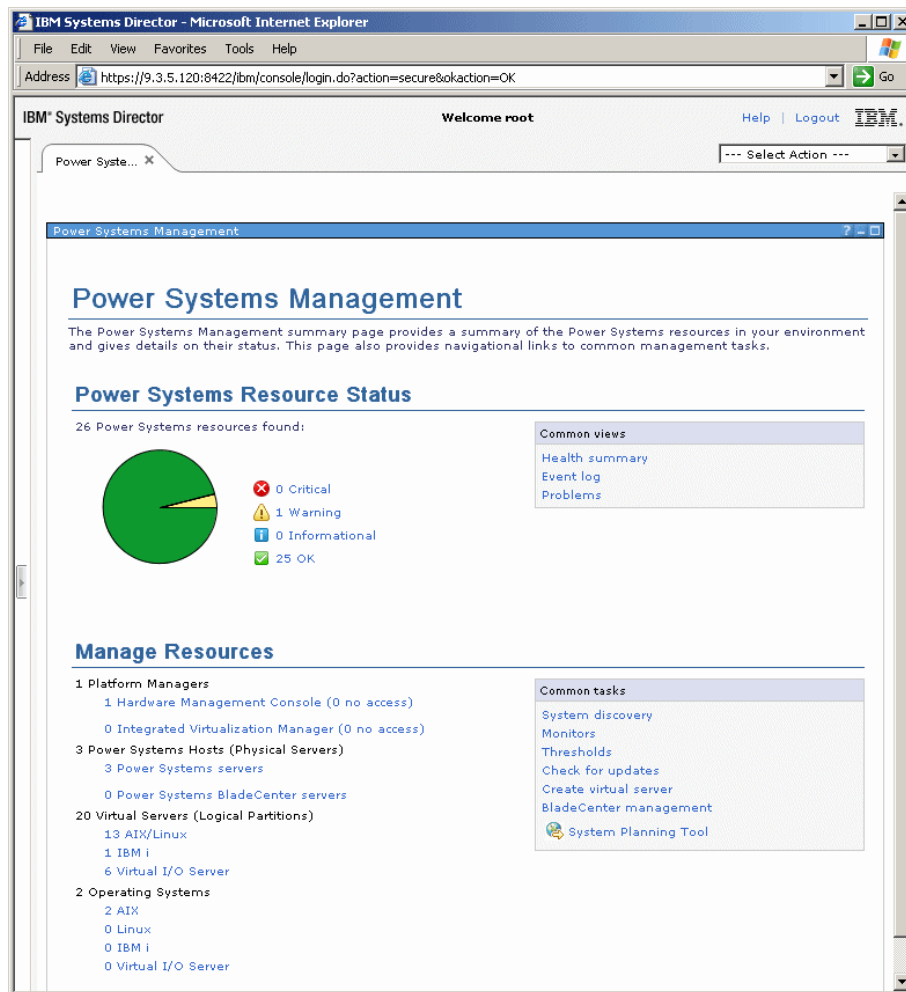


Figure 10-14 IBM Director Power Systems Management summary

This section provides the following information:

Links to the following tasks that you can use to view and manage your resources:

- ▶ Health summary
- ▶ Event log
- ▶ Problems

The **Manage Resources** section provides information about the managed resources as shown in Figure 10-15.

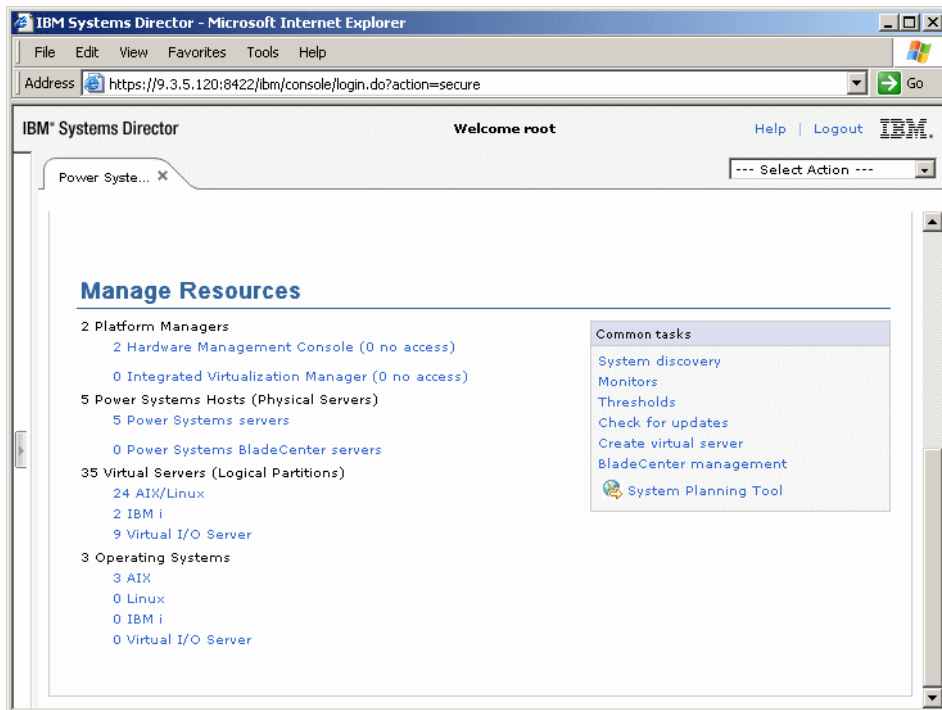


Figure 10-15 IBM Director Power Systems Management Manage resources

IBM Systems Director introduces some new terms, Table 10-1 on page 348 lists terms used in IBM Systems Director:

Table 10-1 Terms for IBM Systems Director

IBM Systems Director term	Power Systems term or concept	Definition
host	server, system, managed system	A physical server that contains physical processors, memory, and I/O resources and which is often virtualized into virtual servers, also known as logical partitions.
virtual server	logical partition, partition	The collection of processor, memory and I/O resources defined to run an operating system and its applications.
utility virtual server	Virtual I/O Server (VIOS)	A virtual server that provide virtualization capabilities for a particular environment.
platform manager	Hardware Management Console (HMC) and Integrated Virtualization Manager (IVM)	A platform manager manages one or more hosts and their associated virtual servers and operating systems. For Power Systems, the platform managers are HMC and IVM.
power on	activate (partition)	Power Systems managed by HMC and IVM, use the term power on with respect to a physical server or host. IBM Systems Director uses the same term, power on, for virtual servers, where Power Systems has used the term activate.
power off	shut down (partition)	Power Systems managed by HMC and IVM, use the term power off with respect to a physical server or host. IBM Systems Director uses the same term, power off, for virtual servers, where Power Systems has used the term shut down.
live relocation	partition mobility, Live Partition Mobility	Moving a running virtual server from one host to another



IBM Systems Director term	Power Systems term or concept	Definition
static relocation	inactive partition mobility, inactive mobility	Moving a virtual server that is powered off from one host to another.
virtual farm	N/A	A virtual farm logically groups like hosts and facilitates the relocation task.

### 10.1.8 IBM Systems Director Virtualization Manager plug-in

With IBM Systems Director virtualization manager, you can work with virtualized environments that are managed by the Hardware Management Console (HMC), the Integrated Virtualization Manager (IVM), Microsoft® Virtual Server, VMware, and Xen virtualization. It is an extension to IBM Director that allows you to discover, visualize, and manage both physical and virtual systems from a single console.

You can use the virtualization manager plug-in to:

- ▶ Work with virtualized environments and tools, including Hardware Management Console (HMC), Integrated Virtualization Manager (IVM), Microsoft Virtual Server, VMware, and Xen virtualization
- ▶ Viewing topology that shows the connections between physical and virtual resources, which can vary dynamically across time
- ▶ Tracking alerts and system status for virtual resources and their resources to easily diagnose problems affecting virtual resources
- ▶ Creating automation plans based on events and actions from virtual and physical resources, such as relocating a virtual server based on critical hardware alerts
- ▶ Create, delete and manage virtual servers and virtual farms for several virtualization technologies in the industry
- ▶ Relocate virtual servers to alternate physical host

To view the Virtualization Manager, in the IBM Systems Director navigation area, click **Welcome**, on the Welcome page, click **Manage**, and click the **Virtualization Manager** section heading. The Virtualization Manager page is displayed as shown in Figure 10-16.

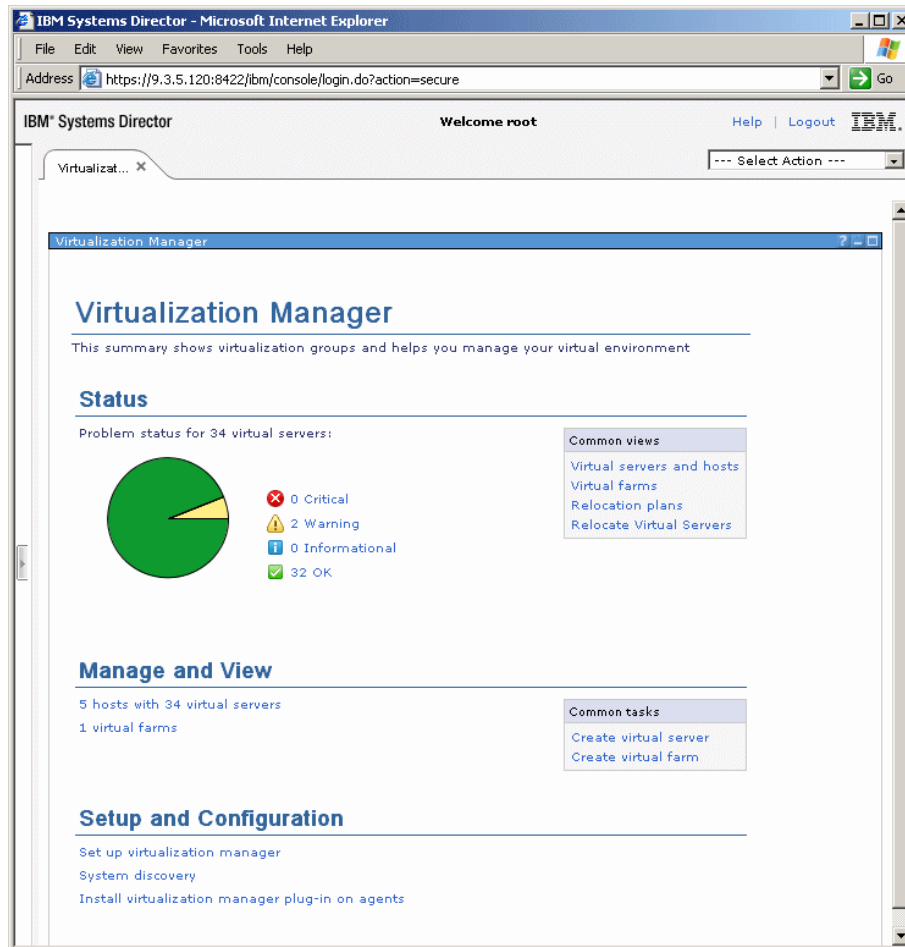


Figure 10-16 IBM Director Virtualization Manager

## 10.1.9 Manage Virtual I/O Server with IBM Systems Director

### Create a group of resources

To make working with a set of resources easier, we create a static group called **PowerVM** which contains all the managed resources used for the redbook.

To create a static group, complete the following steps:

1. In the IBM Systems Director navigation area, click **Navigate Resources** → **Create Group**, click next on the welcome page.

2. On the “name” page, type **PowerVM** for the Name and **“ITSO managed systems”** for the description, click Next
3. On the “Type and Location” page, select **Static** from the Group type list, select **Managed System** from the Member type list, select **Groups** from the Group location list, click Next
4. On the “Define” page, select the resources you want to add to the PowerVM group, click **Add** → **Next**
5. Figure 10-17. show the Group Editor Summary, click **Finish** to create the group.

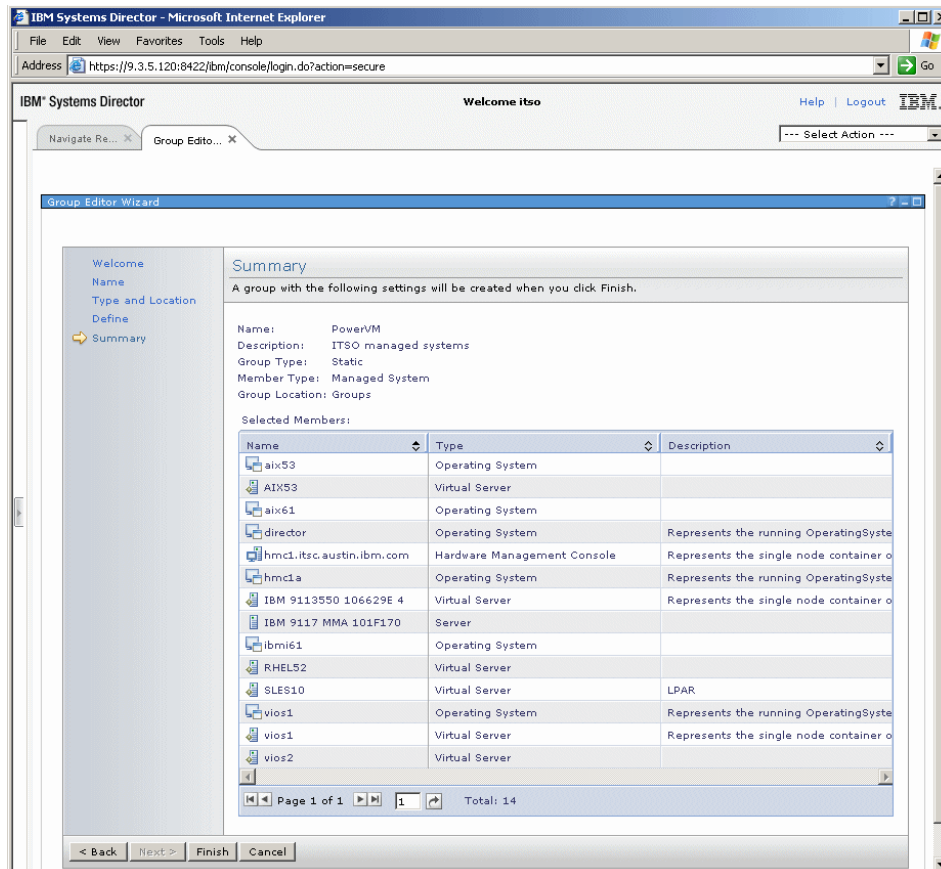


Figure 10-17 IBM Director Create Group Summary

## Create a virtual Server

You can use IBM Systems Director to create and manage virtual servers on the Power system in your environment, it will not install the operating system. It will do the following tasks:

1. Create a Logical Partition on the Power System
2. Create the Virtual SCSI client and Virtual Ethernet adapter on the Logical Partition
3. Create a logical volume on the Virtual I/O Server
4. Create a Virtual SCSI Server adapter on the Virtual I/O Server
5. Map the Logical Volume to the Logical Partition.

In the IBM Systems Director Web interface navigation area, expand **System Configuration**, click **Create Virtual Server**, the Create Virtual Server screen is displayed as shown in Figure 10-18.

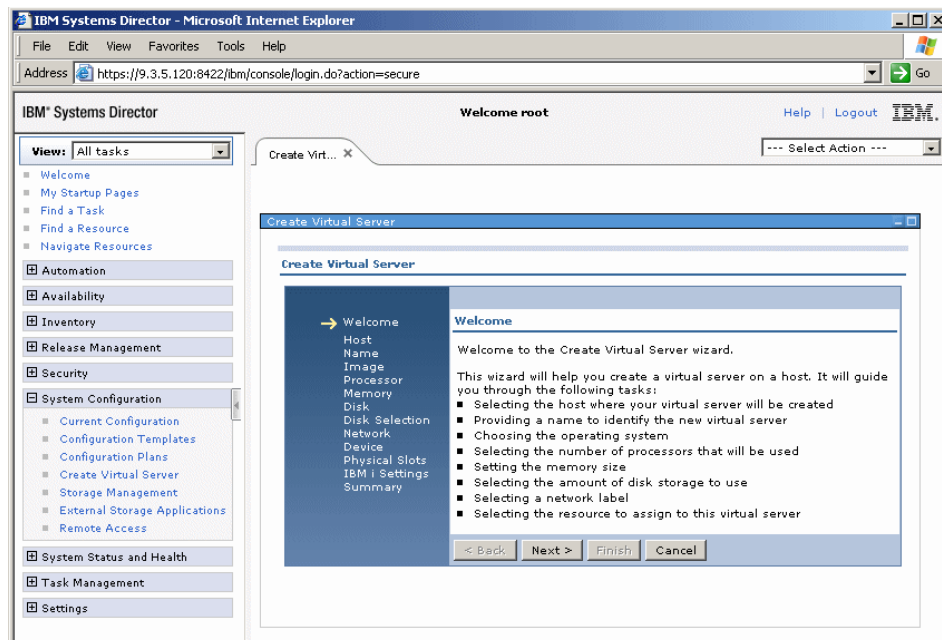


Figure 10-18 IBM Director Create Virtual Server

Click **Next**, select the host where the virtual server will be created **MT\_B\_p570\_MMA\_101F170**, Click **Next**, then type the name of the virtual server that you want to create, and click **Next**, as shown in Figure 10-19.

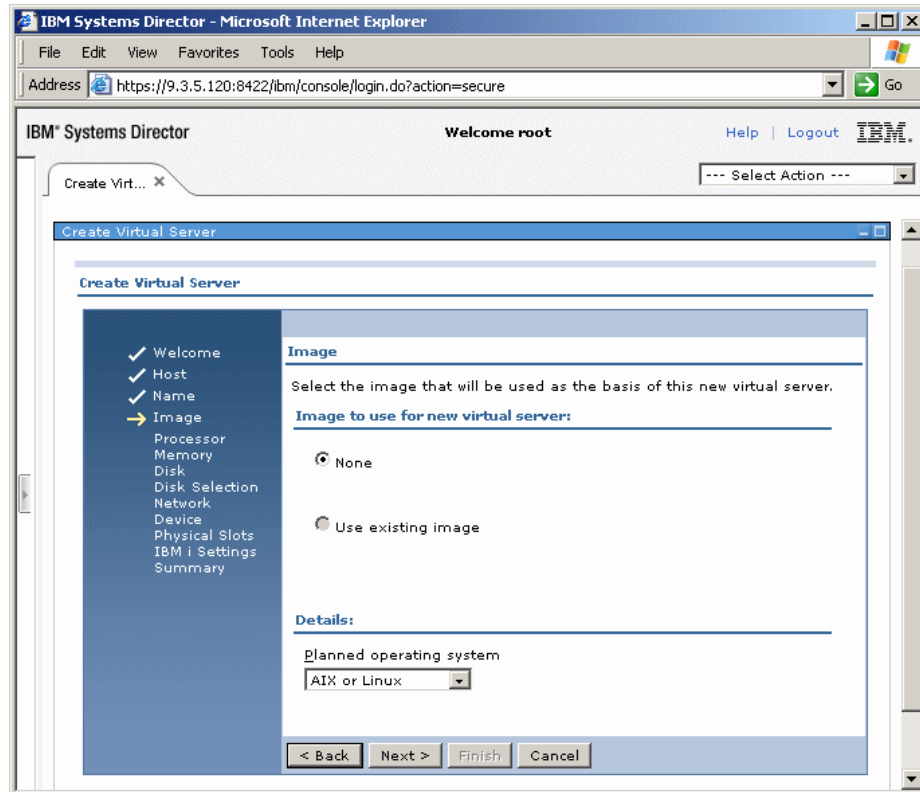


Figure 10-19 IBM Director Create Virtual Server

Select AIX for the **Planned operating system**, Click **Next**, check **Use shared processors** box, assign 1 virtual processor, click **Next**, select the amount of memory to assign, define the disk to be used as shown in Figure 10-20.

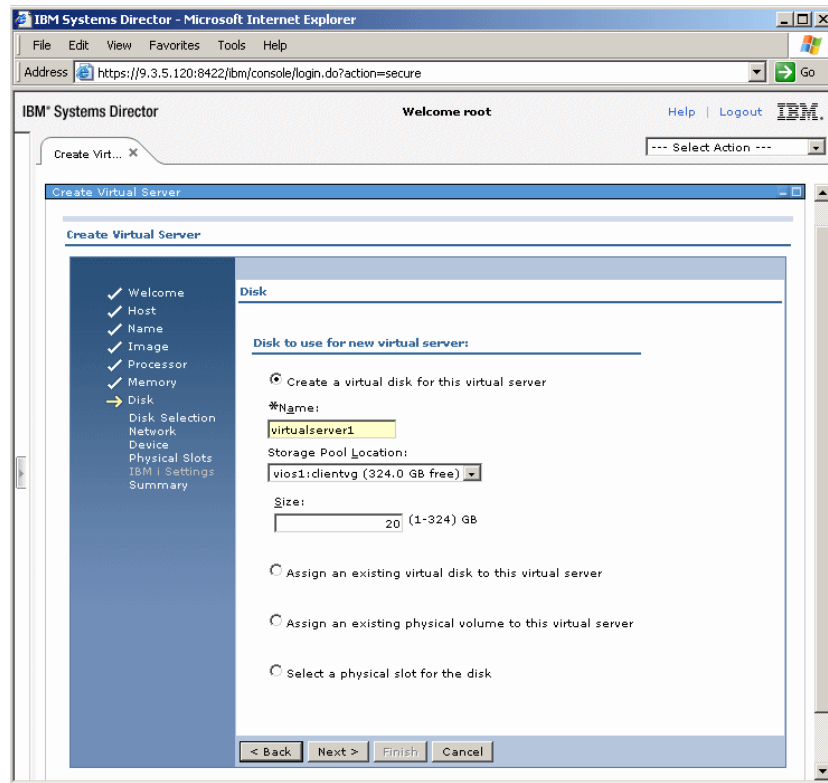


Figure 10-20 IBM Director Create Virtual Server disk

Click **Next**, select the Virtual Network, click **Next** several times to obtain the summary screen, click **Finish** to create the virtual server.

## Show Virtual adapters

You can use IBM Systems Director to view the virtual adapters configured in the Virtual I/O Server.

### Virtual Ethernet

In the IBM Systems Director navigation area.

Click **Navigate Resources**, select the **PowerVM** group, select the **vios1** virtual server, click **Actions** → **Extended Management** → **Hardware Informations** → **Virtual I/O Adapters** → **Ethernet**.

Figure 10-21 show the Virtual LAN Adapters on vios1.

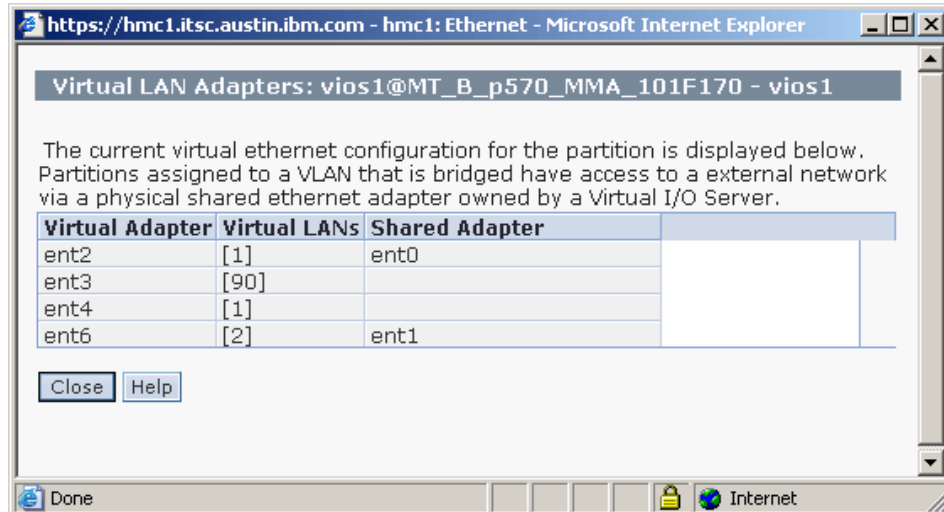


Figure 10-21 IBM Director Virtual LAN Adapters

### Virtual SCSI

In the IBM Systems Director navigation area.

Click **Navigate Resources**, select the **PowerVM** group, select the **vios1** virtual server, click **Actions** → **Extended Management** → **Hardware Informations** → **Virtual I/O Adapters** → **SCSI**.

Figure 10-22 show the Virtual SCSI Adapter configuration vios1.

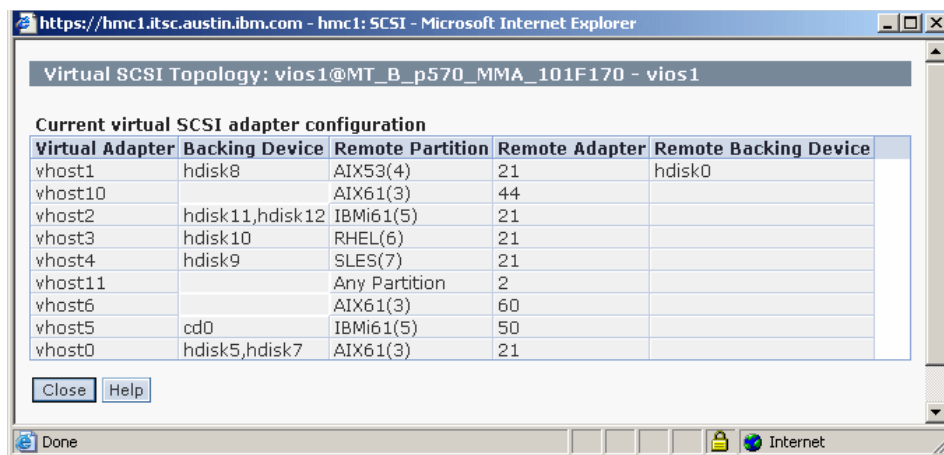


Figure 10-22 IBM Director Virtual SCSI Topology

## Topology map

You can use the topology map to view the relationship between systems.

In the IBM Systems Director navigation area.

Click **Navigate Resources**, select the **PowerVM** group, select the **vios1** virtual server, click **Actions** → **Topology Perspectives** → **Basic**

Figure 10-23 show the Basic Topology map for vios1.

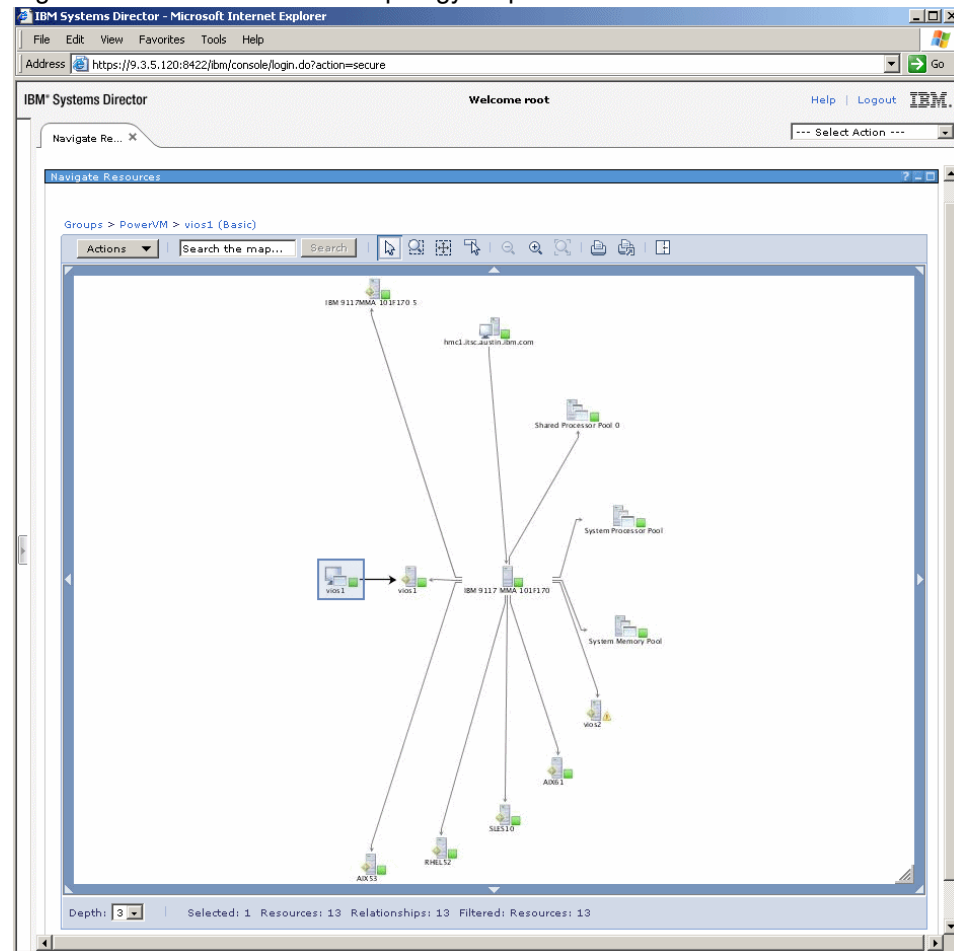


Figure 10-23 IBM Director Basic Topology map

## Inventory

You can use the topology map to view the relationship between systems.

In the IBM Systems Director navigation area.



Click **Navigate Resources**, select the **PowerVM** group, select the **vios1** virtual server, click **Actions** → **Inventory** → **View and Collect Inventory**.

On the “View and Collect Inventory” page, select **vios1** from the Target Systems list, select **All Hardware Inventory** from the Manage inventory profiles list, click **View Inventory**.

Figure 10-24 show the Hardware Inventory for vios1.

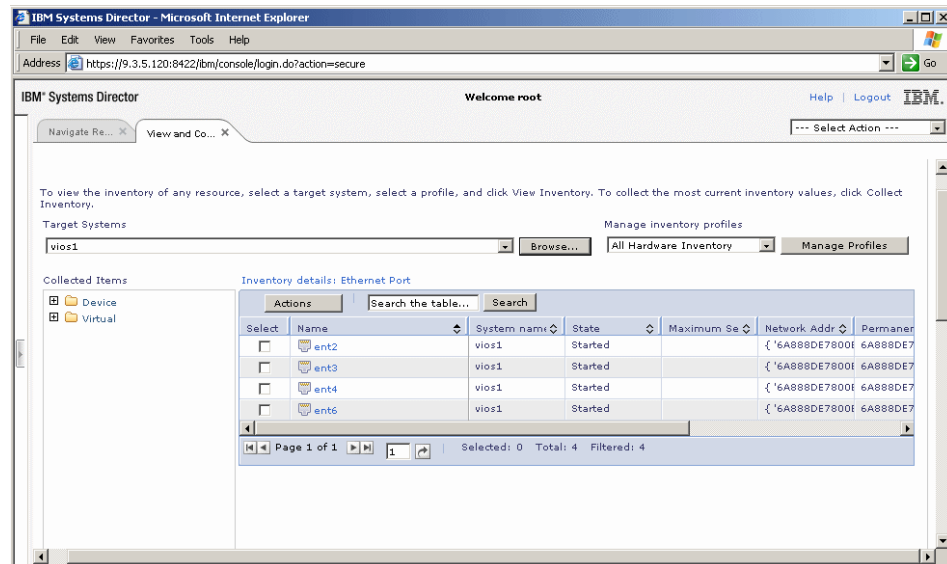


Figure 10-24 IBM Director Hardware Inventory

## Monitor resources

Use the Monitor task to retrieve real-time status and informations about your resources, for example: CPU Utilization. you can also set thresholds for the monitors and graph data.

In the IBM Systems Director navigation area.

Click **Navigate Resources**, select the **PowerVM** group, select the **vios1**, **vios2** virtual server and **MT\_B\_p570\_MMA\_101F170** CEC, click **Actions** → **System Status and Health** → **Monitors**

On the next screen, select **Virtualization Manager Monitors** and click **Show Monitors**

Figure 10-25 show the Virtualization Manager Monitor

This page displays the Virtualization Manager Monitors monitors.

IBM 9117 MMA 101F170, vio...

Select	Name	Monitor Name	Monitor Type	Threshold Stat	Current	Warning	Critical
<input type="checkbox"/>	IBM 9117 MMA 101F170	CPU Utilization %	Individual		1.1%		
<input type="checkbox"/>	IBM 9117 MMA 101F170	Entitled Processing Units	Individual		No Data Available		
<input type="checkbox"/>	IBM 9117 MMA 101F170	Memory (MB)	Individual		32768		
<input type="checkbox"/>	IBM 9117 MMA 101F170	Processors	Individual		4		
<input type="checkbox"/>	vios1	CPU Utilization %	Individual		1.85%		
<input type="checkbox"/>	vios1	Entitled Processing Units	Individual		0.25%		
<input type="checkbox"/>	vios1	Memory (MB)	Individual		1024		
<input type="checkbox"/>	vios1	Processors	Individual		1		
<input type="checkbox"/>	vios2	CPU Utilization %	Individual		1.43%		
<input type="checkbox"/>	vios2	Entitled Processing Units	Individual		0.3%		
<input type="checkbox"/>	vios2	Memory (MB)	Individual		1024		
<input type="checkbox"/>	vios2	Processors	Individual		1		

Page 1 of 1 | Selected: 0 Total: 12 Filtered: 12

Figure 10-25 IBM Director Virtualization Manager Monitor.

To Graph the CPU Utilization of **CEC**, select the **MT\_B\_p570\_MMA\_101F170**, click **Actions** → **Graph**

Figure 10-26 show the CPU Utilization graph for **vios1**.

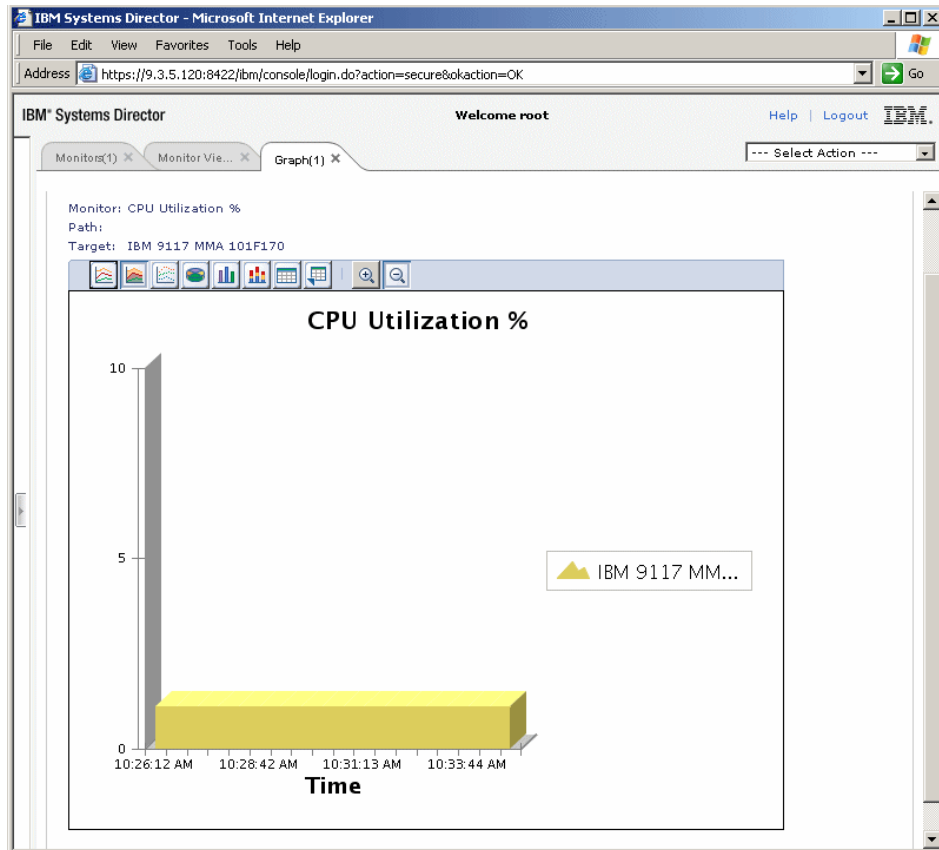


Figure 10-26 IBM Director CPU Utilization graph

## 10.1.10 Active Energy Manager

Active Energy Manager is an interface to the thermal and power management solution called EnergyScale™. This is an architecture that is a system-level power management implementation for Power6 processor-based machines.

### Basic principles of power management

Basic power management can be achieved with a few simple principles. First, the amount of power currently being consumed by individual servers or entire data centers must be assessed. This assessment must be made with the assumption that the UL listing is the worst case estimate of the amount of power used. Next, you need to determine the correct amount of power to allocate for individual servers or the entire data center by taking into account the history of the power

measurements. Finally, you should reduce or cap the amount of power consumed to fit your needs during low periods of power consumption.

## **Features EnergyScale of that can achieve the basic principles**

The features of EnergyScale provide a way to manage power effectively.

### ***Power Trending***

Power Trending is a method used to determine the amount of power that is currently being consumed by servers or an entire data center. The IBM Director Active Energy Manager can display continuous power usage statistics. You can view the statistics and determine trends of high or low power utilization, which can be used to determine power caps.

### ***Power Reduction***

Power can be reduced or capped through Power Saver Mode, power capping, or PCI slot management. Power Saver Mode drops voltage and frequency by a predetermined percentage to save power. Power capping uses power usage limits that are defined by a user. Finally, PCI slot management consists of automatically powering off PCI adapter slots that are not in use. When a PCI slot is empty, not assigned to a partition, assigned to a partition that is powered off, or dynamically removed from a partition, it is considered to be unused.

The Active Energy Manager can interface the POWER6 systems via IBM Director. IBM Director can connect to an HMC-managed system, a non-HMC managed system, or to a BladeCenter® chassis.

More detailed information on EnergyScale and Active Energy Manager can be found at the following links:

<http://www.research.ibm.com/journal/rd/516/mccreary.html>

<http://www-03.ibm.com/systems/management/director/about/director52/extensions/actengmrg.html>

## **10.2 Cluster Systems Management**

Cluster Systems Management (CSM) for AIX and Linux is designed for simple, low-cost management of distributed and clustered IBM Power Systems and Modular Systems™ servers in technical and commercial computing environments. In a virtualized Power Systems environment, CSM can provide a single administration point and improve administration efficiency.

Although virtualized commercial environments are not as homogenic as high performance computing clusters, management through CSM allows you to group

partitions, distribute files, run distributed commands, and monitor the systems. CMS is a cluster administration system, suitable not only for computing clusters but for any group of operating system instances. You should, however, have a reasonable number of systems to administer—at least two.

The CSM management server can scan Power Systems servers for newly created partitions and set up and initialize AIX or Linux installations on them. It can also help to keep partition groups on the same software level, and distribute updates and security patches. CSM monitoring is based on conditions and responses and can be adapted to monitor virtualized devices.

CSM offers a powerful command line interface that is fully scriptable and can be used for deep task automation. CSM deployment on Power Systems is described in detail in *Cluster Systems Management Cookbook for pSeries*, SG24-6859 and in CSM publications at:

<http://publib.boulder.ibm.com/infocenter/clresctr/vxrx/index.jsp?topic=/com.ibm.cluster.csm.doc/clusterbooks.html>

## 10.2.1 CSM architecture and components

CSM uses Reliable Scalable Clustering Technology RSCT as a backbone for a management domain. It connects to existing components for hardware control: Hardware Management Console, Integrated Virtualization Manager, or service processor. Hardware control allows to power on/off systems and check power status. For remote control of partitions the rconsole tool utilizes virtual serial connections. For AIX installation and software distribution, CSM utilizes AIX Network Installation Manager NIM.

CSM enables system administrators to resolve a number of system management challenges. Some of the tasks you can perform from the management server include:

- ▶ Installing and updating software on the cluster nodes
- ▶ Running distributed commands across the cluster
- ▶ Synchronizing files across the cluster
- ▶ Running user-provided customizing scripts during node installation or updates
- ▶ Monitoring the cluster nodes and devices
- ▶ Controlling cluster hardware
- ▶ Managing node or device groups
- ▶ Running diagnostic tools

- ▶ Configuring additional network adapters

For more information, see the CSM home page at:

<http://www-03.ibm.com/systems/clusters/software/csm/index.html>

## 10.2.2 CSM and the Virtual I/O Server

The Virtual I/O Server must be defined not as a CSM node but as a device. To do this, create a hardware definition file (similar to node definition) and run the **definehwdev** command. Example 10-3 shows a sample definition for Virtual I/O Servers.

*Example 10-3 Hardware definition example*

---

```
vios1:
    PowerMethod=hmc
    HWControlPoint=hmc1
    ConsoleMethod=hmc
    ConsoleServerName=hmc1
    RemoteShellUser=padmin
    RemoteShell=/usr/bin/ssh
    RemoteCopyCmd=/usr/bin/scp
    UserComment=""

vios2:
    PowerMethod=hmc
    HWControlPoint=hmc1
    ConsoleMethod=hmc
    ConsoleServerName=hmc1
    RemoteShellUser=padmin
    RemoteShell=/usr/bin/ssh
    RemoteCopyCmd=/usr/bin/scp
    UserComment=""
```

---

Run the **definehwdev** command:

```
# definehwdev -f hw-definition-vios.txt
Defining CSM Devices:
Defining Device "vios1"
```

Now you can exchange the SSH authentication keys with the **updatehwdev** command:

```
# updatehwdev -k -d vios1,vios2
```

After the key exchange you should be able to run the distributed shell on the servers:

```
# dsh -l padmin -d vios1,vios2 date
vios1: Fri Dec 7 09:19:08 CST 2007
vios2: Fri Dec 7 09:18:34 CST 2007
```

Running distributed commands on dual Virtual I/O Server configurations can be particularly useful for determining whether the multipathed virtual disks are mapped to the proper virtual adapter.







## Part 2

# PowerVM virtualization monitoring

This part describes best practices to monitor your advanced PowerVM environment. You will first find an introduction on performance considerations before digging into various monitoring techniques based on common situations. Finally, we present the integration to the IBM Tivoli framework.

Table 10-2 provides an overview of selected tools which can be used for monitoring resources like CPU, memory, storage and network in a Virtual I/O Server virtualized environment including AIX, IBM i, and Linux for POWER virtual I/O clients.

Table 10-2 Tools for monitoring resources in a virtualized environment

Resource / Platform	CPU	Memory	Storage	Network
<b>AIX</b>	<b>topas</b> <b>nmon</b> PM for Power Systems	<b>topas</b> <b>nmon</b> PM for Power Systems	<b>topas</b> <b>nmon</b> <b>iostat</b> <b>fcstat</b> PM for Power Systems	<b>topas</b> <b>nmon</b> <b>entstat</b> PM for Power Systems
<b>IBM i</b>	<b>WRKSYSACT</b> IBM Performance Tools for i5/OS, IBM Systems Director Navigator for i PM for Power Systems	<b>WRKSYSSTS</b> , <b>WRKSHRPOOL</b> IBM Performance Tools for i5/OS, IBM Systems Director Navigator for i PM for Power Systems	<b>WRKDSKSTS</b> IBM Performance Tools for i5/OS, IBM Systems Director Navigator for i PM for Power Systems	<b>WRKTCPSTS</b> IBM Performance Tools for i5/OS, IBM Systems Director Navigator for i PM for Power Systems
<b>Linux</b>	<b>iostat</b> <b>sar</b>	/proc/meminfo	<b>iostat</b>	<b>netstat</b> <b>iptraf</b>
<b>Virtual I/O Server</b>	<b>topas</b> , <b>viostat</b>	<b>topas</b> <b>vmstat</b> <b>svmon</b>	<b>topas</b> <b>viostat</b> <b>fcstat</b>	<b>topas</b> <b>entstat</b>
<b>system-wide</b>	IBM Director (all clients) <b>topas</b> (AIX, Virtual I/O Server), System i Navigator (IBM i)	<b>topas</b> (AIX, Virtual I/O Server), System i Navigator (IBM i)	<b>topas</b> (AIX, Virtual I/O Server), System i Navigator (IBM i)	<b>topas</b> (AIX, Virtual I/O Server), System i Navigator (IBM i)

**Notes:**

For IBM i the IBM Director 6.1 has the IBM Systems Director Navigator for i5/OS fully integrated for all level 0 or higher IBM i 6.1 or IBM i 5.4 agents.

For Linux systems the sysstat RPM needs to be installed for resource performance monitoring.

topas supports cross-partition information for AIX and Virtual I/O Server partitions only.





## Virtual I/O Server monitoring agents

This chapter describes agents that are available on the Virtual I/O Server and how to configure them to interact with an existing environment.

- ▶ IBM Tivoli Monitoring.
- ▶ IBM Tivoli Storage Manager.
- ▶ IBM Tivoli Usage and Accounting Manager.
- ▶ IBM TotalStorage Productivity Center
- ▶ IBM Tivoli Application Dependency Discovery Manager

## 11.1 IBM Tivoli Monitoring

Virtual I/O Server includes the IBM Tivoli Monitoring agent. IBM Tivoli Monitoring enables you to monitor the health and availability of multiple IBM Power System servers from the IBM Tivoli Enterprise Portal.

If you are already using IBM Tivoli Monitoring you can integrate the Virtual I/O Server and client partition agents into your existing Tivoli Enterprise Monitoring Server.

In this chapter a basic Tivoli Monitoring configuration on Virtual I/O Server and integration into the Tivoli Monitoring System are covered.

IBM Tivoli Monitoring Systems Edition for System p is provided at no charge download with one year of non-renewable support. A upgrade to IBM Tivoli Monitoring Version 6.2 is possible for a fee.

The link for more information and to download is:

<http://www-306.ibm.com/software/tivoli/products/monitor-systemp/>

ITM 6.2 Information Center:

<http://publib.boulder.ibm.com/infocenter/tivihelp/v15r1/index.jsp?topic=/com.ibm.itm.doc/welcome.htm>

Tivoli Monitoring Web page:

<http://www-01.ibm.com/software/tivoli/products/monitor/>

### 11.1.1 What to monitor

IBM Tivoli Monitoring enables you to monitor the health and availability of multiple Virtual I/O Server from the Tivoli Enterprise Portal (TEP). ITM gathers data from the Virtual I/O Server (VIOS), including data about physical volumes, logical volumes, storage pools, storage mappings, network mappings, real memory, processor resources, mounted file system sizes, and so on. From the Tivoli Enterprise Portal, you can view a graphical representation of the data, use predefined thresholds to alert you to abnormal conditions detected on key metrics, and resolve issues based on recommendations provided by the Expert Advice feature of ITM.

ITM provides the following functions:

- ▶ Topology and navigation
- ▶ Availability monitoring

- ▶ Health
- ▶ Customizable Workspaces, navigators, eventing and situations
- ▶ Performance and throughput
- ▶ Historical data collection
- ▶ Workflows

### 11.1.2 Agent configuration

The IBM Tivoli Monitoring agent is installed by default on the Virtual I/O Server. Use the following steps to configure it:

1. To list all the attributes associated with the agent configuration:

```
$ cfgsvc -ls ITM_premium
MANAGING_SYSTEM
HOSTNAME
RESTART_ON_REBOOT
```

2. Configure the agent:

```
$ cfgsvc ITM_premium -attr Restart_On_Reboot=TRUE hostname=itm1a
managing_system=hmc1a
```

```
Agent configuration started...
Agent configuration completed...
```

The Restart\_On\_Reboot attribute set to TRUE specifies to restart the ITM agent when the Virtual I/O Server is being rebooted, the hostname attribute specifies the Tivoli Enterprise Monitoring Server (TEMS) hostname and the managing system is the Hardware Management Console hostname.

3. Check the agent configuration

```
$ lssvc ITM_premium
MANAGING_SYSTEM:hmc1a
HOSTNAME:itm1a
RESTART_ON_REBOOT:TRUE
```

4. ssh configuration

ITM requires **ssh** command execution between the Virtual I/O Server and the Hardware Management Console (HMC) to allow the monitoring agent on the VIO Server to gather additional information only accessible from the HMC.

Displays the ssh public key that is generated for a particular agent configuration

```
$ cfgsvc ITM_premium -key
```

```
ssh-rsa
AAAAB3NzaC1yc2EAAAABIwAAAQEAuLNFAiXBDHe2pbZ7T5OHRmfdLAqCzT8PHn2fF1VfV4S
/hTaE06FzkmTCJPfs9dB5ATnRuKC/WF9Zf0BU3hVE/Lm1qyCkK7wym1TK00seaRu+xKKZjG
ZkHf/+BDvz4M/nvAELUGFco5vE+e2pamv9jJyfYvvGMSI4hj6eisRvjTfckcGpr9vBbJN27
Mi7T6RVZ2scN+5rRK80kj+bqvgPZZk4Ild0MLboRossgT01LURo3bGvuih9Xd3rUI0bQdj
8dJK8mVoA5T10+RF/zvPiQU9GT6XvcjmQdKbxLm2mgdATcPaDN4vOSBvXTaNSozpPnNhyxv
pugidtZBohznBDQ== root@vios2
```

Connect to the HMC and add the ssh public key.

```
hscroot@hmc1:~> mkauthkeys --add 'ssh-rsa
AAAAB3NzaC1yc2EAAAABIwAAAQEAuLNFAiXBDHe2pbZ7T5OHRmfdLAqCzT8PHn2fF1VfV4S
/hTaE06FzkmTCJPfs9dB5ATnRuKC/WF9Zf0BU3hVE/Lm1qyCkK7wym1TK00seaRu+xKKZjG
ZkHf/+BDvz4M/nvAELUGFco5vE+e2pamv9jJyfYvvGMSI4hj6eisRvjTfckcGpr9vBbJN27
Mi7T6RVZ2scN+5rRK80kj+bqvgPZZk4Ild0MLboRossgT01LURo3bGvuih9Xd3rUI0bQdj
8dJK8mVoA5T10+RF/zvPiQU9GT6XvcjmQdKbxLm2mgdATcPaDN4vOSBvXTaNSozpPnNhyxv
pugidtZBohznBDQ== root@vios2'
```

5. Start the monitoring agent

```
$ startsvc ITM_premium
Starting Premium Monitoring Agent for VIOS ...
Premium Monitoring Agent for VIOS started
```

### 11.1.3 Using the Tivoli Enterprise Portal

This section walks you through some examples of using the Tivoli Enterprise Portal graphical interface to retrieve important informations about the Virtual I/O Server which have the ITM agent started.

- ▶ Virtual IO Mappings
- ▶ Top Resources
- ▶ System
- ▶ Storage
- ▶ Networking

#### Launching Tivoli Enterprise Portal Client browser application

Point your web browser to the IP address of the IBM Tivoli Monitoring server followed by port 1920 as shown in Figure 11-1 on page 373 and click **IBM Tivoli Enterprise Portal Web Client**. You will be asked for an Authentication applet as shown in Figure 11-2 on page 374. Proceed with login and you will see IBM Portal Client applet.



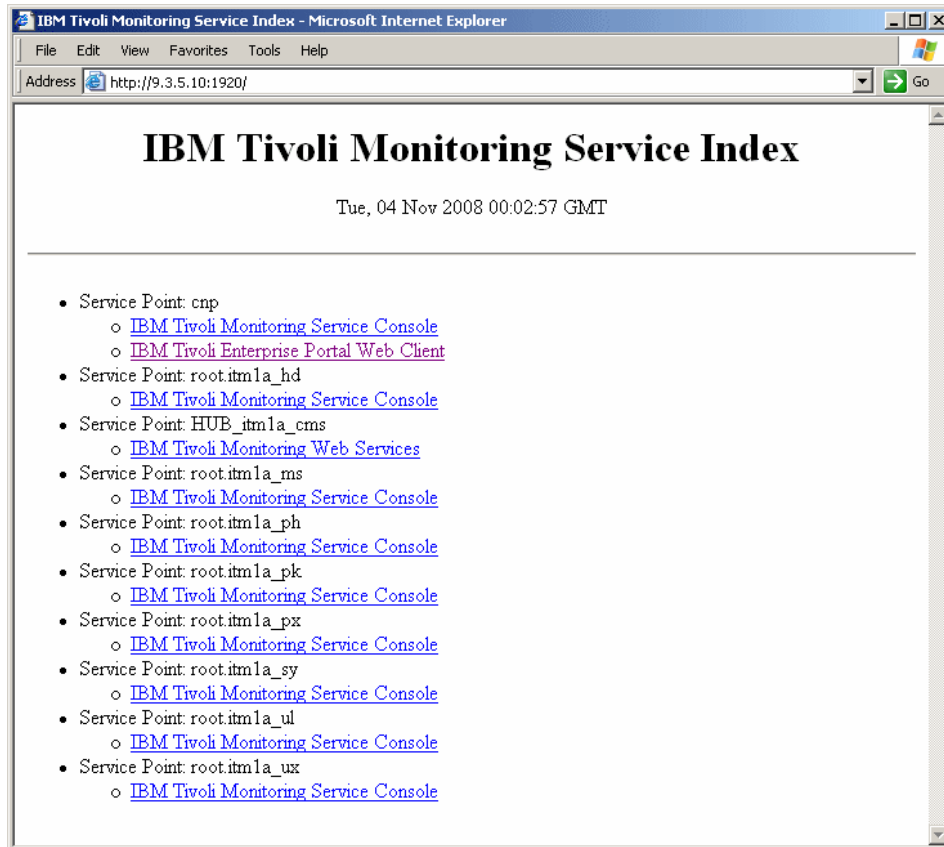


Figure 11-1 Tivoli Enterprise Portal login using web browser

## Launching Tivoli Enterprise Portal Client desktop application

Launch Tivoli Enterprise Portal Client (Figure 11-2 on page 374) using **Start** → **Programs** → **IBM Tivoli Monitoring** → **Tivoli Enterprise Portal**.

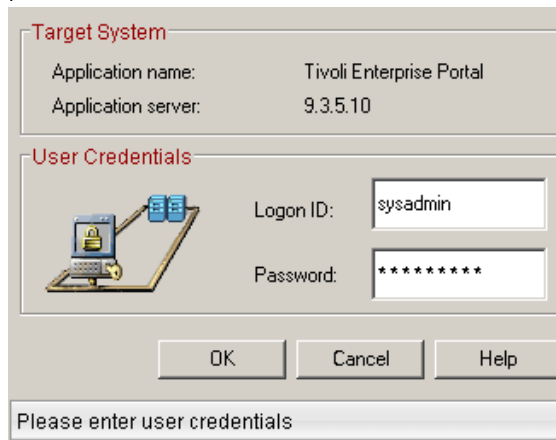


Figure 11-2 Tivoli Enterprise Portal login

## Virtual IO Mappings

Gives information about Virtual SCSI or Virtual Network Mappings based on the selected Workspace. Figure 11-3 shows how to choose a specific Workspace within the Virtual IO Mappings Navigator Item.

### Storage Mappings

In the Navigator window, right-click on Virtual IO Mappings then select **Workspace** → **Storage Mappings**

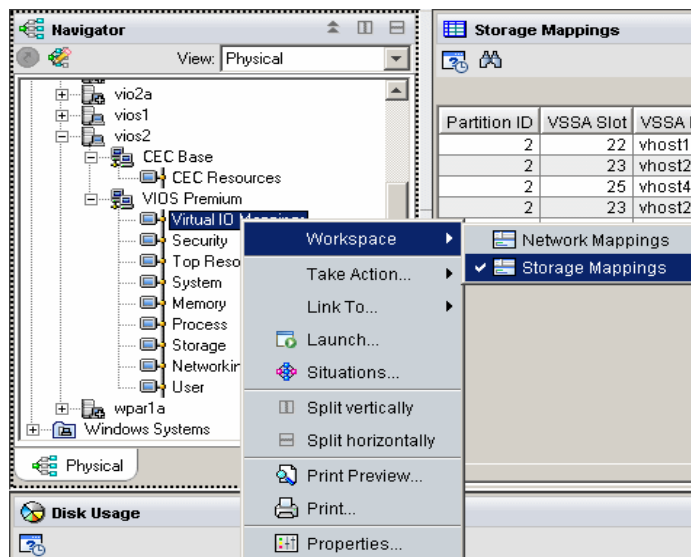


Figure 11-3 Storage Mappings Workspace selection

Storage Mappings provides all the informations related to Virtual SCSI configuration, Figure 11-4 on page 375 shows that the Virtual IO Server `vios2` has a Virtual SCSI Server Adapter `vhost1` with slot number 22 and that Client partition named `AIX53` can access the `hdisk1` physical disk located on `vios2` using a SCSI Client adapter with slot number 22.

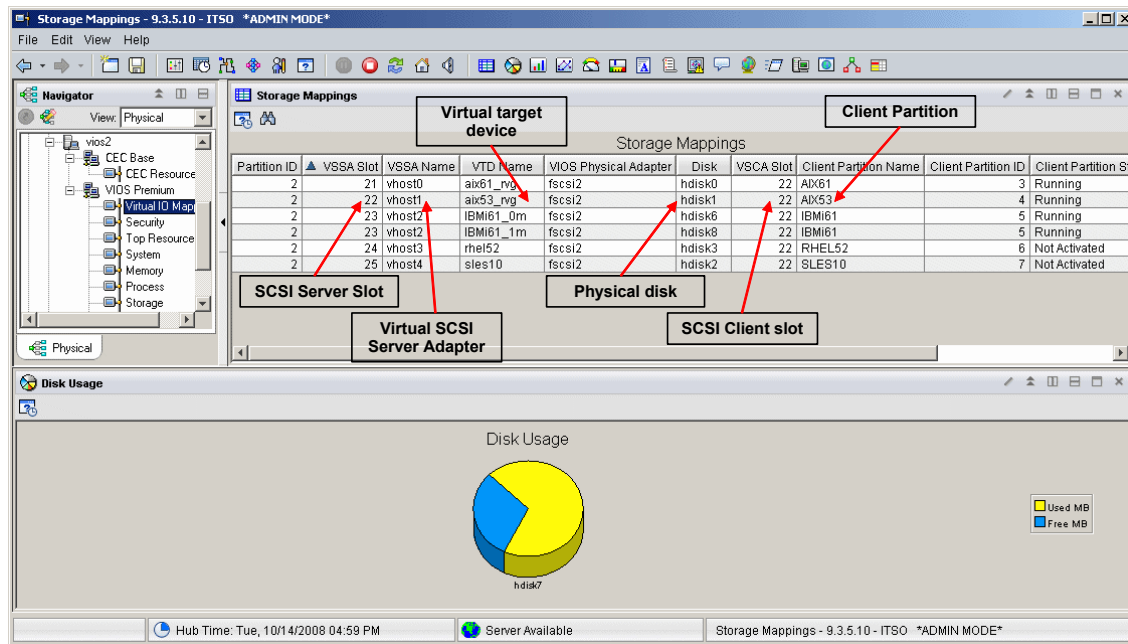


Figure 11-4 ITM panel showing Storage Mappings

### Network Mappings

In the Navigator, right-click on Virtual IO Mappings, select **Workspace** → **Network Mappings**

Network Mappings gives you all the informations related to Virtual Network configuration, Figure 11-5 on page 376 shows that the Virtual IO Server `vios2` has a Shared Ethernet Adapter `ent5` that uses the `ent0` physical Ethernet adapter and the `ent2` Virtual Ethernet adapter with VLAN id 1, there are also 2 other Virtual Ethernet adapter `ent3` and `ent4`.

The screenshot shows the Tivoli Management Console interface. On the left is the Navigator pane showing a tree view of resources under 'Physical' view, including 'vios2', 'CEC Base', 'CEC Resources', 'VIOS Premium', 'Virtual IO Mapping', 'Security', 'Top Resources', 'System', 'Memory', 'Process', 'Storage', 'Networking', and 'User'. The main area is divided into three panels:

- Network Mappings**: A table listing network mappings. The 'Trunk' column for partition '1 vios1' is highlighted in yellow.
- Network Mappings Details**: A detailed table for the selected mappings, showing columns like 'Partition State' (all 'Running') and 'VEA MAC'.
- Network Interfaces**: A panel at the bottom, currently empty.

At the bottom of the console, there is a status bar showing 'Hub Time: Tue, 10/14/2008 06:54 PM', 'Server Available', and 'Network Mappings - 9.3.5.10 - ITSO \*ADMIN MODE\*'.

Figure 11-5 ITM panel showing Network Mappings

## Top Resources

In the Navigator, right-click on Top Resources, select **Workspace** → **Top Resources Usage**

It gives you informations related to File Systems Metrics, processes CPU and Memory consumption, Figure 11-6 on page 377 shows the Top Resources usage Workspace.

The screenshot displays the Tivoli Network Mappings interface. The left pane shows a tree view with 'Virtual I/O Mapping' selected. The main pane is divided into two sections: 'Network Mappings' and 'Network Mappings Details'.

**Network Mappings Table:**

VLAN ID	Partition Name	VEA Slot	Trunk	Shared Ethernet Adapter	Physical Ethernet Adapters	Virtual Ethernet Adapters	Failover	Priority	Host
90	vios2	12				ent3		unavailable	
1	vios2	11	yes	ent5	ent0		auto	2	
1	vios2	13				ent4		unavailable	
90	vios1	12							
1	vios1	13							
1	vios1	11	yes						
1	RHEL52	2							
1	IBMI61	2							
1	ADXB1	2							
1	ADX53	2							

**Network Mappings Details Table:**

VLAN ID	Partition Name	Partition State	Hostname	IP Address	Partition ID	VEA Slot	VEA MAC	VEA IP address	Trunk	Shared Ethernet Adapter	SEA IP Address	SEA MAC	Physical Ethernet
1	IBMI61	Running			5	2	6A888626F102						
1	vios1	Running			1	13	6A888DE7800D						
1	vios1	Running			1	11	6A888DE7800B		yes				
1	ADXB1	Running			3	2	6A8882AA9B02						
90	vios1	Running			1	12	6A888DE7800C						

The status bar at the bottom shows: Hub Time: Tue, 10/14/2008 06:54 PM, Server Available, and Network Mappings - 9.3.5.10 - ITSO \*ADMIN MODE\*

Figure 11-6 ITM panel showing Top Resources usage

## System

In the Navigator, right-click on System, select **Workspace** → **CPU Utilization**

It gives you real time informations related to User, Idle, System and IO Wait CPU percentage Utilization Figure 11-7 on page 378 shows the CPU Utilization Workspace.

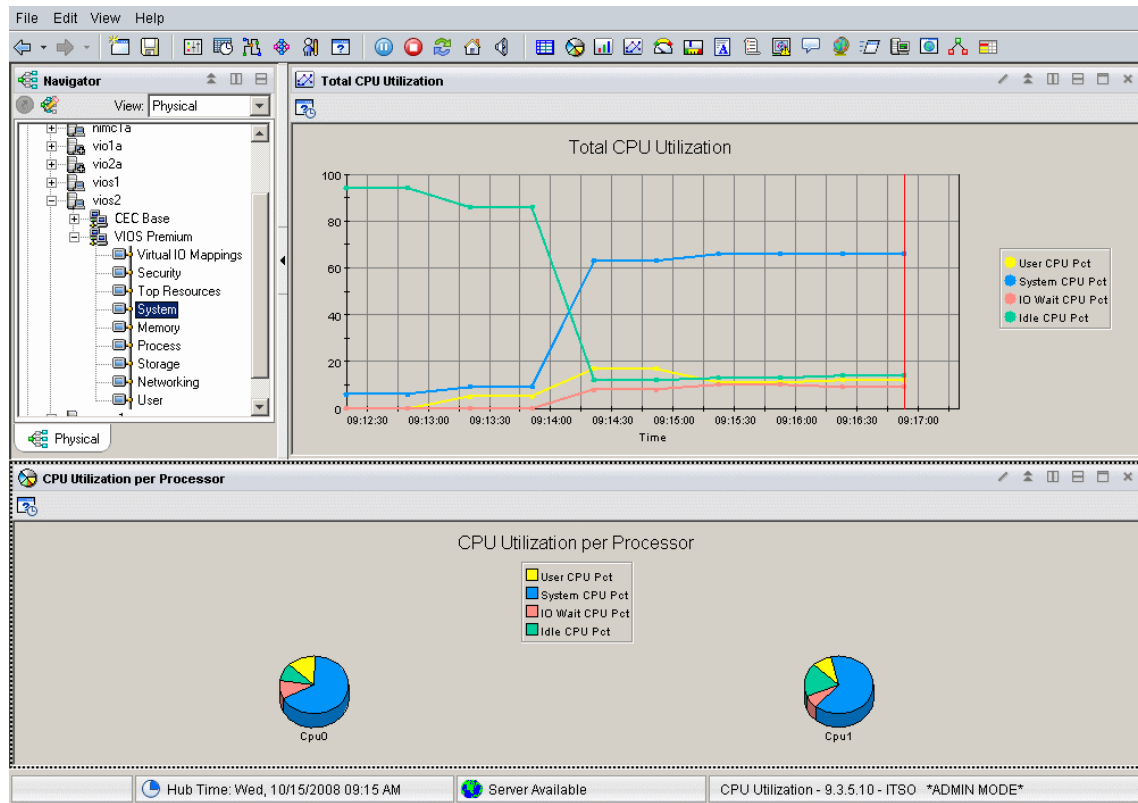


Figure 11-7 ITM panel showing CPU Utilization

## Storage

In the Navigator, right-click on Storage, select **Workspace** → **System Storage Information**

It gives you informations related to disks activity Figure 11-8 on page 379 shows the System Storage Information Workspace.

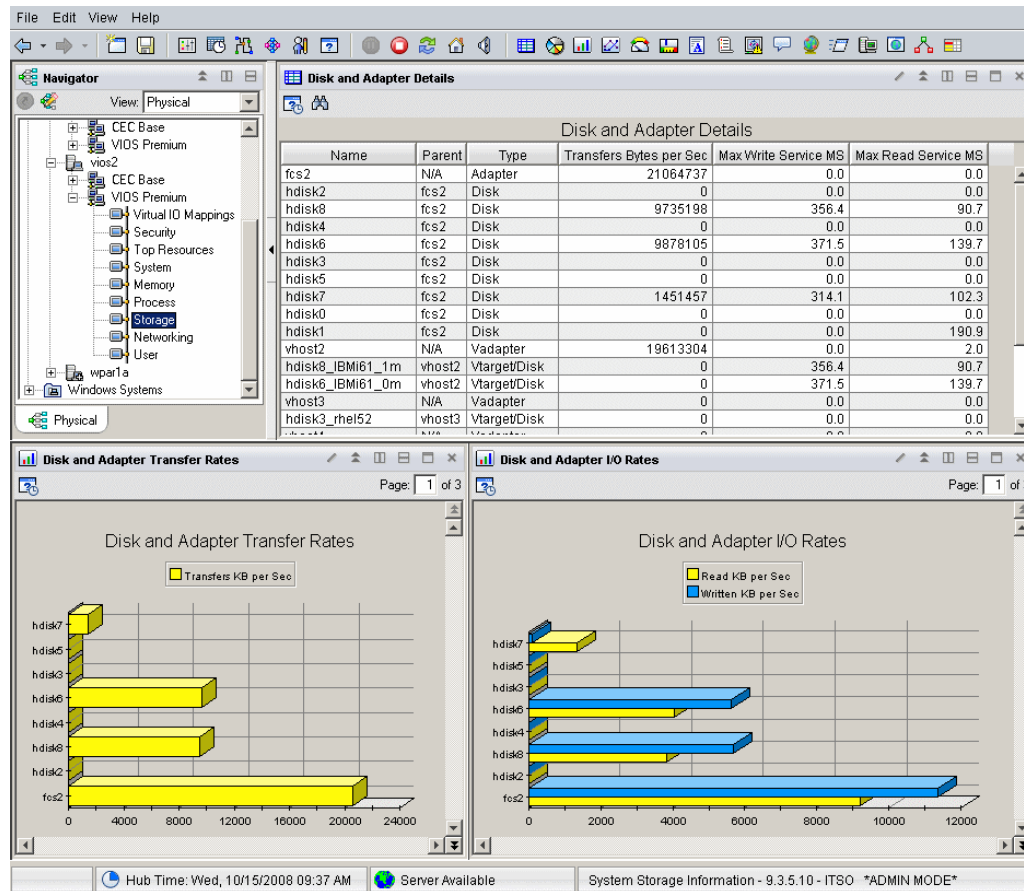


Figure 11-8 ITM panel showing System Storage Information

## Networking

In the Navigator, right-click on Networking, select **Workspace** → **Network Adapter Utilization**

It gives you information related to network adapter activity. Figure 11-9 on page 380 shows the Network Adapter Utilization Workspace.

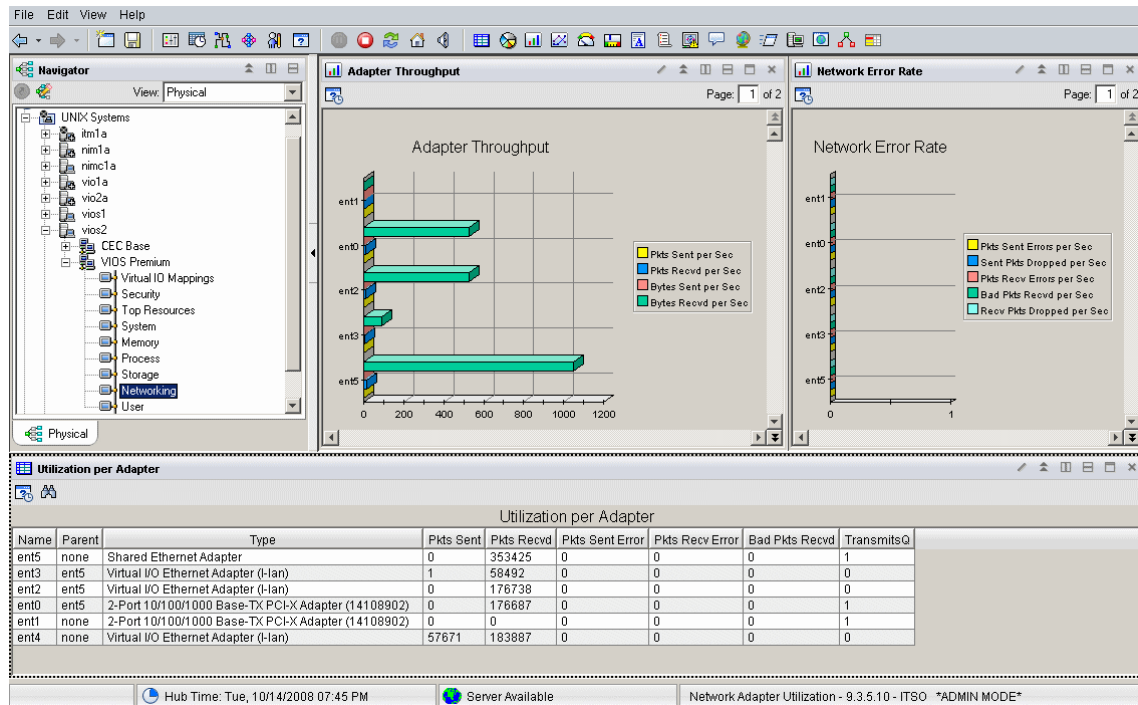


Figure 11-9 ITM panel showing Network Adapter Utilization

## 11.2 Configuring the IBM Tivoli Storage Manager client

Virtual I/O Server includes the IBM Tivoli Storage Manager agent. With Tivoli Storage Manager, you can protect your data from failures and other errors by storing backup and disaster recovery data in a hierarchy of offline storage. Tivoli Storage Manager can help protect computers running a variety of different operating environments, including the Virtual I/O Server, on a variety of different hardware, including IBM Power Systems servers. Configuring the Tivoli Storage Manager client on the Virtual I/O Server enables you to include the Virtual I/O Server in your standard backup framework.

Use the following steps to configure the TSM agent:

1. List all the attributes associated with the agent configuration.

```
$ cfgsvc -ls TSM_base
SERVERNAME
SERVERIP
NODENAME
```



## 2. Configure the agent.

```
$ cfgsvc TSM_base -attr SERVERNAME=hostname SERVERIP=name_or_address  
NODENAME=vios
```

Where:

- ▶ *hostname* is the host name of the Tivoli Storage Manager server with which the Tivoli Storage Manager client is associated.
- ▶ *name\_or\_address* is the IP address or domain name of the Tivoli Storage Manager server with which the Tivoli Storage Manager client is associated.
- ▶ *vios* is the name of the machine on which the Tivoli Storage Manager client is installed. The name must match the name registered on the Tivoli Storage Manager server.

Ask the Tivoli Storage Manager administrator to register the client node, the Virtual I/O Server, with the Tivoli Storage Manager server. To determine what information you must provide to the Tivoli Storage Manager administrator, see the IBM Tivoli Storage Manager documentation at:

<http://publib.boulder.ibm.com/infocenter/tivihelp/v1r1/index.jsp>

After you are finished, you are ready to back up and restore the Virtual I/O Server using the Tivoli Storage Manager. This is described in 5.4.5, “Backing up using IBM Tivoli Storage Manager” on page 203.

## 11.3 IBM Tivoli Usage and Accounting Manager agent

You can configure and start the IBM Tivoli Usage and Accounting Manager agent on the Virtual I/O Server.

IBM Tivoli Usage and Accounting Manager helps you track, allocate, and invoice your IT costs by collecting, analyzing, and reporting on the resources used by entities such as cost centers, departments, and users. IBM Tivoli Usage and Accounting Manager can gather data from multi-tiered data centers including Windows, AIX, Virtual I/O Server, HP/UX Sun™ Solaris™, Linux, i5/OS, and VMware.

Use the following steps to configure the IBM Tivoli Usage and Accounting Manager agent:

1. List all the attributes associated with the agent configuration.

```
$cfgsvc -ls ITUAM_base  
ACCT_DATA0  
ACCT_DATA1
```

```
ISYSTEM  
IPROCESS
```

## 2. Configure the agent.

```
$ cfgsvc ITUAM_base -attr ACCT_DATA0=value1 ACCT_DATA1=value2  
ISYSTEM=value3 IPROCESS=value4
```

Where:

- ▶ *value1* is the size (in MB) of the first data file that holds daily accounting information.
- ▶ *value2* is the size (in MB) of the second data file that holds daily accounting information.
- ▶ *value3* is the time (in minutes) when the agent generates system interval records.
- ▶ *value4* is the time (in minutes) when the system generates aggregate process records.

## 3. Restart the agent.

Finally stop and restart the monitoring agent to use the new configuration. To do so, run the **stopsvc** and **startsvc** commands:

```
$ stopsvc ITUAM_base  
Stopping agent...  
startsvcAgent stopped...  
$ startsvc ITUAM_base  
Starting agent...  
Agent started...
```

After you start the IBM Tivoli Usage and Accounting Manager agent, it begins to collect data and generate log files. You can configure the IBM Tivoli Usage and Accounting Manager server to retrieve the log files, which are then processed by the IBM Tivoli Usage and Accounting Manager Processing Engine. You can work with the data from the IBM Tivoli Usage and Accounting Manager Processing Engine as follows:

- ▶ You can generate customized reports, spreadsheets, and graphs. IBM Tivoli Usage and Accounting Manager provides full data access and reporting capabilities by integrating Microsoft SQL Server® Reporting Services or Crystal Reports with a Database Management System (DBMS).
- ▶ You can view high-level and detailed cost and usage information.
- ▶ You can allocate, distribute, or charge IT costs to users, cost centers, and organizations in a manner that is fair, understandable, and reproducible.

For more information, see one of the following resources:

- ▶ If you are running the IBM Tivoli Usage and Accounting Manager Processing Engine on Windows, then see *IBM Tivoli Usage and Accounting Manager Data Collectors for Microsoft Windows User's Guide*, SC32-1557-02, which is available at:

<http://www.elink.ibm.link.ibm.com/public/applications/publications/cgi-bin/pbi.cgi?CTY=US&FNC=SRX&PBL=SC32-1557-02>

- ▶ If you are running the IBM Tivoli Usage and Accounting Manager Processing Engine on UNIX or Linux, then see *IBM Tivoli Usage and Accounting Manager Data Collectors for UNIX and Linux User's Guide*, SC32-1556-02, which is available at:

<http://www.elink.ibm.link.ibm.com/public/applications/publications/cgi-bin/pbi.cgi?CTY=US&FNC=SRX&PBL=SC32-1556-02>

## 11.4 IBM TotalStorage Productivity Center

IBM TotalStorage Productivity Center is a storage infrastructure management suite. It is designed to help simplify and automate the management of storage devices, storage networks, and capacity utilization of file systems and databases. IBM TotalStorage Productivity Center can help you manage the following activities:

- ▶ Manage capacity utilization of storage systems, file systems and databases
- ▶ Automate file-system capacity provisioning
- ▶ Perform device configuration and management of multiple devices from a single user interface
- ▶ Tune and proactively manage the performance of storage devices on the Storage Area Network (SAN)
- ▶ Manage, monitor and control your SAN fabric

Use the following steps to configure the IBM TotalStorage Productivity Center agent:

1. To list all the attributes associated with an agent configuration, type the following command:

```
$cfgsvc -ls TPC
A
S
devAuth
caPass
```

```

amRegPort
amPubPort
dataPort
devPort
caPort
newCA
oldCA
daScan
daScript
daInstall
faInstall
U

```

## 2. Configure the agent:

The TPC agent is a TotalStorage Productivity Center agent. Agent names are case sensitive. This agent requires that you specify the S, A, devAuth, and caPass attributes for configuration. By default, specifying this agent will configure both the TPC\_data and TPC\_fabric agents, as provided in Table 11-1 on page 384.

*Table 11-1 TPC agent attributes, descriptions, and their values*

Attributes	Description	Value
S	Provides the TotalStorage Productivity Center agent with a TotalStorage Productivity Center server host name.	Host name or IP address
A	Provides the TotalStorage Productivity Center agent with an agent manager host name.	Host name or IP address
devAuth	Sets the TotalStorage Productivity Center device server authentication password.	Password
caPass	Sets the CA authentication password.	Password
caPort	Sets the CA port. The default value is 9510.	Number
amRegPort	Specifies the agent manager registration port. The default value is 9511	Number
amPubPort	Specifies the agent manager public port. The default value is 9513.	Number

Attributes	Description	Value
dataPort	Specifies the TotalStorage Productivity Center data server port. The default value is 9549.	
devPort	Specifies the TotalStorage Productivity Center device server port. The default value is 9550.	
newCA	The default value is true.	True or false
oldCA	The default value is true.	True or false
daScan	The default value is true.	True or false
daScript	The default value is true.	True or false
daInstall	The default value is true.	True or false
faInstall	The default value is true.	True or false
U	Specifies to uninstall the agent.	all   data   fabric

To configure the DIRECTOR\_agent with several attributes, type the following command:

```
$ cfgsvc TPC -attr S=tpc_server_hostname A=agent_manager_hostname
devAuth=password caPass=password
```

The installation wizard appears, type the number next to the language that you want to use for the installation and enter 0. The license agreement panel appears.

```
Initializing InstallShield Wizard.....
Launching InstallShield Wizard.....
```

-----  
Select a language to be used for this wizard.

```
[ ] 1 - Czech
[X] 2 - English
[ ] 3 - French
[ ] 4 - German
[ ] 5 - Hungarian
[ ] 6 - Italian
[ ] 7 - Japanese
[ ] 8 - Korean
[ ] 9 - Polish
[ ] 10 - Portuguese (Brazil)
```

- [ ] 11 - Russian
- [ ] 12 - Simplified Chinese
- [ ] 13 - Spanish
- [ ] 14 - Traditional Chinese

To select an item enter its number, or 0 when you are finished: [0]

Read the license agreement panel. Type 1 to accept the terms of the license agreement. The agent is installed on the Virtual I/O Server according to the attributes specified in the **cfgsvc** command.

Please choose from the following options:

- [x] 1 - I accept the terms of the license agreement.
- [ ] 2 - I do not accept the terms of the license agreement.

To select an item enter its number, or 0 when you are finished: [0]

Installing TPC Agents  
Install Location: /opt/IBM/TPC  
TPC Server Host: tpc\_server\_hostname  
Agent Manager Host: agent\_manager\_hostname

Start the agent:

```
startsvc TPC_fabric
```

```
startsvc TPC_data
```

## 11.5 IBM Tivoli Application Dependency Discovery Manager

IBM Tivoli Application Dependency Discovery Manager (TADDM) discovers infrastructure elements found in the typical data center, including application software, hosts and operating environments (including the Virtual I/O Server), network components (such as routers, switches, load balancers, firewalls, and storage), and network services (such as LDAP, NFS, and DNS). Based on the data it collects, TADDM automatically creates and maintains application infrastructure maps that include runtime dependencies, configuration values, and change history. With this information, you can determine the interdependences between business applications, software applications, and physical components to help you ensure and improve application availability in your environment. For example, you can do the following tasks:

- ▶ You can isolate configuration-related application problems.

- ▶ You can plan for application changes to minimize or eliminate unplanned disruptions.
- ▶ You can create a shared topological definition of applications for use by other management applications.
- ▶ You can determine the effect of a single configuration change on a business application or service.
- ▶ You can see what changes take place in the application environment and where.

TADDM includes an agent-free discovery engine, which means that the Virtual I/O Server does not require that an agent or client be installed and configured in order to be discovered by TADDM. Instead, TADDM uses discovery sensors that rely on open and secure protocols and access mechanisms to discover the data center components.

- ▶ For more information, see:

<http://www-01.ibm.com/software/tivoli/products/taddm/>







## Monitoring global system resource allocations

In order to precisely monitor the various resource consumptions such as CPU, memory, network adapters, or storage adapters, you should first have a clear understanding of the global resources allocation on the system.

Several tools provide some information about the resources allocations. The first one is the Hardware Management Console (HMC), which is also used to manage the resources allocations.

If the system is not managed by an HMC but by an Integrated Virtualization Manager (IVM), this tool can also be used to monitor resources allocations.

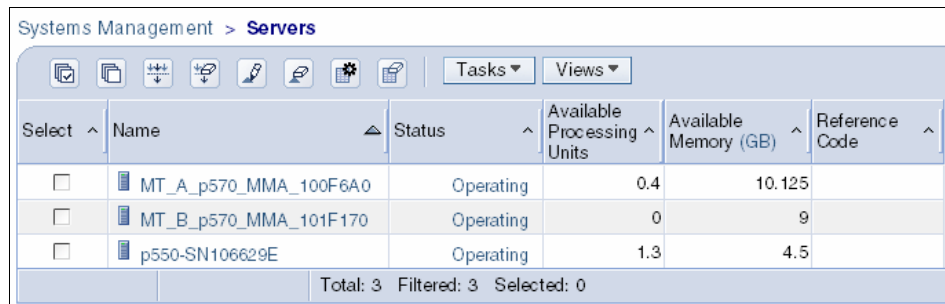
On a virtual I/O client, the **lparstat** command may run on any AIX partition and inspect its current resources allocations. A Linux alternative command is also available.

## 12.1 Hardware Management Console monitoring

The HMC is generally used to set up the system allocations and for system maintenance. Yet it is also very good at monitoring the current resources allocations.

To do so you have to log on to the HMC Web interface. You access the login window by accessing the IP or DNS name of your HMC using https through a Web browser.

You first see the global server allocations by selecting **Systems Management** → **Servers** from the left pane. The list of managed systems then appears in the main area. This is illustrated in Figure 12-1.



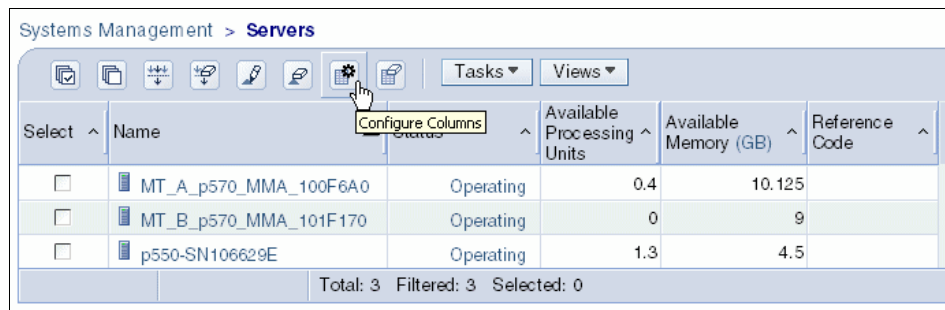
The screenshot shows the HMC web interface for 'Systems Management > Servers'. It features a toolbar with icons for various actions and two dropdown menus for 'Tasks' and 'Views'. Below the toolbar is a table with the following columns: 'Select', 'Name', 'Status', 'Available Processing Units', 'Available Memory (GB)', and 'Reference Code'. The table contains three rows of server data. At the bottom of the table, it displays 'Total: 3 Filtered: 3 Selected: 0'.

Select	Name	Status	Available Processing Units	Available Memory (GB)	Reference Code
<input type="checkbox"/>	MT_A_p570_MMA_100F6A0	Operating	0.4	10.125	
<input type="checkbox"/>	MT_B_p570_MMA_101F170	Operating	0	9	
<input type="checkbox"/>	p550-SN106629E	Operating	1.3	4.5	

Total: 3 Filtered: 3 Selected: 0

Figure 12-1 Available servers managed by the HMC

The unallocated resources are directly visible, which simplifies the administrator's investigations. Moreover it is possible to view additional information by pressing **Configure Columns**, as shown in Figure 12-2.



This screenshot is identical to Figure 12-1, but with a mouse cursor pointing to the 'Configure Columns' icon (a gear with a plus sign) in the toolbar. A tooltip box labeled 'Configure Columns' is visible over the icon.

Select	Name	Status	Available Processing Units	Available Memory (GB)	Reference Code
<input type="checkbox"/>	MT_A_p570_MMA_100F6A0	Operating	0.4	10.125	
<input type="checkbox"/>	MT_B_p570_MMA_101F170	Operating	0	9	
<input type="checkbox"/>	p550-SN106629E	Operating	1.3	4.5	

Total: 3 Filtered: 3 Selected: 0

Figure 12-2 Configuring the displayed columns on the HMC

You can then track the resources allocation information that you are interested in.

## 12.1.1 Partition properties monitoring

To get detailed information for a partition, select its system name and then select the name of the partition. A new window opens and displays the partition properties. Navigating in the different tabs shows the current resources allocations, as shown in Figure 12-3.

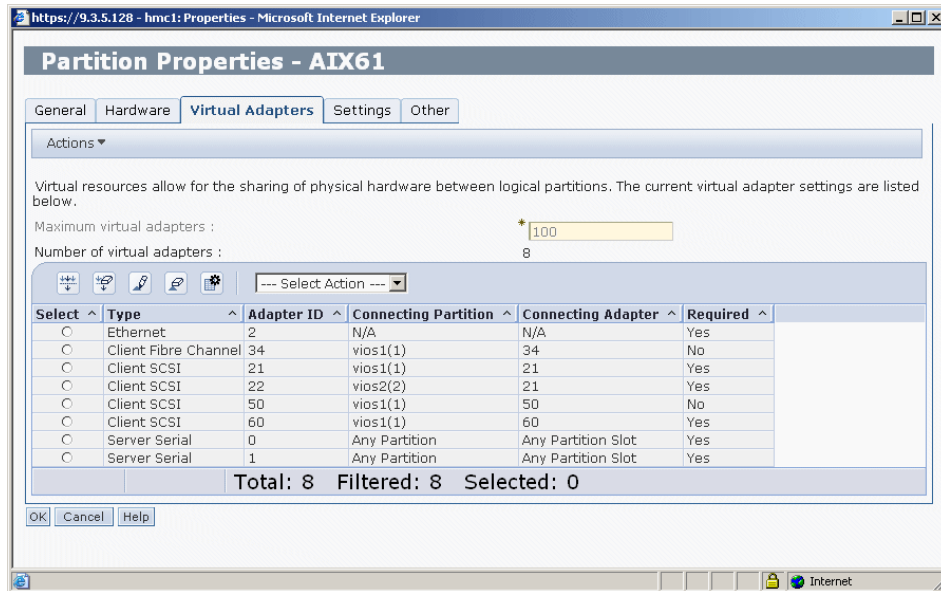


Figure 12-3 Virtual adapters configuration in the partition properties

## 12.1.2 HMC hardware information monitoring

Starting with HMC V7, you can now look at virtual SCSI and virtual LAN topologies of the Virtual I/O Server from the HMC.

**Tip:** The HMC has a feature to aid administrators in looking at virtual SCSI and virtual Ethernet topologies in the Virtual I/O Server.

To do so, select the Virtual I/O Server partition where you want to see the topologies. Select **Hardware Information** → **Virtual I/O Adapters** → **SCSI**, as shown in Figure 12-4 on page 392.

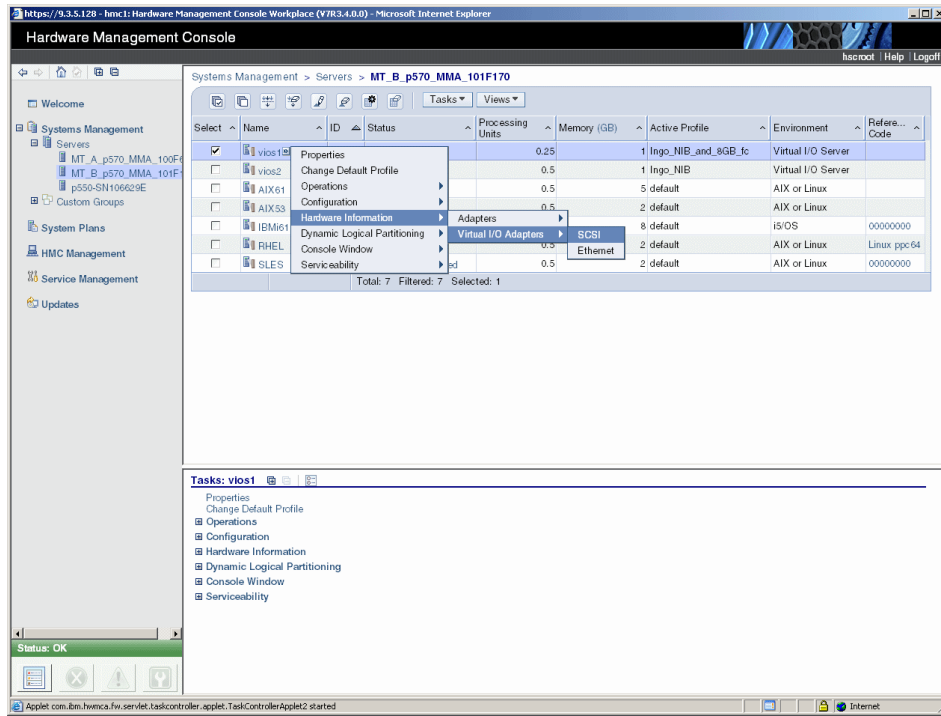


Figure 12-4 Virtual I/O Server hardware information context menu

A window then appears showing the virtual storage topology, as shown on Figure 12-5 on page 392.

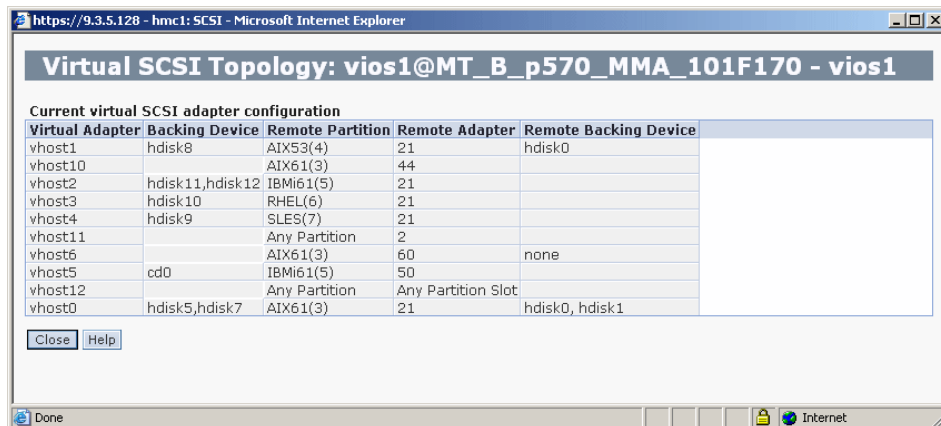


Figure 12-5 The Virtual I/O Server virtual SCSI topology window

### 12.1.3 HMC virtual network monitoring

Starting with HMC V7, you can monitor the virtual network information for each server attached to the HMC.

To do so, select the server where you want to monitor the available virtual networks. Then select **Tasks** → **Configuration** → **Virtual Network Management**, as shown in Figure 12-6.

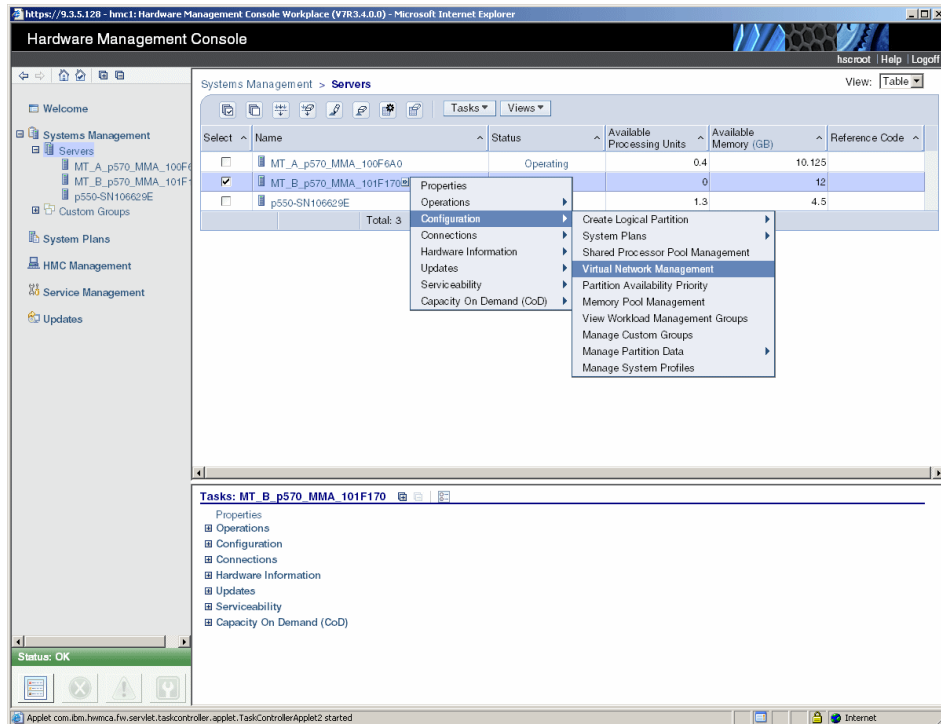


Figure 12-6 Virtual Network Management

A new window appears providing information about the available virtual network topology. If you select a VLAN, you get detailed information about the partitions assign to this VLAN, as shown on Figure 12-7.

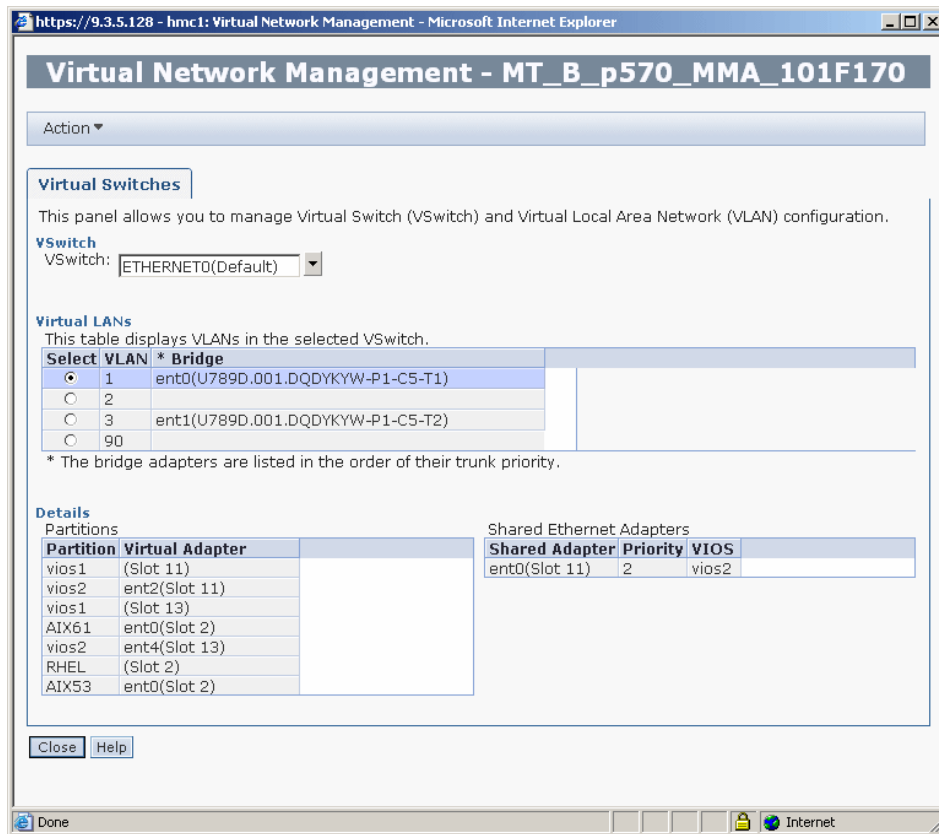


Figure 12-7 Virtual Network Management - detail information

## 12.1.4 HMC shell scripting

It is also possible to log on to the HMC using the `ssh` command. You can then script the HMC commands to retrieve the information provided by the HMC interface.

A practical example of this is the open source `hmcMenu` script that you can find at:

<http://www.the-welters.com/professional/scripts/hmcMenu.txt>

For more information about the HMC V7, refer to *Hardware Management Console V7 Handbook*, SG24-7491.

## 12.2 Integrated Virtualization Manager monitoring

The Integrated Virtualization Manager (IVM) shows detailed partition information in the partitions management panel for servers not managed by an HMC.

Log on to the IVM Web interface of your Virtual I/O Server. Access the login screen by entering the IP address or the DNS name of your Virtual I/O Server in a Web browser.

You then only have to select **View/Modify Partitions** from the left menu.

This is illustrated in Figure 12-8 on page 395. You see the amount of memory, the processing units that are allocated, and those that are available on the machine.

The screenshot shows the IVM web interface in a Microsoft Internet Explorer browser window. The address bar shows the URL `http://9.3.5.112/main.faces`. The page title is "Integrated Virtualization Manager". The left sidebar contains a navigation menu with categories like "Partition Management", "I/O Adapter Management", "Virtual Storage Management", "IVM Management", "System Plan Management", and "Service Management". The main content area is titled "View/Modify Partitions" and includes a "System Overview" section with the following data:

Total system memory:	4 GB	Total processing units:	2
Memory available:	2.2 GB	Processing units available:	1.6
Reserved firmware memory:	304 MB	Processor pool utilization:	0.02 (0.9%)
System attention LED:	Inactive		

Below the system overview is a "Partition Details" section with a table of active partitions:

Select	ID	Name	State	Uptime	Memory	Processors	Entitled Processing Units	Utilized Processing Units	Reference Code
<input type="checkbox"/>	1	<a href="#">10-478DE</a>	Running	3.08 Days	512 MB	2	0.2	0.01	
<input type="checkbox"/>	2	<a href="#">chris1</a>	Running	3.03 Days	1 GB	1	0.2	0.00	

Figure 12-8 IVM partitions monitoring

You can also access the Virtual network configuration by selecting the **View/Modify Virtual Ethernet** item from the left panel.

Figure 12-9 on page 396 illustrates the virtual Ethernet configuration monitoring.

Virtual Ethernet provides Ethernet connectivity among partitions. The table below can show two views of the virtual Ethernets on which partitions have a configured adapter. Select the Partition view for a list of all virtual Ethernets for each partition or select the Virtual Ethernet view for a list of all partitions for each virtual Ethernet. Use the Ethernet tab of the Properties page for the partition to change these settings.

View by:

Partition Name	Virtual Ethernet 1	Virtual Ethernet 2	Virtual Ethernet 3	Virtual Ethernet 4
10-478DE (1)	* <input checked="" type="checkbox"/>	* <input checked="" type="checkbox"/>	* <input checked="" type="checkbox"/>	* <input checked="" type="checkbox"/>
chris1 (2)	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

\* Partition is capable of bridging this virtual Ethernet

Figure 12-9 IVM virtual Ethernet configuration monitoring

Another interesting configuration to monitor is the Virtual storage configuration. Select the **View/Modify Virtual Storage** item from the left menu.

Figure 12-9 illustrates the virtual Ethernet configuration monitoring.

To perform an action on a virtual disk, first select the virtual disk or virtual disks, and then select the task.

\* Create Virtual Disk... Modify partition assignment --- More Tasks ---

Select	Name ^	Storage Pool	Assigned Partition	Size
<input type="checkbox"/>	lp2vd1	chris1 (Default)	chris1 (2)	36 GB

Figure 12-10 IVM virtual storage configuration monitoring

For more information about the IVM interface, refer to *Integrated Virtualization Manager on IBM System p5*, REDP-4061.



## 12.3 Monitoring resources allocations from a partition

When you are logged on a partition, you might prefer to use the command line interface to display the current partition and system wide processor and memory resources allocations.

Some commands are available on AIX or Linux partitions to display these resource allocations.

**Note:** For IBM i system wide resource allocation information use the HMC interface.

### 12.3.1 Monitoring CPU and memory allocations from AIX

In the AIX operating system, run the `lparstat -i` command. Example 12-1 illustrates the command output.

*Example 12-1 lparstat -i command output on AIX*

---

```
# lparstat -i
Node Name                : aix61
Partition Name           : AIX61
Partition Number         : 3
Type                     : Shared-SMT
Mode                     : Uncapped
Entitled Capacity        : 0.50
Partition Group-ID       : 32771
Shared Pool ID           : 0
Online Virtual CPUs      : 1
Maximum Virtual CPUs     : 1
Minimum Virtual CPUs     : 1
Online Memory             : 2048 MB
Maximum Memory           : 6016 MB
Minimum Memory           : 256 MB
Variable Capacity Weight : 128
Minimum Capacity         : 0.10
Maximum Capacity         : 1.00
Capacity Increment       : 0.01
Maximum Physical CPUs in system : 16
Active Physical CPUs in system : 4
Active CPUs in Pool      : 4
Shared Physical CPUs in system : 4
Maximum Capacity of Pool : 400
Entitled Capacity of Pool : 375
Unallocated Capacity     : 0.00
Physical CPU Percentage   : 50.00%
```

Unallocated Weight	: 0
Memory Mode	: Dedicated
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	: -

---

## 12.3.2 Monitoring CPU and memory allocations from Linux

Supported Linux distributions also provide a command line interface to display the partition's CPU and memory resource allocations. This uses the `/proc/ppc64/lparcfg` special device. You only have to read its content, shown in Example 12-2.

### *Example 12-2 Listing partition resources on Linux*

---

```
[root@VIOCRHEL52 ~]# cat /proc/ppc64/lparcfg
lparcfg 1.8
serial_number=IBM,02101F170
system_type=IBM,9117-MMA
partition_id=6
BoundThrds=1
CapInc=1
DisWheRotPer=5120000
MinEntCap=10
MinEntCapPerVP=10
MinMem=128
MinProcs=1
partition_max_entitled_capacity=100
system_potential_processors=16
DesEntCap=50
DesMem=2048
DesProcs=1
DesVarCapWt=128
DedDonMode=0

partition_entitled_capacity=50
group=32774
system_active_processors=4
pool=0
pool_capacity=400
pool_idle_time=0
pool_num_procs=0
```

```
unallocated_capacity_weight=0
capacity_weight=128
capped=0
unallocated_capacity=0
entitled_memory=2147483648
entitled_memory_group_number=32774
entitled_memory_pool_number=65535
entitled_memory_weight=0
unallocated_entitled_memory_weight=0
unallocated_io_mapping_entitlement=0
entitled_memory_loan_request=0
backing_memory=2147483648 bytes
cmo_enabled=0
purr=13825418184
partition_active_processors=1
partition_potential_processors=2
shared_processor_mode=1
```

---





## Monitoring commands on the Virtual I/O Server

The Virtual I/O Server comes with several commands that can monitor its activity. Some look like the standard AIX commands and others are a bit different.

The command parameters are specific to the Virtual I/O Server system. It is therefore good to familiarize yourself with them. The Virtual I/O Server online help can be used with running the **help** command to display the available commands. Generally speaking, running the commands with the **-h** parameter displays the command syntax and provides enough information to correctly use the tool. If more information is required, man pages are available. You can also look at the online reference that can be found at:

<http://publib.boulder.ibm.com/infocenter/systems/scope/hw/topic/iphcg/sa76-0101.pdf>

In this section, the principal monitoring commands available on the Virtual I/O Server are presented, along with some practical uses.

## 13.1 Global system monitoring

To get general system information, use the following commands:

<b>topas</b>	This command is similar to <b>topas</b> in AIX. It presents various system statistics such as CPU, memory, network adapters or disk usage.
<b>sysstat</b>	This command gives you an uptime for the system and the list of logged-on users.
<b>svmon</b>	Captures and analyzes a snapshot of virtual memory.
<b>vmstat</b>	Reports statistics about kernel threads, virtual memory, disks, traps, and processor activity.
<b>wkldout</b>	Provides post-processing of recordings made by wkldagent. The files are located in the /home/ios/perf/wlm path.
<b>lsgcl</b>	Displays the contents of the global command log.
<b>vasistat</b>	Shows VASI device driver and device statistics (used for PowerVM Live Partition Mobility).

Some commands are also useful for configuration inspection:

<b>ioslevel</b>	Gives the version of the Virtual I/O Server.
<b>lssw</b>	Lists the software installed.
<b>lsfware</b>	Displays microcode and firmware levels of the system, adapters and devices.
<b>lslparinfo</b>	Displays LPAR number and LPAR name.
<b>lssvc</b>	Lists available agent names if the parameter is not given. Lists the agent's attributes and values if the agent's name is provided as parameter.
<b>oem_platform_level</b>	Returns the operating system level of the OEM install and setup environment.
<b>chlang</b>	This command is primarily for Japanese locales. Use this command to force messages on the left to appear in English. Without this option, messages during the boot sequence may be corrupted.

## 13.2 Device inspection

Some commands provide device configuration information:

<b>lsdev</b>	Displays devices in the system and their characteristics.
<b>lsmap</b>	Displays the mapping between physical and virtual devices.

## 13.3 Storage monitoring and listing

Some monitoring commands report various storage activities:

<b>viostat</b>	Reports Central Processing Unit statistics, asynchronous input/output, input/output statistics for entire system, adapters, tty devices, disks and CD-ROMs. The parameter <b>-extdisk</b> provides detailed performance statistics information for disk devices.
<b>nmon</b>	Displays local system statistics such as system resource, processor usage etc., Users can use interactive or recording mode.
<b>fcstat</b>	Displays statistics gathered by the specified Fibre Channel device driver.
<b>lsvg</b>	Displays information about volume groups.
<b>lslv</b>	Displays information about a logical volume.
<b>lspv</b>	Displays information about physical volumes.
<b>lssp</b>	Displays information about storage pools.
<b>lsvopt</b>	Lists and displays information about the systems virtual optical devices.
<b>lsrep</b>	Lists and displays information about the Virtual Media Repository.
<b>lspath</b>	Displays information about paths to MultiPath I/O (MPIO)-capable devices.

Additional commands are available with the root login through the **oem\_setup\_env** command:

<b>mpio_get_config -Av</b>	Displays IBM SAN storage information with MPIO multi-path device driver attachment.
<b>fget_config -Av</b>	Displays IBM SAN storage information with RDAC multi-path device driver attachment.
<b>pcmpath query adapter</b>	Displays SAN storage Fibre Channel adapter path information with SDDPCM path control module attachment.

<b>pcmpath query device</b>	Displays SAN storage device information with SDDPCM path control module attachment.
<b>lscfg -v1 <i>hdiskx</i></b>	Displays vendor storage information.

## 13.4 Network monitoring

Some monitoring commands can be used for network monitoring:

<b>netstat</b>	Displays active sockets for each protocol, or routing table information, or displays the contents of a network data structure.
<b>entstat</b>	Shows Ethernet device driver and device statistics.
<b>seastat</b>	Generates a report to view, per client, Shared Ethernet Adapter statistics.

Additional commands report network configuration information:

<b>hostname</b>	Sets or displays the name of the current host system.
<b>lsnetsvc</b>	Gives the status of a network service.
<b>lstcpip</b>	Displays the TCP/IP settings.
<b>optimizenet -list</b>	Lists the characteristics of one or all network tunables.
<b>snmp_info -get -next</b>	Requests values of Management Information Base variables managed by a Simple Network Management Protocol agent.
<b>traceroute</b>	Prints the route that IP packets take to a network host.
<b>showmount</b>	Displays a list of all clients that have remotely mountable file systems.

## 13.5 User ID listing

<b>lsuser</b>	Displays user account attributes.
---------------	-----------------------------------





# CPU monitoring

This chapter will describe shared processor pools, processing units, simultaneous multithreading and PowerVM processor terminologies. It will also explain cross-partition monitoring and various other CPU monitoring tools used to monitor Virtual I/O Servers and virtual I/O clients with AIX, IBM i, and Linux operating systems.

## 14.1 CPU-related terminology and metrics

The POWER5 and POWER6 systems introduced a high level of abstraction of CPU consumption.

### 14.1.1 Terminology and metrics common to POWER5 and POWER6

On a POWER5 system, it is possible to grant either dedicated processors to a logical partition or some processing unit parts of the global shared processor pool.

The Global Shared Processor Pool consists of the processors that are not already assigned to running partitions with dedicated processors.

Consider the example illustrated in Figure 14-1 on page 407. There is a system with 16 processors. Two partitions (A and B) are running. Each partition is using three dedicated processors. The Shared-Processor Pool then contains  $16 - (2 * 3) = 10$  processors.



Figure 14-1 16-core system with dedicated and shared CPUs

Partitions using the Shared-Processor Pool are allocated based on a certain amount of processing units. A processing unit consists of one tenth of a processor. In the example, the Shared-Processor Pool currently contains  $10 * 10 = 100$  processing units.

Partitions using the Shared-Processor Pool are also assigned a certain number of virtual processors. This represents the number of processors that are seen by the operating system running in the partition. The processing units are distributed among the virtual processors depending on their load.

In the same example, we define 2 partitions using the Shared-Processor Pool. The first one (C) has 52 processing units and 7 virtual processors. The second one (D) has 23 processing units and 5 virtual processors. 25 processing units remain unallocated.

Another feature that was introduced with the POWER5 systems is the ability of a partition to use some processing units that were not originally assigned to it. *Capped* shared partitions can not use more processing units than originally assigned. *Uncapped* shared partitions can use additional processing units if they need any and if some are available in the shared processor pool.

The number of processing units currently allocated to a partition represents its *entitlement*. In case a partition is an uncapped shared partition, it may consume more processing units than allocated. Of course, it may use less processing units than allocated. This is why the metrics named *physical consumption* and *entitlement consumption* are defined. *Physical consumption* represents the amount of processing unit currently consumed. *Entitlement consumption* represents the percentage of processing unit currently consumed compared to the number of processing units allocated to the partition. Consequently, uncapped shared partitions can have an entitlement consumption that exceeds 100%.

An additional feature related to CPU consumption was introduced with the POWER5 architecture. The Simultaneous Multi-Threading functionality can be activated in the operating system of a partition. When active, it allows the partition to run two simultaneous threads on a single virtual processor. A virtual processor is then seen as two logical processors. When SMT is not enabled, a virtual processor appears as a single logical processor.

Table 14-1 provides a summary of the CPU terminology and the metrics that are defined to monitor CPU consumption.

Table 14-1 POWER5-based terminology and metrics

Term	Description	Related metrics
dedicated CPU	CPU directly allocated to a partition. Other partitions can not use it.	standard CPU consumption metrics (user, sys, idle, wait etc.)
shared CPU	CPU part of the shared processor pool	physical consumption
processing unit	Power resource of a 10th of the combined CPUs in a shared processor pool	physical consumption
virtual CPU	CPU as seen by a partition when Simultaneous Multithreading is off	<ul style="list-style-type: none"> <li>▶ Simultaneous Multithreading state</li> <li>▶ logical CPU count</li> </ul>
logical CPU	CPU as seen by a partition	<ul style="list-style-type: none"> <li>▶ Simultaneous Multi-Threading state</li> <li>▶ logical CPU count</li> </ul>

Term	Description	Related metrics
Simultaneous multithreading	Capacity for a partition to run two threads simultaneously on a virtual CPU	<ul style="list-style-type: none"> <li>▶ Simultaneous Multi-Threading state</li> <li>▶ Logical CPU count</li> </ul>

## 14.1.2 Terminology and metrics specific to POWER6 systems

Several enhancements were done in POWER6 architecture which introduced new functions and the related metrics.

On a POWER6 systems, you can define Multiple Shared Processor Pool. By default, there is only one Global Shared-Processor Pool.

There are several reasons for using several Shared Processor Pools. First of all, it is possible to define maximum processing unit for a Shared Processor Pool. It ensures that the partitions that are part of this pool will not use more processing unit. This new value is now used to compute the per processor license fees for IBM products.

Second of all, you can define a Multiple Shared Processor Pool processing units reserve. The processing units part of this reserve are distributed among the uncapped shared partitions in the pool.

This is illustrated in the next example in Figure 14-2 on page 410. In this example an additional processor pool is defined. This pool can use up to three processors, which results in 30 processing units. Two partitions (E and F) are defined in this Shared-Processor Pool. E has an entitlement of 0.8 CPU and F has an entitlement of 0.3 CPU.

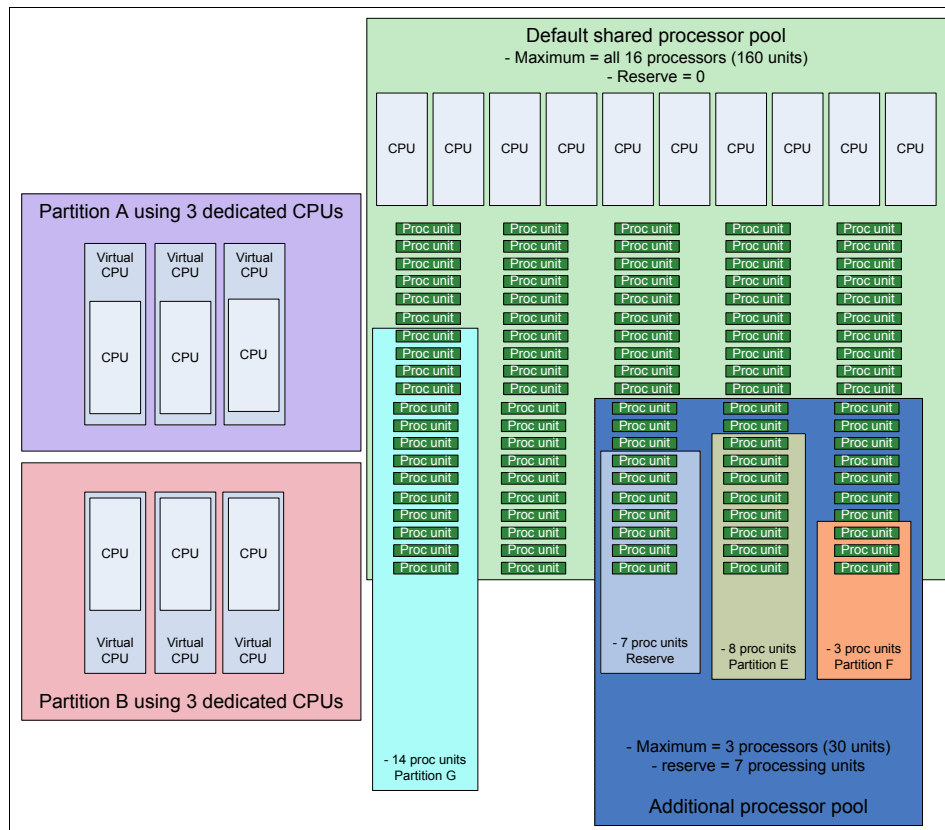


Figure 14-2 A Multiple Shared-Processor Pool example on POWER6

Another enhancement in POWER6 architecture is related to the partitions using dedicated CPU. Upon activation, if a partition does not use 100% of its dedicated processor resources then unused processor resources will be ceded to the Shared-processor pools. These processors are called Donating dedicated CPU.

The new POWER6-specific terminologies and related metrics are shown in Table 14-2.

Table 14-2 POWER6-specific terminology and metrics

Term	Related metric
Multiple shared processor pool	pool size maximum capacity pool entitlement reserve = bucket size available capacity

Term	Related metric
Dedicated shared processors	donation mode donated cycles count

There are a lot of tools that can provide CPU consumption metrics on the Virtual I/O Server and its clients. Depending on the requirements the tools described in subsequent sections can be used.

## 14.2 CPU metrics computation

The previous mechanism of calculating performance based on sampling is not accurate in case of virtual CPU resources as 1 CPU may be shared among more than 1 partition. Its will also fail for SMT enabled CPUs. Moreover, uncapping feature is also tricky as a partition can go beyond its capacity and its difficult to say when a processor is 100 percent busy, unlike capped partitions.

To solve these issues POWER5 family of processors implemented a new performance-specific register called the Processor Utilization of Resources Register (PURR). PURR keeps track of real processor resource usage on a per-thread or per-partition level. AIX performance tools have been updated in AIX V5.3 to show these new statistics.

Traditional performance measurements were based on sampling, at typically, the rate of 100 Hz sample rate (each sample corresponds to a 10ms tick). Each sample was sorted into one of user, system, iowait or idle category based on the code it was executing when interrupted. But this sampling based approach will not work in a virtualized environment because dispatch cycle of each virtual processor is no longer the same (which was the assumption in traditional performance measurement). A similar problem exists with SMT: if one thread is consuming 100 percent of the time on a physical CPU, sample-based calculations would report the system as 50 percent busy (one processor at 100 percent and another at 0 percent), but in fact the processor is really 100 percent busy. To preserve binary compatibility traditional mechanism has not been changed.

### 14.2.1 Processor Utilization of Resources Register (PURR)

PURR is a 64-bit counter with the same units for the timebase and decremter registers that provide per-thread processor utilization statistics. Figure 14-3 shows the relationship between PURR registers in a single POWER5 processor and the two hardware threads and two logical CPUs. With SMT enabled, each hardware thread is seen as a logical processor.

The *timebase* register shown in Figure 14-3 is incremented at each tick. The *decrementer* register provides periodic interrupts.

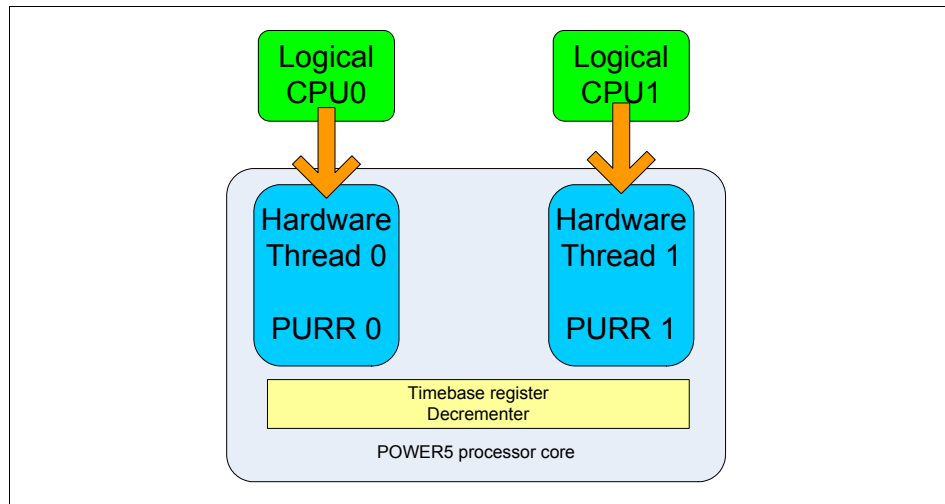


Figure 14-3 Per-thread PURR

At each processor clock cycle the PURR, which last executed or is executing the instruction, is incremented. The sum of two PURRs equals the value of timebase register. This approach is an approximation as SMT allows both threads to run in parallel. Hence, it can not be evolved to distinguish the performance difference between SMT on and SMT off mode.

## 14.2.2 New PURR-based metrics

These new registers provide some new statistics. These might be helpful while looking at the output of performance tools.

### CPU statistics in SMT environment

The ratio of  $(\Delta \text{PURR})/(\Delta \text{timebase})$  over an interval indicates the fraction of physical processor consumed by a logical processor. This is the value returned by the `sar -P ALL` and `mpstat` commands as shown in subsequent sections.

The value of  $(\Delta \text{PURR}/\Delta \text{timebase}) * 100$  over an interval gives percentage of physical processor consumed by logical processor. This value is returned by the `mpstat -s` command, which shows the SMT statistics as explained in subsequent sections.



## CPU statistics in shared-processor partitions

In a shared-processor environment, the PURR measures the time that a virtual processor runs on a physical processor. With SMT on, virtual time base is the sum of the two PURRs; With SMT off, virtual time base is the value stored in the PURR.

### ▶ Capped shared processors

For capped shared processors the *entitled PURR* over an interval is given as *entitlement \* time base*.

So %user time over an interval is (and similarly %sys and %iowait):

$$\%user = (\text{delta PURR in user mode} / \text{entitled PURR}) * 100$$

### ▶ Uncapped shared processors

For uncapped shared processors, the calculations take the variable capacity into account. The *entitled PURR* in the above formula is replaced by the *consumed PURR* whenever the latter is greater than the entitlement. So %user time over an interval is:

$$\%user = (\text{delta PURR in user mode} / \text{consumed PURR}) * 100$$

## Physical processor consumption for a shared processor

A partition's physical processor consumption is the sum of all its logical processor consumption:

$$SUM(\text{delta PURR} / \text{delta TB})$$

## Partition entitlement consumption

A partition's entitlement consumption is the ratio of its physical processor consumption (PPC) to its entitlement expressed as a percentage:

$$(\text{Physical processor consumption} / \text{entitlement}) * 100$$

## Shared-processor pool spare capacity

Unused cycles in a shared-processor pool is spent in POWER Hypervisor's idle loop. The POWER Hypervisor enters this loop when all partition entitlements are satisfied and there are no partitions to dispatch. The time spent in Hypervisor's idle loop, measured in ticks, is called the Pool Idle Count. The Shared-processor pool spare capacity over an interval is expressed as:

$$(\text{delta Pool idle count} / \text{delta timebase})$$

and is measured in numbers of processors. Only partitions with shared-processor pool authority will be able to display this figure.

### Logical processor utilization

It is the sum of traditional 10 ms tick-based sampling of the time spent in %sys and %user. If it starts approaching 100 percent, it indicates that partition can use additional virtual processors.

## 14.2.3 System-wide tools modified for virtualization

The AIX tools **topas**, **lparstat**, **vmstat**, **sar** and **mpstat** now add two extra columns of information when executing in a shared-processor partition:

- ▶ Physical processor consumed by the partition, shown as **pc** or **%physc**.
- ▶ Percentage of entitled capacity consumed by the partition, shown as **ec** or **%entc**.

### Logical processor tools

The logical processor tools are **mpstat** and **sar -P ALL** commands. When running in a partition with SMT enabled, these commands add a column Physical Processor Fraction Consumed (*delta PURR/delta TB*), shown as **physc**. This shows the relative split of physical processor time for each of the two logical processors.

When running in a shared processor partition, these commands add a new column, Percentage of Entitlement Consumed ( $(PPFC/ENT)*100$ ) shown as **%entc**. It gives relative entitlement consumption for each logical processor expressed as a percentage.

## 14.2.4 Scaled Processor Utilization of Resources Register (SPURR)

IBM POWER6 microprocessor chip supports advanced, dynamic power management solutions for managing not just the chip but the entire server. This design facilitates a programmable power management solution for greater flexibility and integration into system and data-center-wide management solutions.

The design of the POWER6 microprocessor provides real-time access to detailed and accurate information about power, temperature, and performance. Altogether, the sensing, actuation and management support available in the POWER6 processor is known as the EnergyScale architecture. It enables higher performance, greater energy efficiency and new power management capabilities such as power and thermal capping and power savings with explicit performance control.

## The EnergyScale architecture and statistics computations

Although the EnergyScale implementation is primarily an out-of-band power management design; managing system-level power and temperature has some effects on in-band software. Basically, power management results in performance variability. Which implies as the power management implementation operates, it can change the effective speed of the processor.

On POWER6 processor-based systems, accounting calculations need to factor to the extent of below-nominal-frequency usage of a processor by a program because of power management actions.

To achieve this the POWER6 processor contains an additional special-purpose register for each hardware thread called as Scaled processor utilization of resources register (SPURR). The SPURR is used to compensate for the effects of performance variability on the OSs: the hypervisor virtualizes the SPURR for each hardware thread so that each OS obtains accurate readings that reflect only the portion of the SPURR count that is associated with its partition. The implementation of virtualization for the SPURR is the same as that for the PURR.

Building on the functions provided by hypervisor, the OSs use SPURR to do same type of accurate accounting that is available on POWER5 processor-based machines. With the introduction of the EnergyScale architecture for POWER6 processor-based machines, not all timebase ticks have the same computational value; some of them represent more-usable processor cycles than others. The SPURR provides a scaled count of the number of timebase ticks assigned to a hardware thread, in which the scaling reflects the speed of the processor (taking into account frequency changes and throttling) relative to its nominal speed.

### ***System-wide tools modified for variable processor frequency***

The EnergyScale architecture may therefore affect some of the performance tools and metrics built with the user-visible performance counters. Many of these counters count processor cycles, and since the number of cycles per unit time is variable, the values reported by unmodified performance monitors would be subject to some interpretation.

The **lparstat** command has been updated in AIX Version 6.1 to display new statistics if the processor is not running at nominal speed. The *%nsp* metric shows the current average processor speed as a percentage of nominal speed. This field is also displayed by the new version of the **mpstat** command.

The **lparstat** command also displays new statistics if turbo mode accounting is disabled (the default) and the processor is running above nominal speed. The *%outcyc* field reflects the total percentage of unaccounted turbo cycles.

Finally, a new command parameter was added to **lparstat**. **lparstat -d** displays unaccounted turbo cycles in user, kernel, idle, and I/O wait modes if turbo mode accounting is disabled and the processor is running above nominal speed.

All of the other stat commands automatically switched to use SPURR-based metrics. Percentages below entitlement become relative to scaled (up and down) entitlement unless turbo mode accounting is off.

## 14.3 Cross-partition CPU monitoring

An interesting capability available in recent versions of the **topas** tool is cross-partition consumption. This command will only see partitions running AIX v5.3 TL3 or later. Virtual I/O Servers with version 1.3 or later are also reported.

**Important:** Cross partition CPU monitoring requires **perfagent.tools** and **bos.perf.tools** to be installed on partitions whose data has to be collected.

To see cross-partition report from Virtual I/O Server run **topas -cecdisp** as shown in Example 14-1.

*Example 14-1 topas -cecdisp command on Virtual I/O Server*

```

Topas CEC Monitor          Interval:  10                Mon Oct 13 14:39:18 2008
Partitions Memory (GB)    Processors
Shr:  3   Mon:11.5 InUse: 3.2 Shr:1.5 PSz:  3   Don:  0.0 Shr_PhysB  0.21
Ded:  1   Avl:   -         Ded:  1 APP:  2.8 Stl:  0.0 Ded_PhysB  0.00

Host      OS  M Mem InU Lp  Us Sy Wa Id  PhysB  Vcsw Ent  %EntC PhI
-----shared-----
NIM_server A61 C 2.0 1.1  4  97  1  0  0   0.20  386  0.20  99.6  0
DB_server  A61 U 4.5 0.8  4   0  0  0  99  0.01  403  1.00  0.6  2
VIO_Server1 A53 U 1.0 0.4  2   0  0  0  99  0.00  212  0.30  1.2  0

Host      OS  M Mem InU Lp  Us Sy Wa Id  PhysB  Vcsw %istl %bstl
-----dedicated-----
Apps_server A61 S 4.0 0.8  2   0  0  0  99  0.00  236   0.00  0.00

```

Once **topas** command runs, keyboard shortcuts behave like they work in standard AIX partition.

To see cross-partition report from virtual I/O client run **topas -C** as shown in Example 14-2.

*Example 14-2 topas -C command on virtual I/O client*

```

Topas CEC Monitor          Interval: 10          Mon Oct 13 17:35:07 2008
Partitions Memory (GB)    Processors
Shr: 4   Mon:11.5 InUse: 4.7 Shr:1.5 PSz: 3   Don: 0.0 Shr_PhysB 1.17
Ded: 1   Avl: -          Ded: 1 APP: 1.8 Stl: 0.0 Ded_PhysB 0.00

Host      OS  M Mem InU Lp  Us Sy Wa Id  PhysB  Vcsw Ent  %EntC PhI
-----shared-----
DB_server  A61 C 4.5 0.9 4  99 0 0 0   1.00  405  1.00  99.7  38
NIM_server A61 C 2.0 2.0 4  50 27 2 19  0.16  969  0.20  80.9  11
VIO_Server1 A53 U 1.0 1.0 2  0  3 0 96  0.01  667  0.30  4.5   0
VIO_Server2 A53 U 1.0 1.0 2  0  1 0 98  0.01  358  0.30  2.7   0
Host      OS  M Mem InU Lp  Us Sy Wa Id  PhysB  Vcsw %istl %bstl
-----dedicated-----
Apps_server A61 S 4.0 0.9 2  0  0 0 99  0.00  332  0.00  0.00

```

You can get lot of information about system and its partitions from this report. Taking an example of Example 14-2 various fields of this reports are described as following:

▶ *Partitions Shr=4, Ded=1*

The system has 5 active partitions; 4 use the shared processor pools.

▶ *Processors Psz = 3*

Three processors are in the same pool as this partition.

▶ *Processors Shr = 1.5*

Within the shared processor pool, 1.5 processor is currently allocated to the partitions.

▶ *Processors Ded = 1*

One processor is used for dedicated partitions.

▶ *Processors APP = 1.8*

Out of 3 processors in the shared processor pool, 1.8 is considered to be idle. Observe that  $Shr = 1.5$  and  $3 - 1.8 < 1.5$ , so it can be concluded that the shared partitions globally do not use all of their entitlements (allocated processing units).

▶ *Processors Shr\_PhysB = 1.17*

This represents the number of processors that are busy for all shared partitions. This confirms what we just observed with the *APP* field. The shared pool CPU consumption is lower than the entitlement of the shared partitions.

▶ *Processors Ded\_PhysB = 0.00*

The dedicated partition is currently idle.

- ▶ *Processors Don* = 0.0

This is related to a new feature introduced with POWER6 architecture. It represents the number of idle processor resources from the dedicated partitions that are donated to the shared processor pools.

To activate idle processor resources donation for dedicated partitions, you have to select the *Processor Sharing* options in the *Hardware* tab from the partition properties on the HMC. At the time of writing this book this feature was only available for POWER6 systems onward. It is also possible to disable *Processor Sharing* when partition is not active on POWER6 system unlike POWER5 system. Figure 14-4 shows how to do that.

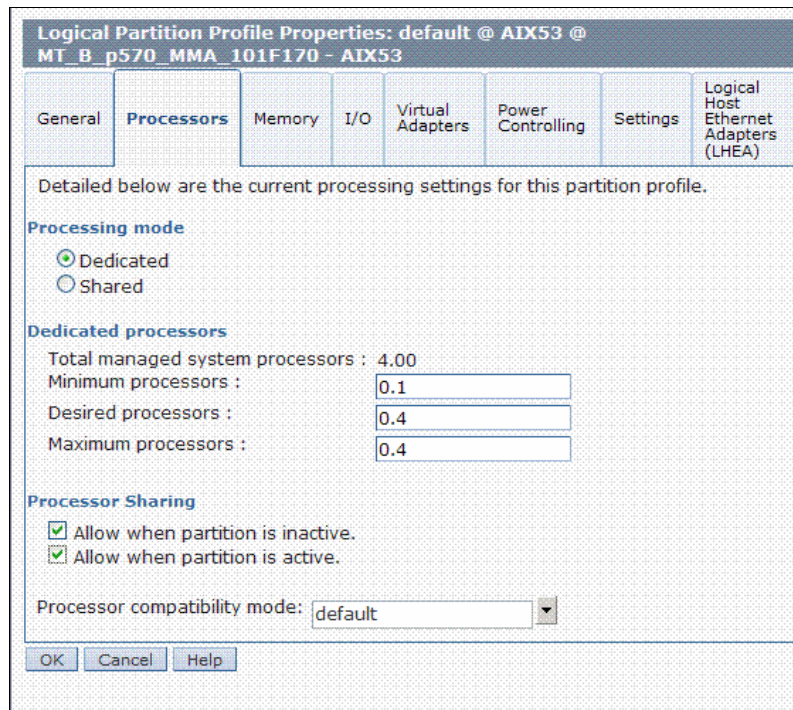


Figure 14-4 Dedicated partition's Processor Sharing properties

- ▶  $Stl = 0$  = Sum of stolen processor cycles from all partitions reported as a number of processors.
- ▶  $\%istl = 0.00$  - This shows the percentage of physical processors that is used while idle cycles are being stolen by the hypervisor. This metric is applicable only to dedicated partitions.

- ▶ *%bstl* = 0.00 - This shows the percentage of physical processors that is used while busy cycles are being stolen by the hypervisor. This metric is applicable only to dedicated partitions.

The partition section lists all the partitions that **topas** command can find in the CEC.

- ▶ The OS column indicates the type of operating system. In Example 14-2 on page 417, A61 indicates AIX Version 6.1.
- ▶ The M column shows the partition mode.
  - For shared partitions:
    - C - SMT enabled and capped
    - c - SMT disabled and capped
    - U - SMT enabled and uncapped
    - u - SMT disabled and uncapped
  - For dedicated partitions:
    - S - SMT enabled
    - d - SMT disabled and donating
    - D - SMT enabled and donating
    - ' ' (blank) - SMT disabled and not donating
- ▶ The other values are equivalent to those provided in the standard view.

Pressing **g** key in the **topas** command window expands global information, as shown in Example 14-3. A number of fields are not filled in this example, such as the total available memory. These fields have been reserved for an update to this command that will allow **topas** to interrogate the HMC to determine their values. It is possible to manually specify some of these values on the command line.

*Example 14-3 topas -C global information with the g command*

```

Topas CEC Monitor          Interval: 10          Mon Oct 13 14:10:19 2008
Partition Info   Memory (GB)   Processor   Virtual Pools :    0
Monitored  : 4   Monitored :12.5   Monitored  :2.7   Avail Pool Proc:  0.1
UnMonitored: -   UnMonitored: -   UnMonitored: -   Shr Physical Busy: 3.95
Shared      : 3   Available  : -   Available  : -   Ded Physical Busy: 0.00
Uncapped    : 2   UnAllocated: -   UnAllocated: -   Donated Phys. CPUs 1.00
Capped      : 2   Consumed   : 5.7   Shared      :1.7   Stolen Phys. CPUs : 0.00
Dedicated   : 1   Dedicated  : 1   Dedicated   : 1   Hypervisor
Donating    : 1   Donated    : 1   Donated     : 1   Virt. Context Switch:1026
2           Pool Size  : 4   Phantom Interrupts : 11

Host      OS  M Mem InU Lp  Us Sy Wa Id  PhysB  Vcsw  Ent  %EntC PhI
-----shared-----
DB_server  A61 U 4.5 0.9 8 99 0 0 0  3.93 9308  1.00 393.4 11
NIM_server A61 C 3.0 2.9 4  0 1 0 98  0.01 556  0.40  2.4  0
VIO_Server1 A53 U 1.0 1.0 2  0 0 0 99  0.00 168  0.30  0.9  0

Host      OS  M Mem InU Lp  Us Sy Wa Id  PhysB  Vcsw  %istl %bstl %bdon %idon
-----dedicated-----
Apps_server A61 D 4.0 0.8 2  0 0 0 99  0.00 230  0.01 0.00 0.00 99.69

```

**Note:** **topas -C** may not be able to locate partitions residing on other subnets. To circumvent this, create a `$HOME/Rsi.hosts` file containing the fully qualified hostnames for each partition (including domains), one host per line.

Cross partition monitoring feature is not provided by other CPU performance tools like **lparstat**, **vmstat**, **sar** and **mpstat**.

### Monitoring Multiple Shared Processor Pools (MSPP)

Multiple Shared-Processor Pools (MSPP) utilization can be monitored using the **topas -C** (**topas -cecdisp** for Virtual I/O Server) command. To do so, press **p** to see the processor pools section, as shown in Example 14-4.

**Note:** The ability to monitor the consumption of other partitions by **topas** is available only when the *Allow performance information collection* item is selected in the partition's profile.



*Example 14-4 Monitoring processor pools with topas -C*

```

Topas CEC Monitor          Interval: 10          Mon Oct 13 14:09:07 2008
Partitions Memory (GB)    Processors
Shr: 0   Mon:12.5 InUse: 6.1 Shr:1.7 PSz: 4   Don: 1.0 Shr_PhysB 0.57
Ded: 4   Avl:   -          Ded: 1 APP: 3.4 Stl: 0.0 Ded_PhysB 0.00
pool  psize  entc  maxc  physb  app  mem  muse
-----
0     4.0    230.0 400.0 0.0    0.00 1.0  1.0
1     2.0    140.0 200.0 0.6    1.52 7.5  3.9

Host      OS  M Mem InU Lp  Us Sy Wa Id  PhysB  Vcsw  %istl %bstl %bdon %idon
-----dedicated-----
Apps_server A61 D 4.0 1.2 2   0  0  0 99  0.00 231   0.00 0.00 0.00 99.66

```

In this report, you see the metrics presented in 14.1.2, “Terminology and metrics specific to POWER6 systems” on page 409.

- ▶ *maxc* represents the maximum pool capacity for each pool.
- ▶ *app* represents the number of available processors in the Shared-Processor Pool (SPP).
- ▶ *physb* represents the summation of physical busy of processors in shared partitions of a SPP.
- ▶ *mem* represents the sum of monitored memory for all shared partitions in the SPP.
- ▶ *muse* represents the sum of memory consumed for all shared partitions in the SPP.

**Note:** The pool with identifier 0 is the default Shared Processor Pool.

When the MSPP section is displayed, you can move using the up and down arrow keys to select a pool (its identifier highlights). Once a pool is selected, pressing the **f** key shows the metrics of the shared partitions belonging to the selected pool, as shown in Example 14-5.

*Example 14-5 Shared pool partitions listing in topas*

```

Topas CEC Monitor          Interval: 10          Mon Oct 13 14:34:05 2008
Partitions Memory (GB)    Processors
Shr: 2   Mon:12.5 InUse: 6.1 Shr:1.7 PSz: 4   Don: 1.0 Shr_PhysB 0.02
Ded: 2   Avl: -          Ded: 1 APP: 4.0 Stl: 0.0 Ded_PhysB 0.00
pool psize entc maxc physb app mem muse
-----
0    4.0   230.0 400.0 0.0   0.00 1.0 1.0
1    2.0   140.0 200.0 0.0   1.98 7.5 3.8

Host      OS  M Mem InU Lp  Us Sy Wa Id  PhysB Vcsw Ent  %EntC PhI
-----shared-----
NIM_server A61 C 3.0 2.9 4   0 0 0 98  0.01 518 0.40 1.9 0
DB_server  A61 U 4.5 0.9 4   0 0 0 99  0.01 401 1.00 0.5 0

Host      OS  M Mem InU Lp  Us Sy Wa Id  PhysB Vcsw %istl %bstl %bdon %idon
-----dedicated-----
Apps_server A61 D 4.0 1.2 2   0 0 0 99  0.00 225 0.00 0.00 0.00 99.65

```

## 14.4 AIX and Virtual I/O Server CPU monitoring

There are several commands to monitor CPU utilization for Virtual I/O Server and virtual I/O client (AIX). Each is covered in a separate section with reason why one should be used over the others. When referring to Virtual I/O Server it is assumed that command is executed from the restricted shell and not the root shell.

### 14.4.1 Monitoring using topas

The **topas** command gathers general information from different metrics on a partition. The major advantage of this tool is that you can see all of the important performance metrics of the partition in real time.

#### Basic display

In basic display mode **topas** is run without any argument as shown in Example 14-6.

*Example 14-6 Basic topas monitoring*

```

Topas Monitor for host:  DB_server          EVENTS/QUEUES  FILE/TTY
Mon Oct 13 15:55:13 2008  Interval: 2      Cswitch       60  Readch   190.9K
                               Syscall       430  Writch   1589
Kernel   0.4  |#                               Reads         54  Rawin    0
User   99.3 |#####                               Writes         3  Ttyout   250
Wait     0.0  |                               Forks          0  Igets    0
Idle    0.3  |#                               Execs         0  Namei    32
Physc = 1.00                               %Entc= 99.9 Runqueue  1.0  Dirblk   0
                               Waitqueue  0.0
Network  KBPS  I-Pack  O-Pack  KB-In  KB-Out          MEMORY
Total    1.2   5.0    1.0    0.3    0.9  PAGING         Real,MB  4608
                               Faults    93  % Comp   18.8
Disk     Busy%  KBPS    TPS  KB-Read  KB-Writ  Steals  0  % Noncomp  2.2
Total    0.0   0.0    0.0   0.0     0.0  Pgspln  0  % Client  2.2
                               Pgspln  0
FileSystem      KBPS    TPS  KB-Read  KB-Writ  PageIn  0  PAGING SPACE
Total            5.3    5.0   5.3     0.0  PageOut  0  Size,MB  512
                               Sios    0  % Used   1.1
Name          PID  CPU%  PgSp  Owner          % Free   99.9
doh         348400 49.7  0.5  root           NFS (calls/sec)
topas         250106  0.2   1.5  root           SerV2     0  WPAR Activ  0
xmgc          49176  0.0   0.4  root           Cliv2     0  WPAR Total  0
gil           69666  0.0   0.9  root           SerV3     0  Press: "h"-help
ksh           335982  0.0   0.5  root           Cliv3     0  "q"-quit
java          209062  0.0  80.3  pconsole
rpc.lock      262310  0.0   1.2  root
aixmibd       200806  0.0   1.0  root

```

You immediately get the most important CPU information for this partition:

- ▶ *%Entc* = 99.9 and remains steady upon refresh (every 2 seconds)  
This partition therefore uses all its entitlement.
- ▶ *Physc* = 1.0  
This partition consumes a whole CPU.
- ▶ *User* = 99.3%  
Most of the computation is spent in user mode.
- ▶ *Runqueue* = 1  
On average over 2s, only one thread is ready to run.
- ▶ *Waitqueue* = 0  
On average over 2s, no thread is waiting for paging to complete.

**Logical Partition display**

In Example 14-7 the process named *doh* uses 49.7% CPU and the other processes almost nothing. Yet *Physc* = 1.00 - Simultaneous multithreading is

probably activated and the **ksh** process is not multithreaded. You can check this by pressing **L** to toggle to the logical partition display as shown in Example 14-7. Alternatively, this view can also be accessed by **topas -L**. The upper section shows a subset of the **lparstat** command statistics while the lower part shows a sorted list of logical processors with a number of the **mpstat** command figures.

*Example 14-7 Logical partition information report in topas (press L)*

```

Interval:      2      Logical Partition: DB_server      Mon Oct 13 16:27:01 2008
Psize:        -      Shared SMT  ON      Online Memory: 4608.0
Ent: 1.00      Mode: Capped      Online Logical CPUs: 4
Partition CPU Utilization      Online Virtual CPUs: 2
%usr %sys %wait %idle physc %entc %lbusy app vcsw phint %hypv hcalls
  99  0  0  0  1.0 99.90 25.00 - 418 26 0.4 397
=====
LCPU minpf majpf intr csw icsw runq lpa scalls usr sys _wt idl pc lcsw
Cpu0 0 0 175 28 14 0 100 13 5 50 0 46 0.00 121
Cpu1 0 0 94 4 2 0 100 3 4 33 0 63 0.00 94
Cpu2 0 0 93 26 18 0 100 13 9 27 0 64 0.00 102
Cpu3 0 0 128 2 2 0 100 0 100 0 0 0 1.00 100

```

This additional report confirms the initial thoughts:

- ▶ *Shared SMT* = ON - This confirms that simultaneous multithreading (SMT) is active on this partition.
- ▶ You also notice that for the 2 virtual CPUs allocated to this partition, the SMT provides 4 logical processors.
- ▶ *%lbusy* = 25% - Only a quarter of the logical processors are effectively in use.
- ▶ The detailed logical CPU activity listing appearing at the bottom of the report also shows that only one of the logical CPUs is in use. This confirms that the **doh** process is probably single-threaded. At this point, the **ps** command would be used to get a deeper analysis of this process consumption.

### Processor subsection display

If you are interested in individual logical processor metrics, typing **c** twice in the main report screen moves to the individual processor activity report, as displayed on Example 14-8.

*Example 14-8 Upper part of topas busiest CPU report*

Topas Monitor for host:		DB_server		EVENTS/QUEUES		FILE/TTY	
Mon Oct 13 16:18:39 2008		Interval: 2		Cswitch	61	Readch	2
				Syscall	35	Writech	115
CPU	User%	Kern%	Wait%	Idle%	Physc	Reads	Writes
<b>cpu2</b>	<b>99.9</b>	0.1	0.0	0.0	<b>1.00</b>	1	Rawin
cpu1	4.6	35.4	0.0	60.1	0.00	1	Ttyout
cpu0	3.6	53.8	0.0	42.6	0.00	0	Igets
cpu3	0.0	17.8	0.0	82.2	0.00	0	Namei
				Runqueue	1.0	Dirblk	0

In this example, a single logical CPU is handling the whole partition load. By observing the values one can pinpoint some performance bottlenecks. Indeed if we consider the first report in Example 14-6 on page 423, *%Entc* is about 100% and does not exceed this value. You can then make the following suppositions:

- ▶ Because this partition also has *Physc* = 1.0, it may run on a dedicated processor.
- ▶ If it runs in a shared processor pool, it may be capped.
- ▶ If it runs in a shared processor pool, it may be uncapped but only have a single virtual processor defined. Indeed, a virtual processor cannot consume more processing units than a physical processor.
- ▶ If it runs in a shared processor pool, it may be uncapped but have reached the maximum processing units for this partition as defined in its active profile.

To move to logical partition report press L. In Example 14-7 on page 424, the partition is capped and therefore runs in shared mode.

In Example 14-7 on page 424, you may have noticed that the *Psize* field is not filled in. This value represents the global shared pool size on the machine, that is, the number of the processors that are not used for running dedicated partitions. To get this value, you have to modify the partition properties from the HMC or the IVM. Then select the Hardware tab and check the *Allow performance information collection* item and validate the change. The *Psize* value will then be displayed by **topas**.

## 14.4.2 Monitoring using nmon

Starting with the Virtual I/O Server release 2.1, AIX 5.3 TL9, and AIX 6.1 TL2, **nmon** is included in the default installation. In order to use **nmon**, in the Virtual I/O Server follow the steps described below:

1. Execute **topas** command at the shell prompt.

```

9.3.5.111 - PuTTY
Topas Monitor for host:   vios1
Mon Nov  3 09:11:28 2008 Interval:  2
EVENTS/QUEUES          FILE/TTY
Cswitch                220  Readch    323
Syscall                70  Writech  721
Reads                  1  Rawin    0
Writes                 3  Ttyout   381
Forks                  0  Igets    0
Execs                  0  Namei    1
Runqueue              0.0  Dirblk   0
Waitqueue              0.0

CPU  User%  Kern%  Wait%  Idle%  Physc  Entc
ALL  0.1    2.4    0.0    97.5   0.01   3.7

Network  KBPS  I-Pack  O-Pack  KB-In  KB-Out
Total    1.5   19.0    2.0     1.0    0.5

Disk  Busy%  KBPS  TPS  KB-Read  KB-Writ
Total 20.0  430.0 81.0 306.0 124.0

FileSystem  KBPS  TPS  KB-Read  KB-Writ
Total        0.3  1.0  0.3  0.0

PAGING          MEMORY
Real,MB        1024
% Comp         67.7
% Noncomp      18.8
% Client       18.8

PAGING SPACE
Size,MB        1536
% Used         0.0
% Free        100.0

NFS (calls/sec)
SerV2          0  WPAR Activ  0
Cliv2          0  WPAR Total  0
SerV3          0  Press: "h"-help
Cliv3          0  "q"-quit

Name      PID  CPU%  PgSp  Owner
syncd    147602  0.3  0.5  root
seaproc  163924  0.3  1.0  root
topas    348296  0.2  1.8  root
getty    213150  0.1  0.6  root
topas    311364  0.1  10.5  root
seaproc  159842  0.1  1.0  root
sched    12294  0.1  0.4  root
gil      57372  0.0  0.9  root
sshd     274632  0.0  0.9  padmin
cimserve 315602  0.0  69.4  root
accesspr 123060  0.0  0.8  root
netm     53274  0.0  0.4  root
IBM.CSMA 364764  0.0  2.5  root
errdemon 81982  0.0  0.7  root
ldmp_pro 86074  0.0  0.5  root
target_k 90172  0.0  0.5  root
acct_wri 94380  0.0  0.6  root
shlap64 98424  0.0  0.4  root
writesrv 102632  0.0  0.3  root
lvmbb    106578  0.0  0.4  root

```

Figure 14-5 topas displaying monitoring information.

- When the topas screen is displayed, type ~ (tilde) and this results in the display of the screen shown in Figure 14-6

```

-----
N   N   H   M   OOOO  N   N   For online help type: h
NN  N  MM  MM  O   O  NN  N   For command line option help:
N N N M MM M  O   O  N N N   quick-hint  nmon -?
N N N M   M  O   O  N N N   full-details nmon -h
N  NN H   M  O   O  N  NN   To start nmon the same way every time?
N   N  M   M  OOOO  N   N   set NMON ksh variable, for example:
-----
                                export NMON=cmt

TOPAS-NMON

                2 - CPUs currently
                2 - CPUs configured
                4208 - MHz CPU clock rate
PowerPC_POWER6 - Processor
                64 bit - Hardware
                64 bit - Kernel
                3,AIX61 - Logical Partition
                6.1.2.1 TL02 - AIX Kernel Version
                aix61 - Hostname
                aix61 - Node/WPAR Name
                101F170 - Serial Number
                IBM,9117-NMA - Machine Type

```

Figure 14-6 Initial screen of NMON application.

**Note:** To switch between the `nmon` and `topas` application displays type `~`.

- To get help on monitoring the available system resources type `?` and this results in the display of the help screen (see Figure 14-7)

```

9.3.5.115 - PuTTY
--HELP--most-keys-toggle-on/off
h = Help information      q = Quit nmon           O = reset peak counts
+ = double refresh time  - = half refresh       r = ResourcesCPU/HW/MHz/AIX
c = CPU by processor     C=upto 128 CPUs        p = LPAR Stats (if LPAR)
l = CPU avg longer term  k = Kernel Internal    # = PhysicalCPU 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) O = 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 proc  "="= Net/Disk KB<-->MB
b = black&white mode    g = User-Defined-Disk-Groups (see cmdline -g)
t = Top-Process --->    1=basic 2=CPU-Use 3=CPU(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)
~ = Switch to topas screen
Need more details? Then stop nmon and use: nmon -?
  
```

Figure 14-7 Display of command help for monitoring system resources

- Now use the letters to monitor the system resources. For example to monitor CPU activity, type `c` (Figure 14-8).

```

9.3.5.115 - PuTTY
--topas nmon -P=PagingSpace--Host=aix61--Refresh=2 secs--13:48.21--
CPU-Utilisation-Small-View EntitledCPU= 0.50 UsedCPU= 0.003
Logical CPUs 0-----25-----50-----75-----100
CPU User% Sys% Wait% Idle%|
0 0.0 0.0 0.0 100.0|>
1 0.0 0.0 0.0 100.0|>
EntitleCapacity/VirtualCPU +-----|-----+
EC 0.1 0.3 0.0 0.3|-----|
VP 0.0 0.2 0.0 0.2|-----|
EC= 0.7% VP= 0.3% +--No Cap---|-----|-----100% VP=1 CPU+
  
```

Figure 14-8 Monitoring CPU activity with NMON

- To monitor additional resources other than the one that you are currently monitoring, just type letter against that system resource (refer to the help screen in Figure 14-7). For example in addition to monitoring the CPU activity, to monitor the network resources type **n** and display adds the network associated resources activity to the existing display as shown in Figure 14-9

```

9.3.5.115 - PuTTY
--topas nmon -v=Verbose-hints Host=aix61 Refresh=2 secs 13:50.59
CPU-Utilisation-Small-View EntitledCPU= 0.50 UsedCPU= 0.003
Logical CPUs 0-----25-----50-----75-----100
CPU User% Sys% Wait% Idle%| | | |
0 0.0 0.0 0.0 100.0|> | | | |
1 0.0 0.0 0.0 100.0|> | | | |
EntitledCapacity/VirtualCPU +-----+-----+-----+-----+
EC 0.1 0.3 0.0 0.3|-----|-----|-----|-----|
VP 0.0 0.2 0.0 0.2|-----|-----|-----|-----|
EC= 0.7% VP= 0.3% +--No Cap--|-----|-----|-----|100% VP=1 CPU+
Network
I/F Name Recv=KB/s Trans=KB/s packin packout insize outside Peak->Recv TransKB
en0 0.4 0.1 7.0 1.0 57.0 129.0 1.0 1.2
lo0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
Total 0.0 0.0 0.0 in Mbytes/second Overflow=0
I/F Name MTU ierror oerror collision Mbits/s Description
en0 1500 0 0 0 2047 Standard Ethernet Network Interface
lo0 16896 0 0 0 0 Loopback Network Interface

```

Figure 14-9 NMON monitoring of CPU and network resources.

- To quit from the `nmon/topas` application type **q** and this will take the user back to the shell prompt.

### 14.4.3 Monitoring using `vmstat`

When you want to focus only on processor and memory consumption, the `vmstat` command is a good alternative to `topas` because you can see the short term evolution in the performance metrics on regular intervals.

This command works on Virtual I/O Server in similar way as it works for virtual I/O client with AIX.

Considering CPU consumption monitoring, the `vmstat` command shows trends for the partition consumption.

In Example 14-9, you see that the values remain stable during the observation period. This reflects steady system consumption. In this example utilization is printed every 5 seconds endlessly. If you want only certain number of iterations (lets say 10) then use `vmstat -wI 5 10`.

*Example 14-9 Monitoring with the `vmstat` command*

```
# vmstat -wI 5
```



System configuration: **lcpu=8** mem=4608MB **ent=1.00**

kthr			memory				page				faults				cpu				
r	b	p	avm	fre	fi	fo	pi	po	fr	sr	in	sy	cs	us	sy	id	wa	pc	ec
8	0	0	247142	943896	0	0	0	0	0	0	7	95	171	99	0	0	0	3.93	393.0
8	0	0	247141	943897	0	0	0	0	0	0	5	42	167	99	0	0	0	3.85	384.6
8	0	0	247141	943897	0	0	0	0	0	0	6	29	161	99	0	0	0	3.94	393.8
8	0	0	247141	943897	0	0	0	0	0	0	9	37	165	99	0	0	0	3.92	391.6
8	0	0	247141	943897	0	0	0	0	0	0	3	34	164	99	0	0	0	3.93	392.6
8	0	0	247141	943897	0	0	0	0	0	0	6	34	161	99	0	0	0	3.78	378.1
8	0	0	247141	943897	0	0	0	0	0	0	5	28	162	99	0	0	0	3.73	372.7

Most of the information can be displayed using **topas** command too (as shown in Example 14-6 and Example 14-7). Fields are described as:

Physical CPU consumption:

*pc* in **vmstat** (*PhysC* in **topas**)

Percentage of entitlement consumed:

*ec* in **vmstat** (*%EntC* in **topas**)

Consumption type: *us/syl/id/wa* in **vmstat** (*User/Kernel/Idle/Wait* in **topas**)

Logical CPUs: *lcpu* in **vmstat** (*Online Logical CPUs* in **topas**)

Partition entitlement: *ent* in **vmstat** (*Ent* in **topas**)

Runqueue: Field *r* in *kthr* column in **vmstat** (*RunQueue* in **topas**). It denotes the number of threads that are runnable, which includes threads that are running and threads that are waiting for the CPU.

Waitqueue: Field *b* in *kthr* column in **vmstat** (*WaitQueue* in **topas**). It denotes the average number of threads that are blocked either because they wait for a file system I/O or they were suspended due to memory load control

Threadqueue for I/O raw devices:

Field *p* in *kthr* column in **vmstat** (no corresponding field in **topas**). It shows the number of threads waiting on I/O to raw devices per second. This does not include threads waiting on I/O to file systems.

Just like in the **topas** tool, the first focus should be on the *r* field value to determine the number of virtual processors that could be used at some time.

#### 14.4.4 Monitoring using **lparstat**

The **lparstat** command provides an easy way to determine whether the CPU resources are optimally used. This command does not work for Virtual I/O Server. Example 14-10 shows how **lparstat** command can be used to display

CPU utilization metrics at an interval of 1 second for 2 iterations. If you want to obtain endless CPU utilization metrics without stopping at an interval of 1 sec use **lparstat -h 1**.

Most of the fields in Example 14-10 are same as explained for **topas** and **vmstat** commands in previous sections. Two unexplained fields are described as:

- ▶ *lbusy* shows the percentage of logical processor utilization that occurs while executing in user and system mode. If this value approaches 100% it may indicate that the partition could make use of additional virtual processors. In this example it is about 25%. Because *lcpu* is 4 and *%entc* is about 100%, the partition is probably running a single-threaded process consuming much CPU.
- ▶ *app* shows the number of available processors in the shared pool. Because in this example *psize* = 2, *ent* = 0.4 and *%entc* = 100, the remaining processing resource on the whole system would be around 1.6. As the *app* field is 1.59, you can conclude that the other partitions on this system consume almost no CPU resources.

**Note:** The *app* field is only available when the *Allow performance information collection* item is selected for the current partition properties.

#### Example 14-10 Monitoring using the lparstat command

```
# lparstat -h 1 2
System configuration: type=Shared mode=Capped smt=0n lcpu=4 mem=4096 psize=2 ent=0.40
%user %sys %wait %idle physc %entc lbusy app vcsw phint %hypv hcalls
-----
84.9 2.0 0.2 12.9 0.40 99.9 27.5 1.59 521 2 13.5 2093
86.5 0.3 0.0 13.1 0.40 99.9 25.0 1.59 518 1 13.1 490
```

You have access to all these metrics using **topas -L** too, but using **lparstat** you can see short term evolution of these values. For more information on **topas -L**, refer to 14.4.1, “Monitoring using topas” on page 422.

### Monitoring variable processor frequency

There are two main reasons for processor frequency to vary as described in [14.2.4, “Scaled Processor Utilization of Resources Register \(SPURR\)” on page 414](#). It can vary in following ways:

- ▶ Down - to control power consumption or fix a heat problem
- ▶ Up - to boost performance

The impact on the system can be monitored using the **lparstat** command. When the processor is not running at nominal speed the `%nsp` field is displayed showing the current average processor speed as a percentage of nominal speed.

This is illustrated in Example 14-11.

*Example 14-11 Variable processor frequency view with lparstat*

---

```
# lparstat
System configuration: type=Shared mode=Uncapped smt=0n lcpu=2 mem=432 psize=2 ent=0.50
%user %sys %wait %idle physc %entc lbusy vcswh phint %nsp %utcyc
-----
 80.5 10.2  0.0  9.3  0.90  0.5  90.5 911944 434236 110  9.1
# lparstat -d
System configuration: type=Shared mode=Uncapped smt=0n lcpu=2 mem=432 psize=2 ent=0.50
%user %sys %wait %idle physc %entc %nsp %utuser %utsys %utidle %utwait
-----
 70.0 20.2  0.0  9.7  0.5 100  110  5.01  1.70  2.30  0.09
```

---

In this example, we see that the processor is running above nominal speed because `%nsp > 100`.

Accounting is disabled by default in Turbo mode, it can be enabled using SMIT panel. If accounting is disabled and the processor is running above nominal speed, the `%utcyc` is displayed. It represents the total percentage of unaccounted turbo cycles. In this example, 9.1% of the cycles were unaccounted. Indeed if turbo accounting mode is disabled, the CPU utilization statistics are capped to the PURR values.

In Example 14-11 we also see new metrics displayed by **lparstat -d** because turbo mode accounting is disabled and the processor is running above nominal speed:

- ▶ `%utuser` shows the percentage of unaccounted turbo cycles in user mode.
- ▶ `%utsys` shows the percentage of unaccounted turbo cycles in kernel mode.
- ▶ `%utidle` shows the percentage of unaccounted turbo cycles in idle state.
- ▶ `%utwait` shows the percentage of unaccounted turbo cycles in I/O wait state.

## 14.4.5 Monitoring using sar

The **sar** command provides statistics for every logical processor. It can be used in 2 ways:

## Real time CPU utilization metrics

It can show real time CPU utilization metrics sampled at an interval for the specified number of iterations. Example 14-12 shows two iterations of CPU utilization for all the processors at an interval of 3 second. It is not possible to show continuous metrics without stopping unlike other commands.

### Example 14-12 Individual CPU Monitoring using the sar command

---

```
# sar -P ALL 3 2

AIX VI0_Server1 3 5 00C0F6A04C00    10/14/08

System configuration: 1cpu=2 ent=0.30 mode=Uncapped

11:33:28 cpu      %usr   %sys   %wio   %idle  physc  %entc
11:33:31  0        3     85     0     12    0.01   2.8
           1        0     36     0     63    0.00   0.4
           U        -     -      0     96    0.29  96.8
           -        0     3      0     97    0.01   3.2
11:33:34  0        4     63     0     33    0.00   1.1
           1        0     35     0     65    0.00   0.4
           U        -     -      0     98    0.30  98.5
           -        0     1      0     99    0.00   1.5

Average  0        3     79     0     18    0.01   1.9
           1        0     36     0     64    0.00   0.4
           U        -     -      0     97    0.29  97.6
           -        0     2      0     98    0.01   2.4
```

---

You see in Example 14-12 that the activity of individual logical processors is reported. The U line shows the unused capacity of the virtual processor.

You have information we already saw using the **topas** command:

- ▶ *physcc* in **sar** = *pc* in **mpstat** = *PhysC* in **topas** = physical CPU consumption
- ▶ *%entc* in **sar** = *%ec* in **mpstat** = *%EntC* in **topas** = percentage of entitlement consumed

## CPU utilization metrics from a file

It can extract and show previously saved CPU utilization metrics which was previously saved in a file (*/var/adm/sa/sadd*, where *dd* refers to current day). The system utilization information is saved by two shell scripts (*/usr/lib/sa/sa1* and */usr/lib/sa/sa2*) running in background. These shell scripts are started by the cron daemon using crontab file */var/spool/cron/crontabs/adm*.

Collection of data in this manner is useful to characterize system usage over a period of time and determine peak usage hours.

Example 14-13 shows **sar** command working on a previously saved file.

*Example 14-13* sar command working a previously saved file

```
# ls -l /usr/adm/sa/
total 112
-rw-r--r-- 1 root system 21978 Nov 03 10:25 sa03
-rw-r--r-- 1 root system 26060 Oct 30 17:04 sa30
-rw-r--r-- 1 root system 780 Nov 03 10:25 sar03
# sar -f /usr/adm/sa/sa03
```

AIX aix61 1 6 00C1F1704C00 11/03/08

System configuration: lcpu=2 ent=0.50 mode=Uncapped

10:25:09	%usr	%sys	%wio	%idle	physc	%entc
10:25:12	1	1	0	98	0.01	2.3
10:25:13	0	1	0	99	0.01	1.1
10:25:14	0	0	0	100	0.00	0.7
10:25:15	0	0	0	100	0.00	0.8
10:25:16	0	0	0	100	0.00	0.7
10:25:17	0	0	0	100	0.00	0.8
10:25:18	0	0	0	100	0.00	0.7
10:25:19	0	0	0	100	0.00	0.8
10:25:20	0	0	0	100	0.00	0.7
10:25:21	0	0	0	100	0.00	0.7

Average 0 0 0 99 0.01 1.1

```
# sar -P ALL -f /usr/adm/sa/sa03
```

AIX aix61 1 6 00C1F1704C00 11/03/08

System configuration: lcpu=2 ent=0.50 mode=Uncapped

10:25:09	cpu	%usr	%sys	%wio	%idle	physc	%entc
10:25:12	0	55	35	0	11	0.01	2.0
	1	0	50	0	50	0.00	0.3
	U	-	-	0	98	0.49	97.7
	-	1	1	0	98	0.01	2.3
10:25:13	0	18	57	0	25	0.00	0.8
	1	0	49	0	51	0.00	0.3
	U	-	-	0	99	0.49	98.9
	-	0	1	0	99	0.01	1.1
10:25:14	0	5	50	0	44	0.00	0.5
	1	0	47	0	53	0.00	0.3
	U	-	-	0	99	0.50	99.3
	-	0	0	0	100	0.00	0.7
10:25:15	0	6	51	0	43	0.00	0.5
	1	0	47	0	53	0.00	0.3

	U	-	-	0	99	0.50	99.2
	-	0	0	0	100	0.00	0.8
.....							
.....							
Average	0	32	43	0	25	0.00	0.8
	1	0	48	0	52	0.00	0.3
	U	-	-	0	99	0.49	98.9
	-	0	0	0	99	0.01	1.1

---

**sar** command does not work for Virtual I/O Server.

## 14.4.6 Monitoring using mpstat

The **mpstat** command provides the same information as **sar** but it also provides additional information on the run queue, page faults, interrupts, and context switches. If you are not interested in that kind of information, using the **sar** command provides a simpler output.

Both commands can also provide a lot more information when using additional parameter flags. Refer to the commands man page for more information.

Note that the **mpstat** command displays these specific metrics:

- ▶ *mig* - number of thread migrations to another logical processor
- ▶ *lpa* - number of re-dispatches within affinity domain 3
- ▶ *ics* - involuntary context switches
- ▶ *mpc* - number of mpc interrupts. These are proactive interrupts used to ensure rapid response to a cross-CPU preemption request when the preempting thread is considered a “real time” thread.
- ▶ *lcs* - logical CPU context switches

In Example 14-14 you see that there are some standard logical CPU context switches. Yet no thread was forced to migrate to another logical processor. In this example it will display output endlessly without stopping. To use **mpstat** for only specified number of iterations (lets say 10) use **mpstat 3 10**.

*Example 14-14 Individual CPU Monitoring using the mpstat command*

---

```
# mpstat 3
```

```
System configuration: 1cpu=2 ent=0.3 mode=Uncapped
```

```
cpu min maj mpc int cs ics rq mig lpa sysc us sy wa id pc %ec lcs
0 0 0 0 141 117 61 0 0 100 45 3 64 0 33 0.00 1.0 145
1 0 0 0 182 0 0 0 0 - 0 0 33 0 67 0.00 0.4 130
```

```

U      -  -  -  -  -  -  -  -  -  -  -  -  -  0 98 0.29 98.0  -
ALL    0  0  0 323 117  61  0  0 100 45 0 1 0 99 0.00 1.4 275
-----
0      1  0  0 158 172  88  0  0 100 53 5 63 0 32 0.00 1.2 169
1      0  0  0 194  0  0  0  0  -  0 0 34 0 66 0.00 0.4 140
U      -  -  -  -  -  -  -  -  -  -  -  -  -  0 98 0.29 97.8  -
ALL    1  0  0 352 172  88  0  0 100 53 0 1 0 99 0.00 1.6 309
-----

```

`mpstat` does not work for Virtual I/O Server.

### 14.4.7 Report generation for CPU utilization

At the time of writing this book, report generation was not an option provided for Virtual I/O server through restricted shell. In order to start report generation you need to enable that from root shell. Report generation will work on virtual I/O client with AIX.

#### Continuous CPU monitoring using `topas`

Data collection in `topas` for continuous monitoring can be enabled, so that statistics can be collected and an idea of the load on the system can be developed. This functionality was introduced in AIX V5.3 TL 4.

This data collection involves several commands. The `xmwl` command is started with the `inittab`. It records data such as CPU usage, memory, disk, network and kernel statistics. The `topasout` command generates reports based on the statistics recorded by the `xmwl` command. The `topas -R` command may also be used to monitor cross-partition activities.

The easiest way to quickly set up data collection with `topas` is to use the associated SMIT interface.

In order to take a report you need to start recording first (unless its already started) as shown in Figure 14-10 using the **smitty topas** command.

```
Topas
Move cursor to desired item and press Enter.

Add Host to topas external subnet search file (Rsi.hosts)
List hosts in topas external subnet search file (Rsi.hosts)

List Available Recordings
Start New Recording
Stop Recording
List completed recordings
Generate Report

F1=Help          F2=Refresh      F3=Cancel      F8=Image
F9=Shell         F10=Exit       Enter=Do
```

Figure 14-10 smitty topas for CPU utilization reporting

From there you can select the kind of recording you want for report generation. In this example we have chosen CEC recording. You need to specify the path



where the recording output file will be saved in **Output Path** field as shown in Figure 14-11.

```

Start Local CEC recording

Type or select values in entry fields.
Press Enter AFTER making all desired changes.

                                     [Entry Fields]
Type of Recording                    cec
Length of Recording                  hour
* Recording intervals in seconds     [60] #
* Number of Samples                  [60] #
Output Path                          [/etc/perf/daily]

F1=Help      F2=Refresh      F3=Cancel      F4=List
F5=Reset     F6=Command     F7=Edit       F8=Image
F9=Shell     F10=Exit         Enter=Do

```

Figure 14-11 Local CEC recording attributes screen

Following types of recordings can be selected through the interface:

- |                 |  |
|-----------------|--|
| CEC Recording   | It monitors cross-partition activity (only available on AIX 5.4 TL5 and later). It starts <b>topas -R</b> daemon. These reports are only available when the <i>Allow performance information collection</i> item is selected for the partition on the HMC. |
| Local Recording | It monitors the current partition activity. It starts the <b>xmwl1m -L</b> daemon. This daemon is active by default.   |

Once it successfully completes you can start generating the report from **smitty topas** → **Generate Report** → **Filename** or **Printer** and specify the path as

shown in Figure 14-12. You might have to wait for some time so that some data is logged into the report file.

```

                                Path to locate the recording File

Type or select a value for the entry field.
Press Enter AFTER making all desired changes.

* Path to Locate the recording file [Entry Fields]
                                     [ /etc/perf/daily ] +

F1=Help      F2=Refresh      F3=Cancel      F4=List
F5=Reset     F6=Command     F7=Edit       F8=Image
F9=Shell     F10=Exit       Enter=Do

```

Figure 14-12 Report generation

After that you need to specify the **Reporting Format** on the next screen. Using **F4** you can see all the available formats as shown in Figure 14-13.

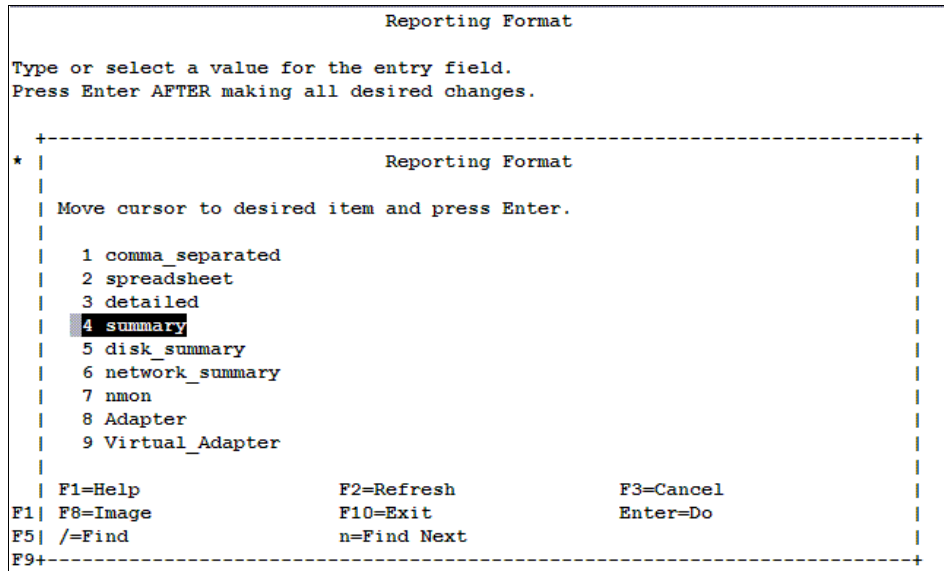


Figure 14-13 Reporting Formats

### Formats of various Reports

#### ► comma\_separated

```

#Monitor: xmtrend recording--- hostname: Apps_server ValueType: mean
Time="2008/14/10 17:25:33", CPU/gluser=19.62
Time="2008/14/10 17:25:33", CPU/glkern=2.48
Time="2008/14/10 17:25:33", CPU/glwait=12.83
. . .

```

#### ► spreadsheet - produces a file that can be opened by spreadsheet software.

```

"#Monitor: xmtrend recording --- hostname: Apps_server" ValueType: mean
"Timestamp" "/Apps_server/CPU/gluser" "/Apps_server/CPU/glkern" . . . .
"2008/14/10 17:25:33" 19.62 2.48 12.83 65.07 1.00 6.10 0.00 . . .
"2008/14/10 17:30:33" 0.04 0.24 0.06 99.66 1.00 6.10 0.00 . . .
. . .

```

#### ► detailed - provides a detailed view of the system metrics.

```

#Report: System Detailed --- hostname: Apps_server version:1.2

```

```

Start:14/10/08 17:25:33 Stop:14/10/08 23:50:26 Int: 5 Min Range: 384 Min

```

```

Time: 17:30:32 -----

```

CONFIG	CPU		MEMORY	PAGING		
Mode	Ded	Kern	Sz,GB	4.0	Sz,GB	0.5
LP	2.0	User	InU	0.8	InU	0.0
SMT	ON	Wait	%Comp	19.2	Flt	2382
Ent	0.0	Idle	%NonC	2.8	Pg-I	404

```

Poolid   -   PhyB   22.1   %CInt   2.8   Pg-0   4
          Entc   0.0

PHYP          EVENTS/QUEUES   NFS
Bdon   0.0   Cswth   602   SrvV2   0
Idon   0.0   Syscl   6553   CltV2   0
Bstl   0.0   RunQ    1   SrvV3   0
Istl   0.0   WtQ    2   CltV3   0
Vcsw   456.4
Phint   0.0

Network   KBPS   I-Pack   O-Pack   KB-I   KB-O
en0       1.6    10.4    5.1    1.0    0.6
lo0       0.2    0.8    0.9    0.1    0.1
. . .

```

- **summary** - presents the consolidated view of system information.

```

Report: System Summary --- hostname: Apps_server          version:1.2
Start:14/10/08 17:25:33 Stop:14/10/08 23:50:26 Int: 5 Min Range: 384 Min
Mem: 4.0 GB Dedicated SMT: ON Logical CPUs: 2
Time      InU  Us  Sy  Wa  Id  PhysB  RunQ  WtQ  CSwitch  Syscall  PgFault  %don  %stl
-----  -
17:30:32  0.8  20  2  13  65  22.10  1.2  1.7    602    6553    2382    0.0  0.0
17:35:32  0.8  0  0  0  100  0.28  0.7  0.2    164     46     6    0.0  0.0
17:40:32  0.8  0  0  0  100  0.36  1.2  0.0    167     74    25    0.0  0.0
. . .

```

- **disk\_summary** - provides information about the amount of data that is read or written to disks.

```

Report: Total Disk I/O Summary --- hostname: Apps_server
version:1.1
Start:14/10/08 17:25:33 Stop:14/10/08 23:50:26 Int: 5 Min Range: 384 Min
Mem: 4.0 GB Dedicated SMT: ON Logical CPUs: 2
Time      InU   PhysB  MBPS   TPS   MB-R   MB-W
17:30:32  0.8   22.1   1.1  132.1  1.0   0.1
17:35:32  0.8    0.3   0.0   1.1   0.0   0.0
17:40:32  0.8    0.4   0.0   0.6   0.0   0.0
. . .

```

- **network\_summary** - provides information about the amount of data that is received or sent by the network interfaces.

```

#Report: System LAN Summary --- hostname: Apps_server          version:1.1
Start:14/10/08 17:25:33 Stop:14/10/08 23:50:26 Int: 5 Min Range: 384 Min
Mem: 4.0 GB Dedicated SMT: ON Logical CPUs: 2
Time      InU   PhysB  MBPS  MB-I  MB-O  Rcvdrp  Xmdtrp
17:30:33  0.8   22.1   0.0   0.0   0.0    0    0
17:35:33  0.8    0.3   0.0   0.0   0.0    0    0
17:40:33  0.8    0.4   0.0   0.0   0.0    0    0
. . .

```

- ▶ **nmon** - generates a nmon analyzer report that can be viewed with the nmon analyzer as described in 14.4.2, “Monitoring using nmon” on page 425.

```
CPU_ALL,CPU Total ,User%,Sys%,Wait%,Idle%,CPUs,
CPU01,CPU Total ,User%,Sys%,Wait%,Idle%,
CPU00,CPU Total ,User%,Sys%,Wait%,Idle%,
CPU03,CPU Total ,User%,Sys%,Wait%,Idle%,
CPU02,CPU Total ,User%,Sys%,Wait%,Idle%,
DISKBUSY,Disk %Busy ,hdisk0,cd0,cd1,hdisk2,hdisk1,cd0,hdisk2,hdisk1,
DISKREAD,Disk Read kb/s ,hdisk0,cd0,cd1,hdisk2,hdisk1,cd0,hdisk2,hdisk1,
. . .
```

You can also use the **topasout** command directly to generate the reports. See its man page for more information.

**Note:** The **topas -R** and **xmwl**m daemons store their recordings into the `/etc/perf` directory by default. **topas -R** stores its reports into files with the `topas_cec.YYMMDD` name format. **xmwl**m uses the `daily/xmwl`m.YYMMDD file name format.

Recordings cover single-day periods (24 hours) and are retained for 2 days before being automatically deleted in AIX V5.3 TL4. This was extended to 7 days in AIX V5.3 TL5. This consequently allows a week’s worth of data to be retained on the system at all times.

The **topasout** command can generate an output file that can be transmitted to the nmon analyzer tool. It uses the file generated by the **xmwl**m daemon as input.

To generate output that can be transmitted to the nmon analyzer tool, run the following commands:

- ▶ To process a **topas** cross-partition report, run **topasout -a** with the report file:

```
# topasout -a /etc/perf/topas_cec.071205
```

The result file is then stored into `/etc/perf/topas_cec.071205.csv`.

- ▶ To process a local **xmwl**m report, run **topasout -a** with the report file:

```
# topasout -a /etc/perf/daily/xmwlm.071201
```

The result file is then stored into `/etc/perf/daily/xmwl`m.071201.csv.

You then FTP the resulting csv file to your station running Microsoft Excel (using the ASCII or TEXT options).

Now open the `nmon_analyser` spreadsheet, select **Analyse nmon data** and select the csv file you just transferred.

It then generates several graphs ready for you to study or write a performance report.

### **Continuous CPU monitoring using sar**

`sar` command can be used to display CPU utilization information in a report format. This report information is saved by two shell scripts which are started in background by a crontab job. For more information refer to the section “CPU utilization metrics from a file” on page 432.

## **14.5 IBM i CPU monitoring**

Since IBM i client CPU allocations and usage are not visible for Virtual I/O Server cross-partition monitoring tools like `topas` we show some examples for IBM i CPU monitoring using native IBM i tools.

### **Real-time CPU monitoring on IBM i**

For real-time CPU monitoring on IBM i the `WRKSYSACT` command can be used shown in Figure 14-14.

```

Work with System Activity                                E101F170
                                                    10/29/08 17:05:43
Automatic refresh in seconds . . . . . 5
Job/Task CPU filter . . . . . .10
Elapsed time . . . . . : 00:00:11  Average CPU util . . . . . : 53.4
Virtual Processors . . . . . : 2      Maximum CPU util . . . . . : 56.3
Overall DB CPU util . . . . . : 25.1  Minimum CPU util . . . . . : 50.4
                                           Current processing capacity: 1.00

Type options, press Enter.
  1=Monitor job   5=Work with job

      Job or
Opt  Task      User      Number  Thread  Pty   CPU   Total  Total  DB
      Task      User      Number  Thread  Pty   Util  I/O    I/O    Util
      ASP010001 QDEXUSER  013366  00000002  90   20.2  1257  14917  12.1
      ASP010003 QDEXUSER  013368  00000001  9    18.6  1146  24916  10.7
      ASP010002 QDEXUSER  013367  00000001  9    4.8   746   13793  1.0
      ASP010004 QDEXUSER  013369  00000001  10   3.9   374   11876  1.4
      SMIOSTCPGF          0      .3     0     0     .0
      QTSMTPLTD  QTCP      013335  00000002  35   .1    0     0     .0
                                           More...
F3=Exit  F10=Update list  F11=View 2  F12=Cancel  F19=Automatic refresh
F24=More keys

```

Figure 14-14 IBM i WRKSYSACT command output

The Work with System Activity screen shows the current physical and virtual processor allocation to the IBM i partition and CPU utilization percentages for average, maximum and minimum utilizations as well as the CPU utilization by all jobs performing database processing work (Overall DB CPU util). Additionally a list of jobs or system tasks sorted with the ones consuming the most processing time listed at the top is shown. Selecting **F10=Update** list refreshes the display with current information – the refresh interval should be 5s or longer so the sampled data is statistically meaningful. Using the **F19=Automatic refresh** function for automatically refreshing the display with the specified interval duration in **Automatic refresh in seconds** allows to easily identify jobs consuming a lot of processing time as they would appear repeatedly in the top part of the list.

**Note:** The CPU utilization can actually exceed 100% for an *uncapped* IBM i partition using a shared processor pool and reach up to the percentage value for the number of virtual processors.

For further information about IBM i performance management tools refer to *IBM eServer iSeries Performance Management Tools*, REDP-4026.

## Long-term CPU monitoring on IBM i

For long-term monitoring of CPU usage the *IBM Performance Tools for i5/OS* licensed program (5761-PT1) can be utilized which allows to generate a variety of reports from QAPM\* performance database files created from Collection Services data. IBM Performance Tools for i5/OS functions are accessible on IBM i 5250 sessions via a menu using the **STRPFRT** or **GO PERFORM** command and via native CL commands like **PRTSYSRPT**, **PRTCPTTRPT**, **PRTACTRPT** etc. Example 14-15 shows a spool file output from a *component report for component interval activity* we created via the following command:

```
PRTCPTTRPT MBR(Q302160328) TYPE(*INTERVAL)
```

### Example 14-15 IBM i Component Report for Component Interval Activity

Component Report															102908 22:17:3 Component Interval Activity									
Page															Member . . . : Q302160328 Model/Serial . . : MMA/10-1F170 Main storage . . : 8192.0 MB Started . . . . : 10/28/08									
16:03:2															Library . . . : QPFRDATA System name . . :E101F170 Version/Release : 6/ 1.0 Stopped . . . . : 10/28/08									
17:15:0															Partition ID : 005 Feature Code . :5622-5622 Int Threshold . : .00 %									
Virtual Processors: 2 Processor Units : 1.00															Uncap Int Int DB ----- Disk I/O ----- High Pool									
Excp	Itv	Tns	Rsp	DDM	-CPU Utilization-			CPU	Feat	CPU	Cpb	----- Per Second -----		-- Util --	-Faults/Sec-									
End	/Hour	/Tns	I/O	Total	Inter	Batch	Avail	Util	>Thld	Util	Sync	Async	Disk	Unit	Mch	User	ID							
Second	-----																							
16:05	0	.00	0	72.2	.0	72.2	199.9	.0	0	8.2	1127.4	189.6	95	0002	99	1081	02							
114.4																								
16:10	1056	.07	0	8.7	.0	8.7	200.0	.0	0	.0	155.6	26.1	16	0002	4	92	02							
12.6																								
16:15	492	.14	0	8.0	.0	8.0	199.9	.0	0	2.1	31.1	78.9	5	0002	1	5	02							
8.5																								
16:20	300	.04	0	52.8	.0	52.8	200.0	.0	0	42.4	121.2	6287.0	78	0001	0	5	02							
3.4																								
16:25	0	.00	0	48.8	.0	48.8	200.0	.0	0	44.0	133.2	6601.0	78	0001	0	5	02							
1.2																								
16:30	24	.00	0	50.0	.0	50.0	200.0	.0	0	45.5	129.2	6764.7	78	0001	0	0	02							
1.5																								
16:35	876	.08	0	49.6	.0	49.6	200.0	.0	0	45.0	137.6	6814.3	78	0001	1	2	02							
.9																								
16:40	348	.00	0	48.5	.0	48.5	200.0	.0	0	44.2	132.4	6628.2	79	0001	0	0	03							
.8																								
16:45	204	.00	0	48.5	.0	48.5	199.9	.0	0	43.9	128.9	6618.3	78	0001	0	0	02							
.5																								
More...																								
...																								
Itv End	-- Interval end time (hour and minute)																							
Tns /Hour	-- Number of interactive transactions per hour																							
Rsp /Tns	-- Average interactive transaction response time in seconds																							
DDM I/O	-- Number of logical DB I/O operations for DDM server jobs																							



Total CPU Utilization -- Percentage of available CPU time used by interactive and batch jobs. This is the average of all processors

Inter CPU Utilization -- Percentage of available CPU time used by interactive jobs. This is the average of all processors

Batch CPU Utilization -- Percentage of available CPU time used by batch jobs. This is the average of all processors

Uncap CPU Avail -- Percentage of CPU time available to this partition in the shared processors pool during the interval in addition to its configured CPU. This value is relative to the configured CPU available for this partition.

Int Feat Util -- Percentage of interactive feature used by all jobs

Int CPU >Thld -- Interactive CPU time (in seconds) over threshold

DB Cpb Util -- Percentage of database capability used to perform database processing

...

It is recommended to regularly monitor the CPU utilization to be able to take proactive measures like ending jobs or adding processor allocations to prevent response time impacts for the users. As a performance rule of thumb the IBM i average CPU utilization for high priority jobs (RUNPTY of 25 or less) depending on the number of available physical processors as derived from queuing theory should be below the percentage values shown in Table 14-3.

Table 14-3 IBM i CPU utilization guidelines

Number of Processors	CPU utilization guideline
1-way	70
2-way	76
3-way	79
4-way	81
6-way	83
8-way	85
12-way	87
16-way	89
18-way	90
24-way	91
32-way	93

To monitor the cumulative CPU utilization according to job type run priority values the IBM Performance Tools for i5/OS *system report for resource utilization expansion* can be used (available via the IBM Performance Tools for i5/OS Manager Feature, 5761-PT1 option 1) as shown in Example 14-16 we created using the following command:

PRTSYSRPT MBR(Q302160328) TYPE(\*RSCEXPN)

Our example resource utilization expansion report e.g. shows that all high run priority jobs with run priority equal or less than 25 consumed 7.9% CPU utilization while all jobs and system threads including the lower priority ones used 44.8% of available processing time in average for the selected report time frame on Oct. 28th between 16:03 and 17:15.

*Example 14-16 IBM i System Report for Resource Utilization Expansion*

System Report		10/29/08 22:25:3							
		Resource Utilization Expansion							
Page 000									
Member . . . : Q302160328 Model/Serial . . : MMA/10-1F170		Main storage . . : 8192.0 MB		Started . . . . : 10/28/08		16:03:2			
Library . . . : QPFRDATA		System name . . : E101F170		Version/Release : 6/ 1.0		Stopped . . . . : 10/28/08		17:15:0	
Partition ID : 005		Feature Code . . : 5622-5622		Int Threshold . . : .00 %					
Virtual Processors: 2		Processor Units : 1.00							
Job		CPU	Cum	----- Disk I/O -----		---- CPU Per I/O ----		----- DIO	
/Sec	-----	Util	Util	Faults	Sync	Async	Sync	Async	Sync
Pty	Type								
Async									
---	-----	-----	-----	-----	-----	-----	-----	-----	-----
000	System	4.9	4.9	6,065	18,744	16,033	.0113	.0132	4.3
3.7									
001	Batch	.0	4.9	7,348	10,090	699	.0000	.0007	2.3
.1									
009	System	.0	4.9	13	61	12	.0003	.0016	.0
.0									
010	Batch	.1	5.0	519	715	110	.0072	.0472	.1
.0									
015	System	.0	5.0	73	88	16	.0004	.0023	.0
.0									
016	Batch	.1	5.2	5,049	5,205	67	.0014	.1163	1.2
.0									
	System	.0	5.2	11	209	95	.0000	.0000	.0
.0									
020	PassThru	.0	5.2	1,735	2,661	148	.0001	.0030	.6
.0									
	Batch	.1	5.4	6,230	11,810	3,821	.0006	.0021	2.7
.8									
	AutoStart	.0	5.4	0	0	0	.0000	.0000	.0
.0									
	System i Access-Bch	.1	5.5	3,046	47,192	10,462	.0000	.0004	11.0
2.4									
	System	.0	5.5	9,509	16,482	2,890	.0000	.0002	3.8
.6									
021	Batch	.0	5.5	277	779	221	.0002	.0007	.1
.0									
	AutoStart	.0	5.5	87	137	104	.0002	.0002	.0
.0									
025	Batch	2.3	7.9	47,916	87,144	15,399	.0011	.0065	20.3
3.5									
	AutoStart	.0	7.9	0	0	0	.0000	.0000	.0
.0									
030	Batch	.0	7.9	0	0	0	.0000	.0000	.0
.0									
035	Batch	.0	7.9	547	651	67	.0029	.0283	.1
.0									
036	System	.0	7.9	0	0	0	.0000	.0000	.0
.0									

060	System	.0	8.4	0	0	0	.0000	.0000	.0
.0									
090	Batch	36.3	44.7	1,743	459,864	24,100,002	.0033	.0000	107.1
5617.7									
099	System	.0	44.8	268	1,790	53,395	.0008	.0000	.4
12.4									
Total				120,951	713,429	24,225,864			166.3
5647.0									

For further information about IBM Performance Tools for i5/OS refer to *IBM eServer iSeries Performance Tools for iSeries*, SC41-5340.

The *IBM Systems Director Navigator for i5/OS* graphical user interface which is accessible via [http://IBM\\_i\\_server\\_IP\\_address:2001](http://IBM_i_server_IP_address:2001) can be utilized to graphically display long-term monitoring data like CPU utilization. For generating the CPU utilization and waits overview graph shown in Figure 14-15 we selected **i5/OS Management** → **Performance** → **Collections**, chose a collection services DB file, selected **Investigate Data** from the popup menu and selected **CPU Utilization and Waits Overview** from the Select Action menu of the Resource Utilization Rates diagram.

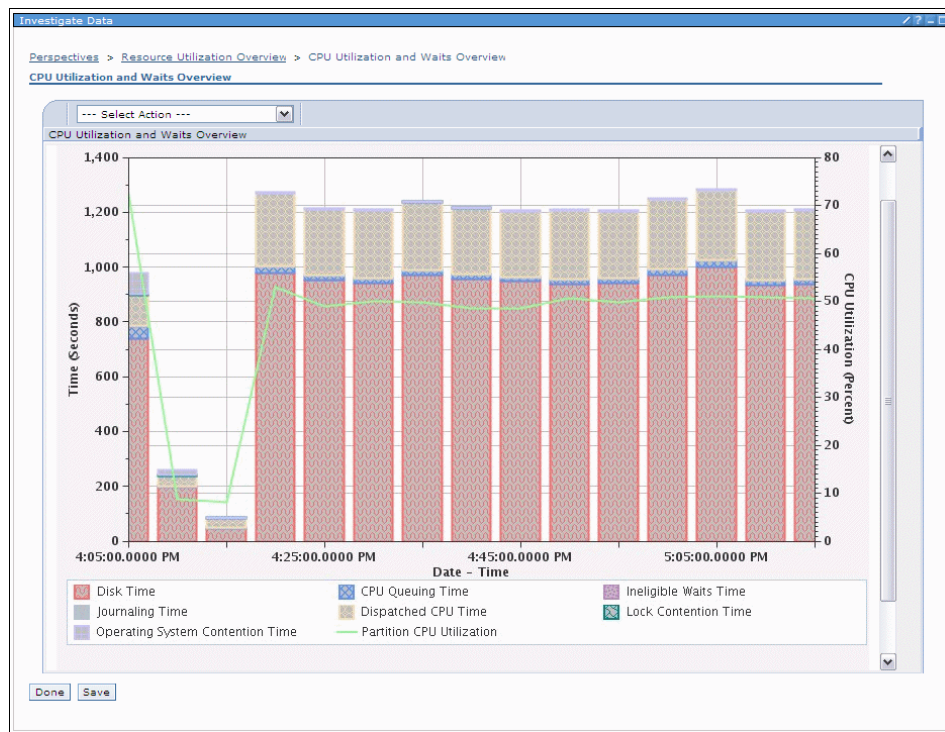


Figure 14-15 IBM i CPU Utilization and Waits Overview

As shown in the chart we can see the CPU utilization over time stabilized at around 50% as represented by the green line. Looking at where our jobs spent the most time we can see that we were running a disk I/O intensive workload with jobs having spent most of their time for disk I/O and only a fraction of about 20% on CPU processing (dispatched CPU time). If you were experiencing performance problems looking at where your jobs spent most of their time easily shows the area (e.g. disk vs. CPU resource, lock contention etc.) for which improvements would help most.

Another approach for long-term CPU monitoring for IBM i which also allows IBM i cross-partition monitoring is using the *System i Navigator's* Management Central system monitors function. The user may define a threshold for the CPU utilization of an IBM i system or group of systems which when exceeded for a specified number of intervals could trigger an event like sending a system operator message.

For further information about using System i Navigator for performance monitoring refer to *Managing OS/400 with Operations Navigator V5R1 Volume 5: Performance Management*, SG24-6565.

## 14.6 Linux for Power CPU monitoring

Processor activity can be monitored using the **iostat**, **mpstat**, **sar** commands in the **sysstat** utility. For a brief description of **sysstat** utility users can refer to “Sysstat utility” on page 507. Besides the **sysstat** utility, users might use the **top** command to monitor the dynamic real time view of a running system. Linux for Power include the **top** utility in the default install. **nmon** is another widely used monitoring tool for Unix systems. For more information on **nmon** refer to “nmon utility” on page 503.

The **mpstat** command reports processor activities for each available processor. The first processor is represented as processor 0. An example of the output is shown below in Figure 14-16. The figure also shows that the **mpstat** command can take monitoring interval (in secs) and count can be used for dynamic monitoring.

```

9.3.5.117 - PuTTY
linux-SLES:/tmp/sysstat-7.0.4 # mpstat

17:50:51 CPU %user %nice %sys %iowait %irq %soft %steal %idle
      intr/s
17:50:51 all 0.06 0.00 0.03 2.05 0.00 0.00 0.01 97.85
      5.91
linux-SLES:/tmp/sysstat-7.0.4 # mpstat 5 2

17:51:29 CPU %user %nice %sys %iowait %irq %soft %steal %idle
      intr/s
17:51:34 all 0.00 0.00 0.00 0.00 0.00 0.00 0.00 100.00
      4.25
17:51:39 all 0.00 0.00 0.00 0.00 0.00 0.00 0.00 100.00
      3.34
Average: all 0.00 0.00 0.00 0.00 0.00 0.00 0.00 100.00
      3.80
linux-SLES:/tmp/sysstat-7.0.4 # █

```

Figure 14-16 *mpstat* command output

The `iostat` command takes the `-c` flag to output only the CPU activity report. The usage of interval and count parameters are the same like other `sysstat` utility components. An example of the output of the CPU activity output is shown below in Example 14-17.

Example 14-17 *Usage of iostat for CPU monitoring*

```

linux-SLES:/tmp/sysstat-7.0.4 # iostat -c 5 2

avg-cpu:  %user   %nice %system %iowait  %steal   %idle
           0.05    0.00    0.03    1.75    0.01   98.16

avg-cpu:  %user   %nice %system %iowait  %steal   %idle
           0.00    0.00    0.00    0.00    0.00  100.00

linux-SLES:/tmp/sysstat-7.0.4 #

```





# Memory monitoring

This chapter covers monitoring memory resources across partitions in a virtualized IBM Power Systems server environment.

First show cross-partition memory monitoring from the Virtual I/O Server for Virtual I/O Server and AIX partitions is discussed, and then examples for client partition memory monitoring for IBM i, and Linux partitions are given.

## 15.1 Cross-partition memory monitoring

Monitoring memory in a virtualized environment on a Virtual I/O Server or an AIX partition is quite simple. The amount of memory allocated to the partition can only vary between the *minimum* and *maximum* values entered in the active profile upon a user request.

Standard memory monitoring tools such as **svmon**, **vmstat**, **topas**, **schedtune**, **vm tune**, and so on, monitor memory almost in the same manner as in standard environments.

One noticeable difference is that tools such as **topas** are aware of dynamic reconfiguration events. Any change to the amount of memory currently allocated to the partition is dynamically reflected in the monitoring tool.

Another interesting feature is the capacity for **topas** to monitor the memory consumption of the other partitions.

**Note:** This capacity for **topas** to monitor the consumption of other partitions is only available when the *Allow performance information collection* item is selected for the current partition properties.

To do so, run the **topas -C** command on the partition as illustrated in Example 15-1.

*Example 15-1 Cross-partition memory monitoring with topas -C*

```

Topas CEC Monitor                Interval: 10                Wed Nov 28 14:09:07 2007
Partitions Memory (GB)          Processors
Shr: 0   Mon:12.5 InUse: 6.1   Shr:1.7 PSz: 4   Don: 1.0 Shr_PhysB 0.57
Ded: 4   Avl: -                Ded: 1 APP: 3.4 Stl: 0.0 Ded_PhysB 0.00
pool  psize  entc  maxc  physb  app  mem  muse
-----
0     4.0    230.0 400.0 0.0    0.00 1.0  1.0
1     2.0    140.0 200.0 0.6    1.52 7.5  3.9

Host      OS  M Mem InU Lp  Us Sy Wa Id  PhysB  Vcsw  %istl %bstl %bdon %idon
-----dedicated-----
Apps_server A61 D 4.0 1.2 2  0  0  0 99  0.00 231  0.00 0.00 0.00 99.66

```

In this report, you see the global memory consumption metrics:

- ▶ *Mon* represents the total amount of memory allocated to the partitions.
- ▶ *mem* represents the sum of monitored memory for all shared partitions in the pool.



- ▶ *muse* represents the sum of memory consumed for all shared partitions in the pool.
- ▶ *Mem* represents the amount of memory allocated to the partition.
- ▶ *InU* represents the amount of memory consumed by the partition.

You can compare these values to the global memory allocations as described in Chapter 12, “Monitoring global system resource allocations” on page 389 and dynamically change the memory allocations if necessary.

**Note:** At the time of writing IBM i, and Linux partitions are not cross-monitored by **topas** and consequently the *Mon* value does not include the memory allocated to IBM i, and Linux partitions.

## 15.2 IBM i memory monitoring

Since IBM i client memory allocations are not visible for cross-partition monitoring tools like *topas* some examples for IBM i memory monitoring using native IBM i tools are given.

### Real-time memory monitoring on IBM i

For nearly real-time memory monitoring on IBM i the **WRKSYSSTS** command can be used shown in Figure 15-1.

```

Work with System Status                                E101F170
                                                    10/27/08 15:56:35
% CPU used . . . . . :      55.6  Auxiliary storage:
% DB capability . . . . :      50.2  System ASP . . . . . :    38.17 G
Elapsed time . . . . . : 00:00:09  % system ASP used . . :    86.5060
Jobs in system . . . . . :    2858  Total . . . . . . . :    38.17 G
% perm addresses . . . . :      .007  Current unprotect used :    4250 M
% temp addresses . . . . :      .010  Maximum unprotect . . :    4250 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     410.54   206.05   +++++   .0     .0     .1     .1
  2    7248.22    11.38    350     .0    5417    .0     .0
  3     403.10    <.01     101     .0     .0     .0     .0
  4         .25     .00       5       .0     .0     .0     .0

                                                    Bottom

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

```

Figure 15-1 IBM i WRKSYSSTS command output

The current memory usage is shown for each IBM i storage pool. Selecting **F5=Refresh** updates the display with current information and **F10=Restart** restarts the statistics from the last displayed time.

**Note:** When using IBM i automatic performance adjustment which is enabled by default via the system value setting QPFRADJ=2, the memory distribution and activity level for the pools are managed dynamically by the system depending on the workload.

As a performance rule of thumb the total page faults, i.e. database and non-database faults, within an IBM i storage pool should be below the following value at least in average for non-peak workloads, otherwise the partition may suffer a memory shortage:

$100 \times (\% \text{ CPU used} / 100) \times \text{number of physical processors in this partition}$

For further information about IBM i performance management tools refer to *IBM eServer iSeries Performance Management Tools*, REDP-4026.

## Long-term memory monitoring on IBM i

For long-term monitoring of storage pools the *IBM Performance Tools for i5/OS* licensed program (5761-PT1) can be utilized which allows to generate a variety of reports from QAPM\* performance database files created from Collection Services data. IBM Performance Tools for i5/OS functions are accessible on IBM i 5250 sessions via a menu using the **STRPFRT** or **GO PERFORM** command and via native CL commands like **STRPFCOL**, **ENDPFCOL**, **CRTPFRDTA**, **PRTCPTRPT** etc. Example 15-2 shows a spool file output from a *component report for storage pool activity* we created via the following command:

```
PRTCPTRPT MBR(Q302174235) TYPE(*POOL)
```

### Example 15-2 IBM i Component Report for Storage Pool Activity

Component Report		10/28/08 18:05:1 Storage Pool Activity										
Page												
Member . . . .	Q302174235	Model/Serial . .	MMA/10-1F170	Main storage . . .	8192.0 MB	Started . . . . .	10/28/08					
17:42:3												
Library . . . .	QPFRDATA	System name . .	:E101F170	Version/Release . .	6/ 1.0	Stopped . . . . .	10/28/08					
18:04:0												
Partition ID . .	005	Feature Code . .	:5622-5622	Int Threshold . . .	.00 %							
Virtual Processors:	2	Processor Units . .	1.00									
<b>Pool identifier . . . .</b>	<b>02</b>	Expert Cache . . . .	0									
Pool	Avg		----- Avg Per Second -----		----- Avg Per							
Minute ----												
Itv	Size	Act	Total	Rsp	CPU	----- DB -----	----- Non-DB -----	-----	Act-	Wait-		
Act-	(MB)	Level	Tns	Time	Util	Faults	Pages	Faults	Pages	Wait	Incl	
End												
Incl												
-----												
17:43	6,823	350	0	.00	15.0	.0	3877	10.2	29	23950	0	
17:44	6,873	350	0	.00	43.7	.3	6598	3.7	6	23795	0	
17:45	6,920	350	0	.00	43.2	.0	5593	.3	0	24031	0	
17:46	6,922	350	0	.00	43.0	.0	6625	.6	1	26356	0	
17:47	6,922	350	0	.00	42.6	.0	5975	.2	0	26960	0	
17:48	6,966	350	0	.00	43.3	.2	5757	20.9	29	21712	0	
17:49	7,009	350	0	.00	45.2	.1	5914	1.0	3	11170	0	
17:50	7,045	350	0	.00	44.3	.0	6556	.0	0	11130	0	
17:51	7,025	350	0	.00	45.0	.0	6322	.0	0	11187	0	
17:52	7,065	350	0	.00	91.9	.0	2794	450.4	1701	32681	0	
17:53	7,103	350	0	.00	69.9	.3	3649	155.8	1262	20327	0	
17:54	7,134	350	0	.00	44.5	.1	5200	105.1	423	16498	0	
17:55	7,160	350	0	.00	96.8	.0	2888	419.6	2431	29363	0	
17:56	7,186	350	0	.00	64.9	.2	4119	14.2	163	22461	0	
17:57	7,210	350	0	.00	49.6	.0	6082	4.5	11	16891	0	
17:58	7,210	350	0	.00	43.1	.0	5757	.3	0	12972	0	
17:59	7,240	350	0	.00	46.3	.1	7021	.5	1	13448	0	
18:00	7,262	350	0	.00	56.3	.1	5228	3.4	6	21026	0	
More...												

Beginning with the interval end at 17:52 we can see a sudden increase in non-database page faults which is related to a sudden increase in threads we generated by restarting the QHTTSPVR subsystem.

For further information about IBM Performance Tools for i5/OS refer to *IBM eServer iSeries Performance Tools for iSeries*, SC41-5340.

The *IBM Systems Director Navigator for i5/OS* graphical user interface which is accessible via [http://IBM\\_i\\_server\\_IP\\_address:2001](http://IBM_i_server_IP_address:2001) can be utilized to graphically display long-term monitoring data like storage pool page faults. For generating the page fault overview graph shown in Figure 15-2 we selected **i5/OS Management** → **Performance** → **Collections**, chose a collection services DB file, selected **Investigate Data** from the popup menu and selected **Page Faults Overview** from the Select Action menu of the Resource Utilization Rates diagram.

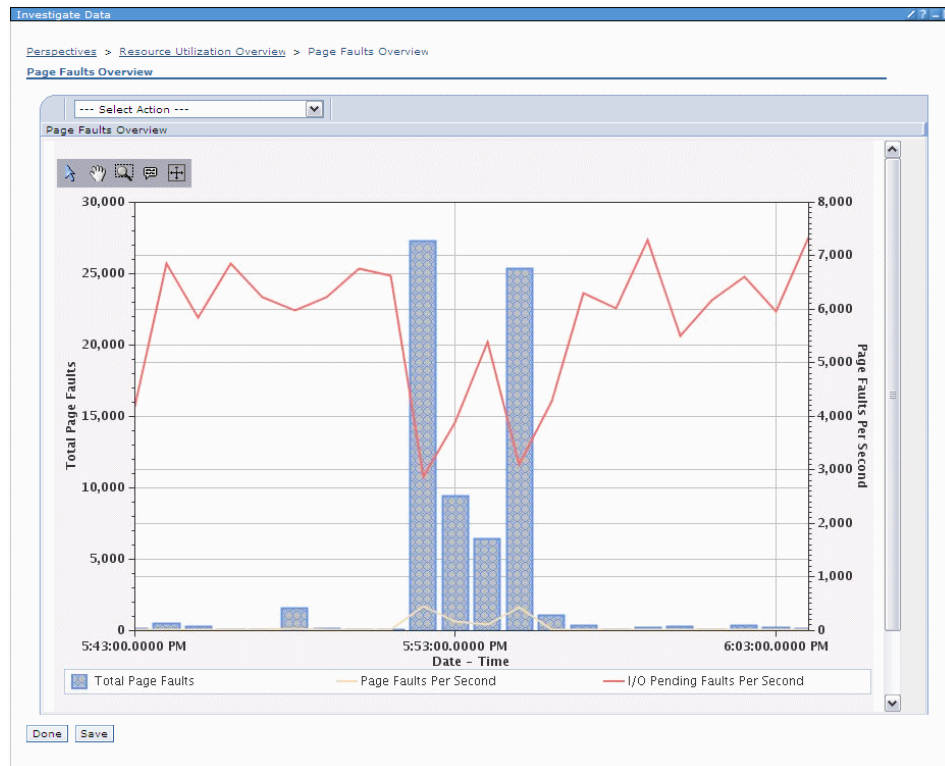


Figure 15-2 IBM i System Director Navigator Page fault overview

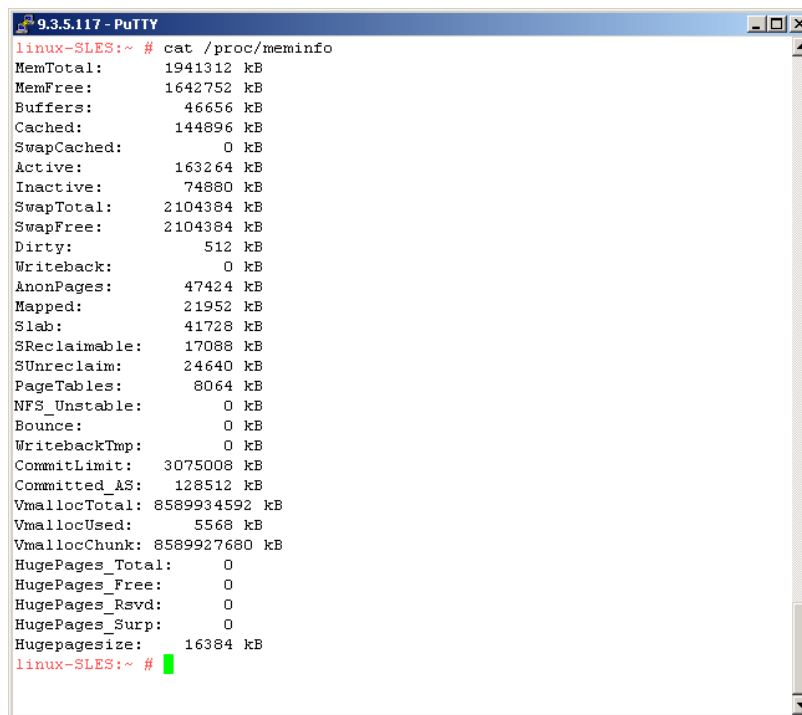
Another approach for long-term memory monitoring for IBM i which also allows IBM i cross-partition monitoring is using the System i Navigator's Management

Central System monitors function. Based on the experienced average page fault rate the user may define a threshold to be notified about an unusual high amount page faults.

For further information about using System i Navigator for performance monitoring refer to *Managing OS/400 with Operations Navigator V5R1 Volume 5: Performance Management*, SG24-6565.

## 15.3 Linux for Power memory monitoring

Red Hat and Novell Linux distributors include the `sysstat` utility in their default installation package. The `sysstat` utility is a collection of performance monitoring tools for Linux. These tools include `sar`, `sadf`, `mpstat`, `iostat`, `pidstat` and `sa` commands which help the user to monitor a number of metrics on the system resource utilization. These tools are discussed in Chapter 18., “Third-party monitoring tools for AIX and Linux” on page 503. Besides the `iostat` command, users can also use the `cat /proc/meminfo` command to get the statistics on memory, virtual memory, hugepages, swap, paging, and page faults. An example output of the above command is show in Figure 15-3.



```
9.3.5.117 - PuTTY
linux-SLES:~ # cat /proc/meminfo
MemTotal:      1941312 kB
MemFree:      1642752 kB
Buffers:       46656 kB
Cached:       144896 kB
SwapCached:    0 kB
Active:       163264 kB
Inactive:     74880 kB
SwapTotal:    2104384 kB
SwapFree:     2104384 kB
Dirty:        512 kB
Writeback:    0 kB
AnonPages:   47424 kB
Mapped:      21952 kB
Slab:        41728 kB
SReclaimable: 17088 kB
SUnreclaim: 24640 kB
PageTables:   8064 kB
NFS_Unstable: 0 kB
Bounce:      0 kB
WritebackTmp: 0 kB
CommitLimit: 3075008 kB
Committed_AS: 128512 kB
VmallocTotal: 8589934592 kB
VmallocUsed:  5568 kB
VmallocChunk: 8589927680 kB
HugePages_Total: 0
HugePages_Free: 0
HugePages_Rsvd: 0
HugePages_Surp: 0
Hugepagesize: 16384 kB
linux-SLES:~ #
```

Figure 15-3 Linux monitoring memory statistics using `meminfo`



# Virtual storage monitoring

This chapter describes how to check and monitor virtual storage health and performance for Virtual I/O Server and virtual I/O client partitions.

## 16.1 Virtual I/O Server storage monitoring

In this section we provide some guidance for how to check the storage health and performance on the Virtual I/O Server.

### 16.1.1 Checking storage health on the Virtual I/O Server

On the Virtual I/O use the following commands to check its disk status:

<b>lsvg rootvg</b>	Check for stale PPs and no stale PV.
<b>lsvg -pv rootvg</b>	Check for missing disks.
<b>errlog</b>	Check for disk related errors in the error log
<b>lsvg -p rootvg</b>	On the AIX virtual I/O client, check for missing disks.

If using IBM SAN storage with the Virtual I/O Server:

<b>lspath</b>	Check for missing paths.
<b>mpio_get_config -Av</b>	Check all SAN storage, LUNs are detected.
<b>lsdev -type disk</b>	Check all expected disks are available.

### 16.1.2 Monitoring storage performance on the Virtual I/O Server

The **viostat** command can be very helpful in tracing system activity with regard to questions related to I/O. It allows for relatively fine-grained measurements of different types of adapters and attached disks as well as the usage of paths to redundant attached disks, including virtual adapters and virtual disks as well as their backing devices.

The output of a measurement while disk I/O occurred on one client with a virtual disk, is shown in Example 16-1.

#### *Example 16-1 Monitoring I/O performance with viostat*

```
$ viostat -extdisk
System configuration: lcpu=2 drives=15 paths=22 vdisks=23

hdisk8      xfer: %tm_act    bps    tps    bread    bwrtn
           94.7    9.0M   65.9   14.7    9.0M
           read:    rps  avgserv  minserv  maxserv  timeouts  fails
           0.0    1.6    0.1    5.0     0         0
           write:  wps  avgserv  minserv  maxserv  timeouts  fails
           65.8   27.0   0.2    3.8S   0         0
           queue: avgtime  mintime  maxtime  avgqsz  avgqsz  sqfull
           0.0    0.0    1.8S   0.0    0.9    0.0
hdisk11     xfer: %tm_act    bps    tps    bread    bwrtn
```



		94.8	9.0M	64.7	14.7	9.0M	
	read:	rps	avgserv	minserv	maxserv	timeouts	fails
		0.0	2.0	0.1	10.6	0	0
	write:	wps	avgserv	minserv	maxserv	timeouts	fails
		64.7	27.7	0.2	3.5S	0	0
	queue:	avgtime	mintime	maxtime	avgqsz	avgqsz	sqfull
		0.0	0.0	263.5	0.0	0.9	0.0
hdisk9	xfer:	%tm_act	bps	tps	bread	bwrtn	
		0.2	0.0	0.0	0.0	0.0	
	read:	rps	avgserv	minserv	maxserv	timeouts	fails
		0.0	0.0	0.0	0.0	0	0
	write:	wps	avgserv	minserv	maxserv	timeouts	fails
		0.0	0.0	0.0	0.0	0	0
	queue:	avgtime	mintime	maxtime	avgqsz	avgqsz	sqfull
		0.0	0.0	0.0	0.0	0.0	0.0
hdisk6	xfer:	%tm_act	bps	tps	bread	bwrtn	
		2.6	134.6K	8.9	118.5K	16.1K	
	read:	rps	avgserv	minserv	maxserv	timeouts	fails
		6.5	6.6	0.1	1.9S	0	0
	write:	wps	avgserv	minserv	maxserv	timeouts	fails
		2.3	0.9	0.2	220.5	0	0
	queue:	avgtime	mintime	maxtime	avgqsz	avgqsz	sqfull
		0.2	0.0	44.2	0.0	0.0	0.7

---

## 16.2 AIX virtual I/O client storage monitoring

In this section we provide some guidance for how to check the storage health and performance on the AIX virtual I/O client.

## 16.2.1 Checking storage health on the AIX virtual I/O client

In this section we differentiate between checking the storage health on an AIX virtual I/O client in a MPIO and a LVM mirroring environment.

### AIX virtual I/O client MPIO environment

The following procedure applies to a dual Virtual I/O Server environment using MPIO on the AIX virtual I/O client partition as shown in Figure 16-1.

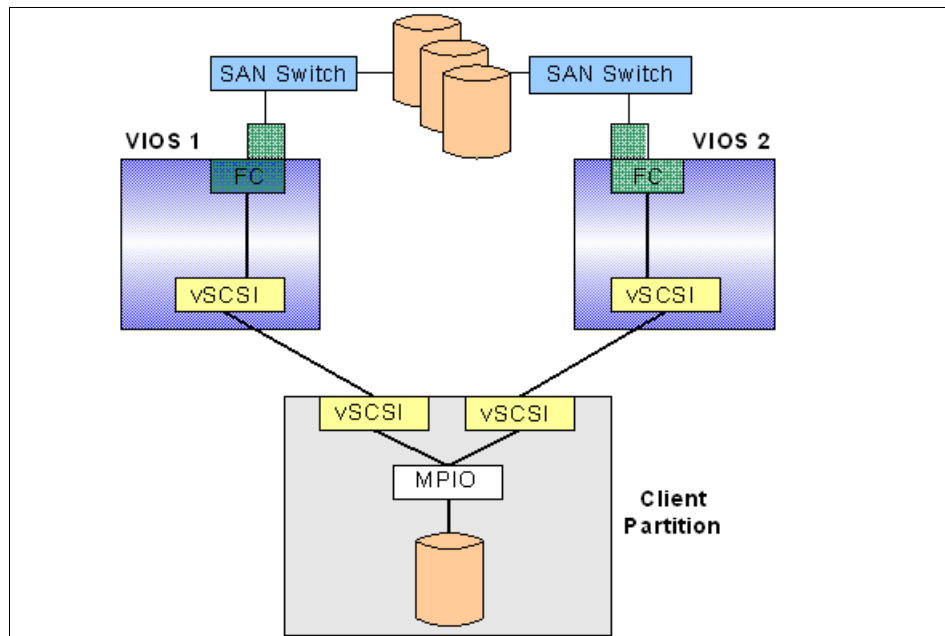


Figure 16-1 AIX virtual I/O client using MPIO

Run the following commands on the AIX virtual I/O client to check its storage health:

**lspath** Check all the paths to the disks. They should all be in the enabled state as shown in Example 16-2.

*Example 16-2 AIX lspath command output*

```
# lspath
Enabled hdisk0 vscsi0
Enabled hdisk0 vscsi1
```

**lsattr -E1 hdisk0** Check the MPIO heartbeat for hdisk0, the attribute hcheck\_mode is set to nonactive, and hcheck\_interval is

60. If you run IBM SAN storage, check that `reserve_policy` is `no_reserve`, other storage vendors might require other values for `reserve_policy`. Example 16-2 shows output of `lsattr` command.

*Example 16-3 AIX client `lsattr` command to show `hdisk` attributes.*

---

```
# lsattr -El hdisk0
PCM                PCM/friend/vscsi          Path Control Module      False
algorithm          fail_over                 Algorithm                 True
hcheck_cmd         test_unit_rdy            Health Check Command     True
hcheck_interval    60                       Health Check Interval    True
hcheck_mode        nonactive                 Health Check Mode        True
max_transfer       0x40000                  Maximum TRANSFER Size    True
pvid               00c1f170e327afa70000000000000000 Physical volume identifier False
queue_depth        3                         Queue DEPTH              True
reserve_policy     no_reserve                Reserve Policy           True
```

---

If the Virtual I/O Server was rebooted earlier and the `health_check` attribute is not set, you may need to enable it. There are instances when the path shows up as failed though the path to the Virtual I/O Server is actually up. The way to correct this is to set the `hcheck_interval` and `hcheck_mode` attributes using the `chdev` command as shown in Example 16-4.

*Example 16-4 Using the `chdev` command for setting `hdisk` recovery parameters*

---

```
# chdev -l hdisk0 -a hcheck_interval=60 -a hcheck_mode=nonactive -P
```

---

`lsvg -p rootvg` Check for a missing `hdisk`.

## AIX virtual I/O client LVM mirroring environment

In the following we discuss storage health monitoring for an AIX virtual I/O client software mirroring environment as shown in Figure 16-2.

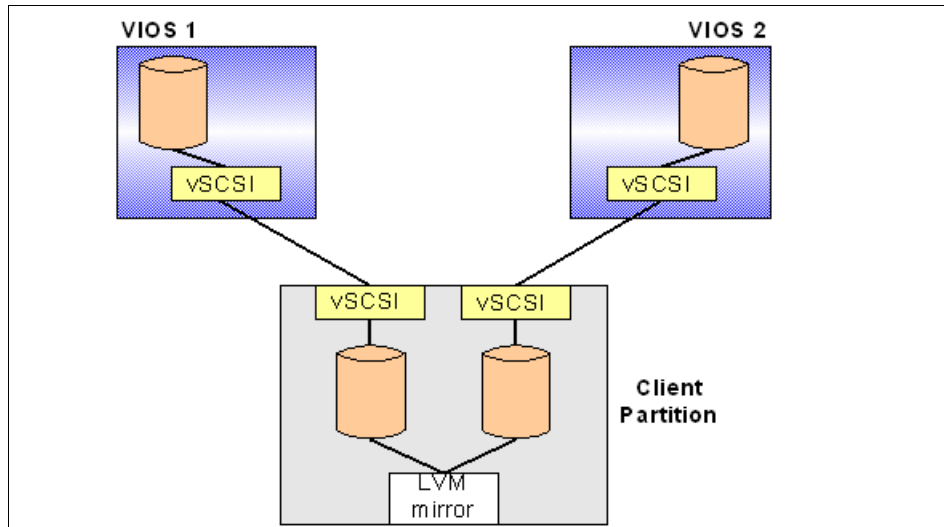


Figure 16-2 AIX virtual I/O client using LVM mirroring

Check for any missing disk using the `lsvg -p rootvg` command as shown in Example 16-3.

*Example 16-5 Check missing disk*

```
# lsvg -p rootvg
rootvg:
PV_NAME          PV STATE          TOTAL PPs   FREE PPs   FREE
DISTRIBUTION
hdisk0           active            511         488
102..94..88..102..102
hdisk1           missing         511         488
102..94..88..102..102
```

If one disk is missing there will be stale partitions which can be verified using the `lsvg rootvg` command. If there are stale partitions and all disks are available it is important to resynchronize the mirror using the `varyonvg` command and the `syncvg -v` command on the volume groups that use virtual disks from the Virtual I/O Server environment. Example 16-6 shows output of this commands.

*Example 16-6 AIX command to recover from stale partitions*

```
# varyonvg rootvg
# syncvg -v rootvg
# lsvg -p rootvg
rootvg:
```

```

PV_NAME          PV STATE          TOTAL PPs   FREE PPs   FREE
DISTRIBUTION
hdisk0           active           511         488
102..94..88..102..102
hdisk1           active          511         488
102..94..88..102..102
# lsvg rootvg
VOLUME GROUP:    rootvg           VG IDENTIFIER:
00c478de00004c00000
00006b8b6c15e
VG STATE:        active           PP SIZE:     64 megabyte(s)
VG PERMISSION:   read/write      TOTAL PPs:   1022 (65408
megabytes)
MAX LVs:         256            FREE PPs:    976 (62464
megabytes)
LVs:            9              USED PPs:    46 (2944
megabytes)
OPEN LVs:        8              QUORUM:      1
TOTAL PVs:       2              VG DESCRIPTORS: 3
STALE PVs:    0              STALE PPs:    0
ACTIVE PVs:      2              AUTO ON:     yes
MAX PPs per VG:  32512
MAX PPs per PV:  1016            MAX PVs:     32
LTG size (Dynamic): 256 kilobyte(s)  AUTO SYNC:   no
HOT SPARE:       no              BB POLICY:   relocatable
#

```

For a configuration with a large number of AIX virtual I/O client partitions it would be time consuming and error-prone to check the storage health individually. A sample script using distributed shell for automating the health checking and recovery for a group of AIX virtual I/O client partitions is provided in Appendix A, "Sample script for disk and NIB network checking and recovery on AIX virtual clients" on page 511.

## 16.2.2 Monitoring storage performance on the AIX virtual I/O client

The **iostat** command can be very helpful in tracing system activity with regard to questions related to I/O. It allows for relatively fine-grained measurements of different types of adapters and attached disks as well as the usage of paths to redundant attached disks.

The output of a measurement while disk I/O occurred on hdisk0 disk, is shown in Example 16-9.

*Example 16-7 Monitoring disk performance with iostat*

```
# iostat -d hdisk0 2
```

System configuration: 1cpu=2 drives=2 paths=3 vdisks=3

Disks:	% tm_act	Kbps	tps	Kb_read	Kb_wrtn
hdisk0	0.0	0.0	0.0	0	0
hdisk0	0.0	0.0	0.0	0	0
hdisk0	85.2	538.8	589.5	614	512
hdisk0	97.4	744.9	709.2	692	768
hdisk0	90.7	673.2	672.2	693	724
hdisk0	92.9	723.1	704.6	718	768
hdisk0	100.0	654.5	674.0	669	640
hdisk0	100.0	669.5	704.0	699	640

## 16.3 IBM i virtual I/O client storage monitoring

In this section we describe monitoring the storage health and performance from an IBM i virtual I/O client.

### 16.3.1 Checking storage health on the IBM i virtual I/O client

There are basically two things to check on the IBM i virtual I/O client regarding storage health which is verifying that the used capacity doesn't reach the auxiliary storage pool (ASP) limit and that all disk units are available.

To check the used storage capacity in the system ASP (ASP 1) use the **WRKSYSSTS** command check the value displayed for *% system ASP used*, for user ASPs the used capacity can be displayed from System Service Tools (SST) by entering the command **STRSST** and choosing the options **3. Work with disk units** → **1. Display disk configuration** → **2. Display disk configuration capacity**.

By default the ASP usage threshold is 90% which if reached causes a system operator message CPF0907 notification. User ASPs reaching 100% used capacity will by default overflow into the system ASP while independent ASPs (IASPs) running out of capacity might get varied off. A system ASP becoming filled up would likely cause a system crash so it is important to monitor the IBM i storage usage and take preventive measures like data housekeeping or adding additional storage.

To check if all disk units are available run the command **STRSST** to log into System Service Tools and select **3. Work with disk units** → **1. Display disk configuration** → **1. Display disk configuration status**. Check if all disk units are shown with a status of either *Configured* for a single Virtual I/O Server

environment or *Active* for dual Virtual I/O Server environment with using IBM i mirroring across two Virtual I/O Servers as shown in Figure 16-3.

If there are units shown as suspended, not connected or missing from the configuration resolve the problem, typically checking first if the corresponding volumes are available and accessible on the Virtual I/O server – refer to “IBM i Virtual SCSI disk configuration tracing” on page 36, and check whether IBM i mirroring state changes to resuming – if not, resume the units manually using the SST function **3. Work with disk units** → **3. Work with disk unit recovery** → **4. Resume mirrored protection**.

```

Display Disk Configuration Status

      Serial
ASP Unit Number      Type Model Name      Status
  1
    1 YYUUH3U9UELD    6B22 050 DD004    Active
    1 YD598QUY5XR8    6B22 050 DD003    Active
    2 YTM3C79KY4XF    6B22 050 DD002    Active
    2 Y3WUTVVQMM4G    6B22 050 DD001    Active

Press Enter to continue.

F3=Exit      F5=Refresh      F9=Display disk unit details
F11=Disk configuration capacity  F12=Cancel

```

Figure 16-3 IBM i Display Disk Configuration Status

## 16.3.2 Monitoring storage performance on the IBM i virtual I/O client

In the following we show some examples for IBM i storage monitoring using native IBM i tools.

### Real-time storage monitoring on IBM i

For nearly real-time storage monitoring on IBM i the **WRKDSKSTS** command can be used shown in Figure 16-4.

Work with Disk Status		E101F170		10/31/08 14:04:33						
Elapsed time:		00:00:01								
Unit	Type	Size (M)	% Used	I/O Rqs	Request Size (K)	Read Rqs	Write Rqs	Read (K)	Write (K)	% Busy
1	6B22	19088	76.0	125.1	5.7	30.2	94.9	4.5	6.1	3
1	6B22	19088	76.0	120.5	7.1	40.0	80.5	4.5	8.3	2
2	6B22	19088	75.5	213.4	5.2	121.0	92.3	4.5	6.2	4
2	6B22	19088	75.5	184.1	5.4	124.6	59.5	4.5	7.4	3

Bottom

Command  
 ===>  
 F3=Exit F5=Refresh F12=Cancel F24=More keys

Figure 16-4 IBM i WRKDSKSTS command output

The current disk I/O workload statistics are shown for each IBM i disk unit. Selecting **F5=Refresh** updates the display with current information and **F10=Restart statistics** restarts the statistics from the last displayed time.

*Collection Services* on IBM i is the core component for system-wide performance data collection at specified intervals (default 15 min.) and is enabled by default. To administer Collection Services the CL commands **CFGPFRCOL**, **STRPFRCOL** and **ENDPFRCOL** can be used or the menu interface of IBM Performance Tools for i5/OS licensed program. To get information about IBM i disk performance statistics collection services data from the QAPMDISK database file would need to be analyzed. This could be done using either native SQL commands or more easily by using IBM Performance Tools for i5/OS as described in the following section.

For further information about IBM i performance management tools refer to *IBM eServer iSeries Performance Management Tools*, REDP-4026.





16:05	1,005.7	886.6	119.0	8.5	59.7	65.4	0002B	.0047	0002A
66.327									
16:10	769.1	510.6	258.4	16.8	71.9	77.4	0002B	.0115	0002A
66.018									
16:15	656.1	630.2	25.9	7.3	51.8	53.0	0002B	.0042	0002A
66.634									
16:20	560.3	479.7	80.5	11.0	55.9	61.1	0002B	.0065	0002A
66.912									
16:25	691.0	469.7	221.3	15.8	70.0	74.7	0002B	.0118	0001A
66.110									
16:30	722.5	679.8	42.7	9.1	52.9	54.7	0002B	.0042	0002A
66.954									
16:35	728.0	649.6	78.4	9.4	55.4	60.6	0002B	.0050	0002A
66.905									
16:40	666.1	478.2	187.8	20.7	63.2	70.2	0002B	.0140	0002A
66.940									
16:45	1,032.1	972.6	59.5	6.4	53.0	55.8	0002B	.0030	0002A
66.517									
16:50	765.3	596.8	168.5	16.1	61.5	69.4	0002B	.0105	0001A
65.983									
More...									
-----	-----	-----	-----	-----					
Average:	795.4	677.9	117.4	12.1	59.0				

For monitoring IBM i storage performance it is useful to look at a system report and resource report for the same time frame together as the system report provided useful information about the average disk response time (= service time + wait time) and the resource report shows the amount of I/O workload for the time frame. If the resource report shows changing workload characteristics in terms of I/O throughput or I/O transfer size like small I/O transfers typical for interactive workload during the day and large I/O transfers typical for batch or save workload during the night it might be useful to generate a separate system report for each workload period to get a better view of the disk performance for the different workloads.

For further information about IBM Performance Tools for i5/OS refer to *IBM eServer iSeries Performance Tools for iSeries*, SC41-5340.

The *IBM Systems Director Navigator for i5/OS* graphical user interface which is accessible via `http://IBM_i_server_IP_address:2001` can be utilized to graphically display long-term monitoring data for disk performance. For generating the page fault overview graph shown in Figure 16-5 we first selected **i5/OS Management** → **Performance** → **Collections**, chose a collection services DB file, selected **Investigate Data** from the popup menu and selected **Disk Overview for System Disk Pool** from the Select Action menu of the diagram. We still added information for the I/O workload by having used the option **Add data series** from the Select Action menu as disk performance should be looked at in context of the I/O workload.

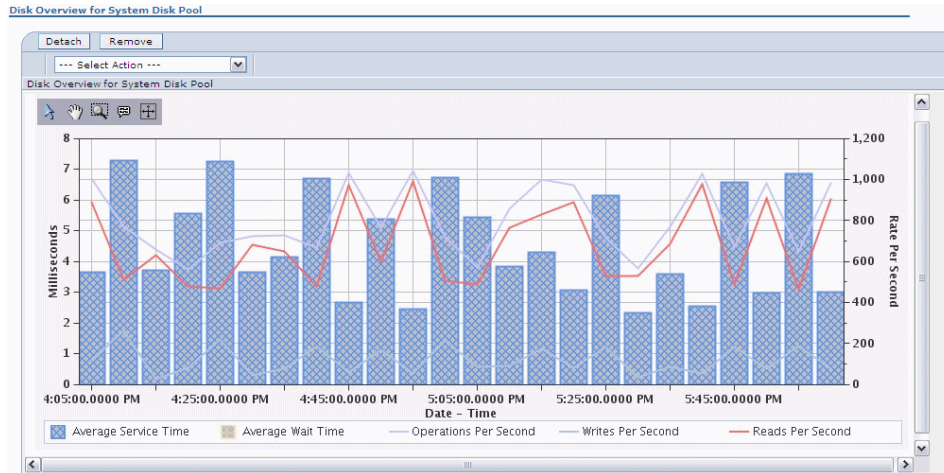


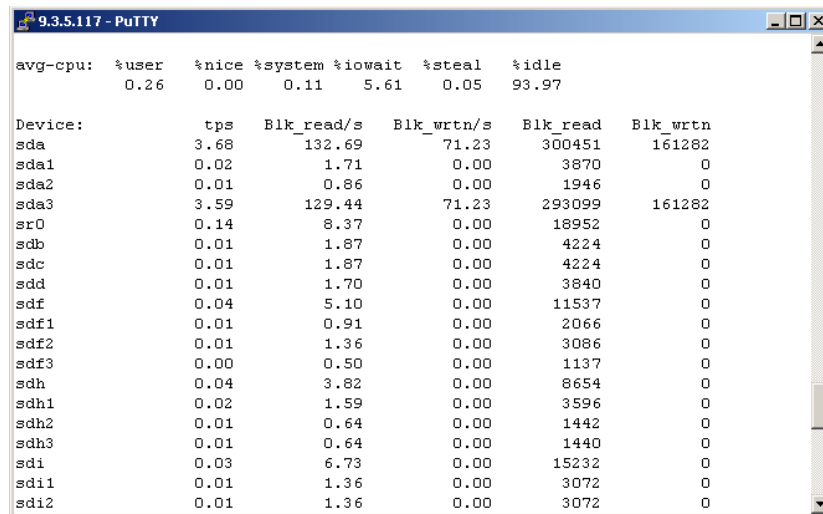
Figure 16-5 IBM i Navigator Disk Overview for System Disk Pool

Another approach for long-term disk monitoring for IBM i which also allows IBM i cross-partition monitoring is using the System i Navigator's Management Central monitors function. There is no metric for disk response time however based on the experienced average disk arm utilization the user may define a threshold to be notified about an unusual high disk arm utilization to be alerted for potential disk performance problems.

For further information about using System i Navigator for performance monitoring refer to *Managing OS/400 with Operations Navigator V5R1 Volume 5: Performance Management*, SG24-6565.

## 16.4 Linux for Power virtual I/O client storage monitoring

I/O activity across internal and external disks can be monitored using **iostat** command included in the **sysstat** package in Linux distributions. A brief description of the **sysstat** utility can be found in 18.2, “Sysstat utility” on page 507. Users can obtain I/O transfer rate and other I/O statistics with the **iostat** command. An example output is shown below in Figure 16-6



```

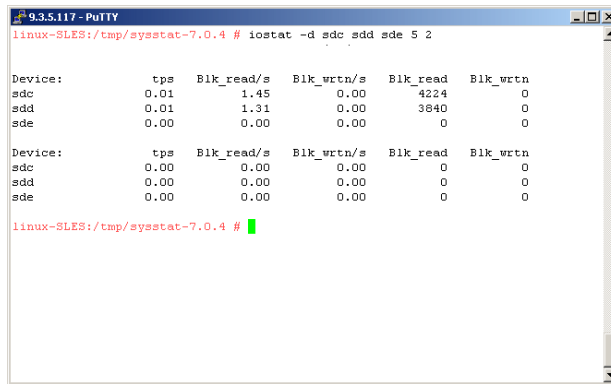
9.3.5.117 - PuTTY
avg-cpu:  %user   %nice %system %iowait  %steal   %idle
           0.26    0.00    0.11    5.61    0.05   93.97

Device:            tps    Blk_read/s    Blk_wrtn/s    Blk_read    Blk_wrtn
sda                 3.68         132.69         71.23      300451      161282
sda1                0.02          1.71          0.00         3870         0
sda2                0.01          0.86          0.00         1946         0
sda3                3.59         129.44         71.23     293099      161282
sr0                 0.14          8.37          0.00        18952         0
sdb                 0.01          1.87          0.00         4224         0
sdc                 0.01          1.87          0.00         4224         0
sdd                 0.01          1.70          0.00         3840         0
sdf                 0.04          5.10          0.00        11537         0
sdf1                0.01          0.91          0.00         2066         0
sdf2                0.01          1.36          0.00         3086         0
sdf3                0.00          0.50          0.00         1137         0
sdh                 0.04          3.82          0.00         8654         0
sdh1                0.02          1.59          0.00         3596         0
sdh2                0.01          0.64          0.00         1442         0
sdh3                0.01          0.64          0.00         1440         0
sdi                 0.03          6.73          0.00        15232         0
sdi1                0.01          1.36          0.00         3072         0
sdi2                0.01          1.36          0.00         3072         0

```

Figure 16-6 *iostat* command output showing the i/o output activity

When the user first executes the **iostat** command as shown in Figure 16-6, the utility generates the report of the I/O statistics on all disk devices since the boot time. Each subsequent execution of the command will only report the activity since the previous execution of the command/report. Users can also use the **-d** flag to monitor the I/O activity of specific disks as shown in Figure 16-7:



```
9.3.5.117 - PuTTY
linux-SLES:/tmp/sysstat-7.0.4 # iostat -d sdc sdd sde 5 2

Device:            tps    Blk_read/s    Blk_wrtn/s    Blk_read    Blk_wrtn
sdc                 0.01         1.45           0.00         4224         0
sdd                 0.01         1.31           0.00         3840         0
sde                 0.00         0.00           0.00          0            0

Device:            tps    Blk_read/s    Blk_wrtn/s    Blk_read    Blk_wrtn
sdc                 0.00         0.00           0.00          0            0
sdd                 0.00         0.00           0.00          0            0
sde                 0.00         0.00           0.00          0            0

linux-SLES:/tmp/sysstat-7.0.4 #
```

Figure 16-7 *iostat* output with *-d* flag and 5 sec interval as a parameter.





## Virtual network monitoring

Once a virtualized environment is set up, it is important to check whether all the contingencies in place for network connectivity will work in case of a failure. In this chapter we discuss:

- ▶ “Monitoring the Virtual I/O Server”
  - “Error logs”
  - “IBM Tivoli Monitoring”
  - “Testing your configuration”
- ▶ “Virtual I/O Server networking monitoring”
- ▶ “AIX client network monitoring”
- ▶ “IBM i client network monitoring”

## 17.1 Monitoring the Virtual I/O Server

You can monitor the Virtual I/O Server using error logs or IBM Tivoli Monitoring.

### 17.1.1 Error logs

AIX, IBM i, and Linux client logical partitions log errors against failing I/O operations. Hardware errors on the client logical partitions associated with virtual devices usually have corresponding errors logged on the Virtual I/O server. The error log on the Virtual I/O Server is displayed using the `errlog` command. However, if the failure is within the client partition, there are typically no errors logged on the Virtual I/O Server. Also, on Linux client logical partitions, if the algorithm for retrying SCSI temporary errors is different from the algorithm used by AIX, the errors might not be recorded on the server.

### 17.1.2 IBM Tivoli Monitoring

Beginning with Virtual I/O Server V1.3.0.1 (fix pack 8.1), you can install and configure the IBM Tivoli Monitoring System Edition for the System p agent on the Virtual I/O Server. Tivoli Monitoring System Edition for System p enables you to monitor the health and availability of multiple IBM System p servers (including the Virtual I/O Server) from the Tivoli Enterprise Portal. It gathers following network related data from Virtual I/O Server:

- ▶ Network adapter details
- ▶ Network adapter utilization
- ▶ Network interfaces
- ▶ Network protocol views
- ▶ Shared Ethernet
- ▶ Shared Ethernet adapter high availability details
- ▶ Shared Ethernet bridging details

You can monitor the above mentioned metrics from Tivoli Enterprise Portal Client. For more information on Network monitoring using Tivoli Enterprise Portal refer to “Networking” on page 379 of Chapter “Virtual I/O Server monitoring agents”.



### 17.1.3 Testing your configuration

One thing that can be done to prevent errors is to test your environment configuration periodically. The tests can be performed on the network configuration as well as on the virtual SCSI configuration.

#### Testing the network configuration

In this section the steps to test a Shared Ethernet Adapter Failover and a Network Interface Backup are shown. It is assumed that the SEA and NIB are already configured in an environment with two Virtual I/O Servers (Virtual I/O Server 1 and Virtual I/O Server 2). For more information about how to configure this environment, refer to *PowerVM on IBM System p: Introduction and Configuration*, SG24-7940.

#### Testing the SEA Failover

Perform the following steps to test whether the SEA Failover configuration works as expected.

1. Open a remote session from a system on the external network to any of the AIX, IBM i or Linux client partitions. If your session gets disconnected during any of the tests, your configuration is not highly available. You might want to run a `ping` command to verify that you have continuous connectivity.
2. On the Virtual I/O Server 1 check whether the primary adapter is active using the `entstat` command. In this example the primary adapter is ent4.

```
$ entstat -all ent4 | grep Active
Priority: 1 Active: True
```

3. Perform a manual failover using the `chdev` command to switch to the standby adapter:

```
chdev -dev ent4 -attr ha_mode=standby
```

4. Check whether the SEA failover was successful using the `entstat` command. On Virtual I/O Server 1 the `entstat` output should look like this:

```
$ entstat -all ent4 | grep Active
Priority: 1 Active: False
```

You should also see the following entry in the errorlog when you issue the `errlog` command:

```
40D97644 1205135007 I H ent4 BECOME BACKUP
```

On Virtual I/O Server 2, the `entstat` command output should look like this:

```
$ entstat -all ent4 | grep Active
Priority: 2 Active: True
```

You should see the following entry in the errorlog when you issue the **errlog** command.

```
E136EAFA 1205135007 I H ent4          BECOME PRIMARY
```

**Note:** You may experience up to 30 seconds delay in failover when using SEA failover. The behavior depends on the network switch and the spanning tree settings.

5. On Virtual I/O Server 1, switch back to the primary adapter and verify that the primary adapter is active using these commands:

```
$ chdev -dev ent4 -attr ha_mode=auto
ent4 changed
$ entstat -all ent4 | grep Active
Priority: 1 Active: True
```

6. Unplug the link of the physical adapter on Virtual I/O Server 1. Use the **entstat** command to check whether the SEA has failed over to the standby adapter.

Re-plug the link of the physical adapter on Virtual I/O Server 1 and verify that the SEA has switched back to the primary.

### **Testing NIB**

Perform the following steps to verify that the NIB (network interface backup or Etherchannel) configuration works as expected for an AIX virtual I/O client partition.

**Note:** NIB or Etherchannel is not available on IBM i. A similar concept on IBM i using virtual IP address (VIPA) failover is currently not supported for use with Virtual Ethernet adapters.

These steps show how failover works when the network connection of a Virtual I/O Server is disconnected. The failover works in the same fashion when a Virtual I/O Server is rebooted. You can test this by rebooting Virtual I/O Server 1 instead of unplugging the network cable in step 2.

1. Do a remote login to the client partition using telnet or SSH. Once you are logged in, check whether the primary channel that is connected to Virtual I/O Server 1 is active using the **entstat** command, as shown in Example 17-1.

*Example 17-1 Verifying the active channel in an EtherChannel*

---

```
# entstat -d ent2 | grep Active
Active channel: primary channel
```

---

2. Unplug the network cable from the physical network adapter that is connected to Virtual I/O Server 1.
3. As soon as the EtherChannel notices that it has lost connection it should perform a switchover to the backup adapter. You will see a message as shown in Example 17-2 in the errorlog.

**Important:** Your telnet or SSH connection should not be disconnected. If it is disconnected your configuration fails High Availability criterion.

*Example 17-2 Errorlog message when the primary channel fails*

---

LABEL: ECH\_PING\_FAIL\_PRMRY  
IDENTIFIER: 9F7B0FA6

Date/Time: Fri Oct 17 19:53:35 CST 2008  
Sequence Number: 141  
Machine Id: 00C1F1704C00  
Node Id: NIM\_server  
Class: H  
Type: INFO  
WPAR: Global  
Resource Name: ent2  
Resource Class: adapter  
Resource Type: ibm\_ech  
Location:

Description  
PING TO REMOTE HOST FAILED

Probable Causes  
CABLE  
SWITCH  
ADAPTER

Failure Causes  
CABLES AND CONNECTIONS

Recommended Actions  
CHECK CABLE AND ITS CONNECTIONS  
IF ERROR PERSISTS, REPLACE ADAPTER CARD.

Detail Data  
FAILING ADAPTER  
PRIMARY

```
SWITCHING TO ADAPTER
```

```
ent1
```

```
Unable to reach remote host through primary adapter: switching over
to backup adapter
```

---

As shown in Example 17-3, the **entstat** command will also show that the backup adapter is now active.

*Example 17-3 Verifying the active channel in an EtherChannel*

---

```
# entstat -d ent2 | grep Active
Active channel: backup adapter
```

---

4. Reconnect the physical adapter in Virtual I/O Server 1.
5. The EtherChannel will not automatically switch back to the primary channel. In the SMIT menu, there is an option to “Automatically Recover to Main Channel”. It is set to Yes by default and this is the behavior when using physical adapters. However, virtual adapters do not adhere to this. Instead, the backup channel is used until it fails and then switches to the primary channel. You have to manually switch back using the **/usr/lib/methods/ethchan\_config -f** command as shown in Example 17-4. A message will appear in the errorlog when the EtherChannel recovers to the primary channel.

*Example 17-4 Manual switch to primary channel using entstat*

---

```
# /usr/lib/methods/ethchan_config -f ent2
# entstat -d ent2 | grep Active
Active channel: primary channel
# errpt
8650BE3F 1123195807 I H ent2          ETHERCHANNEL
RECOVERY
```

---

**Note:** You can use the **dsh** command to automate the check or switch back if you have a lot of client partitions, or use the script in Appendix A, “Sample script for disk and NIB network checking and recovery on AIX virtual clients” on page 511.

### ***Testing the bonding device configuration on Linux***

Perform the following steps to verify that the bonding device configuration on Linux is working as expected. Example 17-5 shows how to check for the link failure count in an interface.

#### *Example 17-5 Checking for the Link Failure count*

---

```
# cat /proc/net/bonding/bond0
Ethernet Channel Bonding Driver: v2.6.3-rh (June 8, 2005)

Bonding Mode: fault-tolerance (active-backup)
Primary Slave: None
Currently Active Slave: eth0
MII Status: up
MII Polling Interval (ms): 0
Up Delay (ms): 0
Down Delay (ms): 0

Slave Interface: eth0
MII Status: up
Link Failure Count: 0
Permanent HW addr: ba:d3:f0:00:40:02

Slave Interface: eth1
MII Status: up
Link Failure Count: 0

Permanent HW addr: ba:d3:f0:00:40:03
```

---

## **17.2 Virtual I/O Server networking monitoring**

In this section we present a monitoring scenario where the intention is to identify which adapter from a Link Aggregation (or EtherChannel) was used to transfer data when a certain amount of data was transferred via ftp from a server through the Virtual I/O Server when link aggregation was used as Shared Ethernet Adapter backing device.

### **17.2.1 Describing the scenario**

The scenario to be analyzed is presented in Figure 17-1. In this scenario there is a Virtual I/O Server, a logical partition, and an external Linux server. A given amount of data will be transferred from the Linux server to the logical partition

through the Virtual I/O Server via ftp, and the interface used to transfer this data on the Virtual I/O Server will be identified.

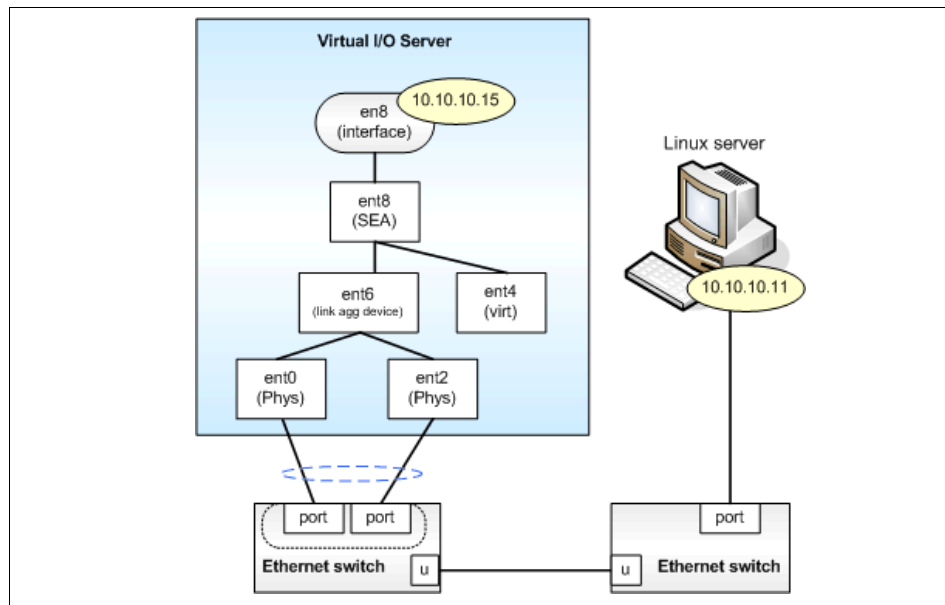


Figure 17-1 Network monitoring testing scenario

To set up the environment the Virtual I/O Server was connected to a switch that supports Link Aggregation. We used a 4-port Ethernet card where port0 and port2 were connected to the switch. These ports were recognized as ent0 and ent2 on the Virtual I/O Server.

The Link Aggregation device (ent6) was created using the following command (or alternatively you can use **smitty etherchannel** from root shell):

```
$ mkvdev -lnagg ent0,ent2 -attr mode=8023ad hash_mode=src_dsc_port
ent6 Available
```

When operating under IEEE 802.3ad mode, as defined by the mode=8023ad flag, the hash\_mode attribute determines how the outgoing adapter for each packet is chosen. In this case the src\_dsc\_port was chosen, which means that both the source and the destination TCP or UDP ports will be used for that connection to determine the outgoing adapter.

The hash\_mode attribute was introduced in IY45289 (devices.common.IBM.ethernet.ret 5.2.0.13). The hash attribute can also be set to default, src\_port and dst\_port. We can use this attribute, for example, to define

that http traffic should go through by one specific adapter while ftp traffic should go through by the other adapter.

You can find more information about SEA attributes at:

[http://publib.boulder.ibm.com/infocenter/systems/scope/hw/topic/iphb1\\_vios\\_managing\\_sea\\_attr.htm](http://publib.boulder.ibm.com/infocenter/systems/scope/hw/topic/iphb1_vios_managing_sea_attr.htm)

Once the Link Aggregation device is created, the Shared Ethernet Adapter can be configured. To create a Shared Ethernet Adapter use the following command:

```
$ mkvdev -sea ent6 -vadapter ent4 -default ent4 -defaultid 1
ent8 Available
```

The next step is to configure the IP address on the SEA adapter. In order to do that, execute the following command:

```
$ mktcpip -hostname 'VI0_Server1' -inetaddr '10.10.10.15' -netmask
'255.0.0.0' -interface 'en8'
```

With these steps done, the Virtual I/O Server is ready to perform file transfer tests.

Before starting the transfer tests, reset all the statistics for all adapters on the Virtual I/O Server. One way to do that is using a simple for loop as follows:

```
$ for i in 0 1 2 3 4 5 6 7 8
> do
> entstat -reset ent$i
> done
```

You can now check the statistics of adapter ent8 for the first time as shown in Example 17-6. All the values should be low since they have just been reset.

*Example 17-6 Output of entstat on SEA*

---

```
$ entstat ent8
-----
ETHERNET STATISTICS (ent8) :
Device Type: Shared Ethernet Adapter
Hardware Address: 00:11:25:cc:80:38
Elapsed Time: 0 days 0 hours 0 minutes 10 seconds

Transmit Statistics:                                Receive Statistics:
-----
Packets: 9                                           Packets: 10
Bytes: 788                                           Bytes: 830
Interrupts: 0                                         Interrupts: 10
Transmit Errors: 0                                    Receive Errors: 0
Packets Dropped: 0                                   Packets Dropped: 0
Bad Packets: 0                                        Bad Packets: 0
```

```

Max Packets on S/W Transmit Queue: 1
S/W Transmit Queue Overflow: 0
Current S/W+H/W Transmit Queue Length: 1

Elapsed Time: 0 days 0 hours 0 minutes 0 seconds
Broadcast Packets: 1
Multicast Packets: 9
No Carrier Sense: 0
DMA Underrun: 0
Lost CTS Errors: 0
Max Collision Errors: 0
Late Collision Errors: 0
Deferred: 0
SQE Test: 0
Timeout Errors: 0
Single Collision Count: 0
Multiple Collision Count: 0
Current HW Transmit Queue Length: 1

Broadcast Packets: 1
Multicast Packets: 9
CRC Errors: 0
DMA Overrun: 0
Alignment Errors: 0
No Resource Errors: 0
Receive Collision Errors: 0
Packet Too Short Errors: 0
Packet Too Long Errors: 0
Packets Discarded by Adapter: 0
Receiver Start Count: 0

General Statistics:
-----
No mbuf Errors: 0
Adapter Reset Count: 0
Adapter Data Rate: 0
Driver Flags: Up Broadcast Running
                Simplex 64BitSupport ChecksumOffload
                DataRateSet

```

---

The **entstat -all** command can be used to provide all the information related to ent8 and all the adapters integrated to it as shown in Example 17-7.

*Example 17-7 entstat -all command on SEA*

---

```

$ entstat -all ent8 |grep -E "Packets:|ETHERNET"
ETHERNET STATISTICS (ent8) :
Packets: 111
Broadcast Packets: 8
Multicast Packets: 103
ETHERNET STATISTICS (ent6) :
Packets: 18
Broadcast Packets: 8
Multicast Packets: 10
ETHERNET STATISTICS (ent0) :
Packets: 5
Broadcast Packets: 0
Multicast Packets: 5
ETHERNET STATISTICS (ent2) :
Packets: 13
Broadcast Packets: 8

Packets: 101
Bad Packets: 0
Broadcast Packets: 8
Multicast Packets: 93
Packets: 93
Bad Packets: 0
Broadcast Packets: 0
Multicast Packets: 93
Packets: 87
Bad Packets: 0
Broadcast Packets: 0
Multicast Packets: 87
Packets: 6
Bad Packets: 0
Broadcast Packets: 0

```



Multicast Packets: 5	Multicast Packets: 6
<b>ETHERNET STATISTICS (ent4) :</b>	
Packets: 93	Packets: 8
Broadcast Packets: 0	Bad Packets: 0
Multicast Packets: 93	Broadcast Packets: 8
Invalid VLAN ID Packets: 0	Multicast Packets: 0
Switch ID: ETHERNET0	

**Note:** Because the Link Aggregation Control Protocol (LACP) is being used in this example, it is possible to see some packets flowing as the switch and the logical partitions negotiate configurations.

Note that you can see the statistics of the Shared Ethernet Adapter (ent8), of the Link Aggregation device (ent6), the physical devices (ent0 and ent2), and the virtual Ethernet adapter (ent4).

Now execute the first data transfer and check the statistics again. To transfer the data, log in to the Linux server box and do an ftp to the Virtual I/O Server. To do that, execute the following commands:

```
ftp> put "| dd if=/dev/zero bs=1M count=100" /dev/zero
local: | dd if=/dev/zero bs=1M count=100 remote: /dev/zero
229 Entering Extended Passive Mode (|||32851|)
150 Opening data connection for /dev/zero.
100+0 records in
100+0 records out
104857600 bytes (105 MB) copied, 8.85929 seconds, 11.8 MB/s
226 Transfer complete.
104857600 bytes sent in 00:08 (11.28 MB/s)
```

In this operation 100 MB were transferred from the Linux server to the Virtual I/O Server. In fact no file was transferred since the **dd** command was used to create 100 packets of 1 MB each, fill them up with zeros, and transfer them to /dev/zero on the Virtual I/O Server.

You can check which adapter was used to transfer the file. Execute the **entstat** command and see the number of packets, as shown in Example 17-8. Compared to the number of packets shown in Example 17-7 on page 484 the number increased after the first file transfer.

*Example 17-8 entstat -all command after file transfer attempt 1*

---

```
$ entstat -all ent8 |grep -E "Packets:|ETHERNET"
ETHERNET STATISTICS (ent8) :
Packets: 35485
Packets: 74936
Bad Packets: 0
```

Broadcast Packets: 23	Broadcast Packets: 23
Multicast Packets: 238	Multicast Packets: 214
ETHERNET STATISTICS (ent6) :	
Packets: 35270	Packets: 74914
	Bad Packets: 0
Broadcast Packets: 22	Broadcast Packets: 1
Multicast Packets: 24	Multicast Packets: 214
ETHERNET STATISTICS (ent0) :	
<b>Packets: 14</b>	Packets: 74901
	Bad Packets: 0
Broadcast Packets: 0	Broadcast Packets: 1
Multicast Packets: 12	Multicast Packets: 201
<b>ETHERNET STATISTICS (ent2) :</b>	
<b>Packets: 35256</b>	Packets: 13
	Bad Packets: 0
Broadcast Packets: 22	Broadcast Packets: 0
Multicast Packets: 12	Multicast Packets: 13
ETHERNET STATISTICS (ent4) :	
Packets: 215	Packets: 22
	Bad Packets: 0
Broadcast Packets: 1	Broadcast Packets: 22
Multicast Packets: 214	Multicast Packets: 0
Invalid VLAN ID Packets: 0	
Switch ID: ETHERNET0	

---

Note the packet count for the ent2 interface. It is now 35256. This means that physical adapter ent2 was chosen to transfer the file this time.

Go back to the ftp session and transfer another 100 MB and check again for the adapter statistics.

The ftp session should look like this:

```
ftp> put "| dd if=/dev/zero bs=1M count=100" /dev/zero
local: | dd if=/dev/zero bs=1M count=100 remote: /dev/zero
229 Entering Extended Passive Mode (|||32855|)
150 Opening data connection for /dev/zero.
100+0 records in
100+0 records out
104857600 bytes (105 MB) copied, 8.84978 seconds, 11.8 MB/s
226 Transfer complete.
104857600 bytes sent in 00:08 (11.29 MB/s)
ftp> quit
221 Goodbye.
```

Check the adapter statistics one more time as shown in Example 17-9.

*Example 17-9 entstat -all command after file transfer attempt 2*

---

```
$ entstat -all ent8 |grep -E "Packets:|ETHERNET"
ETHERNET STATISTICS (ent8) :
Packets: 70767                               Packets: 149681
                                              Bad Packets: 0
Broadcast Packets: 37                       Broadcast Packets: 37
Multicast Packets: 294                     Multicast Packets: 264
ETHERNET STATISTICS (ent6) :
Packets: 70502                               Packets: 149645
                                              Bad Packets: 0
Broadcast Packets: 36                       Broadcast Packets: 1
Multicast Packets: 30                     Multicast Packets: 264
ETHERNET STATISTICS (ent0) :
Packets: 17                                 Packets: 149629
                                              Bad Packets: 0
Broadcast Packets: 0                       Broadcast Packets: 1
Multicast Packets: 15                     Multicast Packets: 248
ETHERNET STATISTICS (ent2) :
Packets: 70485                               Packets: 16
                                              Bad Packets: 0
Broadcast Packets: 36                       Broadcast Packets: 0
Multicast Packets: 15                     Multicast Packets: 16
ETHERNET STATISTICS (ent4) :
Packets: 265                               Packets: 36
                                              Bad Packets: 0
Broadcast Packets: 1                       Broadcast Packets: 36
Multicast Packets: 264                     Multicast Packets: 0
Invalid VLAN ID Packets: 0
Switch ID: ETHERNET0
```

---

Note that this time the ent2 interface was used again and has increased to 70485.

Open a new ftp session to the Virtual I/O Server, transfer another amount of data, and then verify which interface was used.

On the Linux server open a new ftp session and transfer the data:

```
ftp> put "| dd if=/dev/zero bs=1M count=100" /dev/zero
local: | dd if=/dev/zero bs=1M count=100 remote: /dev/zero
229 Entering Extended Passive Mode (|||32858|)
150 Opening data connection for /dev/zero.
100+0 records in
100+0 records out
104857600 bytes (105 MB) copied, 8.85152 seconds, 11.8 MB/s
226 Transfer complete.
```

```
104857600 bytes sent in 00:08 (11.28 MB/s)
ftp>
```

On the Virtual I/O Server check the interface statistics to identify which interface was used to transfer the data this time as shown in Example 17-10.

*Example 17-10 entstat -all command after file transfer attempt 3*

---

```
$ entstat -all ent8 |grep -E "Packets:|ETHERNET"
ETHERNET STATISTICS (ent8) :
Packets: 106003                               Packets: 224403
                                                Bad Packets: 0
Broadcast Packets: 44                         Broadcast Packets: 44
Multicast Packets: 319                       Multicast Packets: 287
ETHERNET STATISTICS (ent6) :
Packets: 105715                               Packets: 224360
                                                Bad Packets: 0
Broadcast Packets: 43                         Broadcast Packets: 1
Multicast Packets: 32                       Multicast Packets: 287
ETHERNET STATISTICS (ent0) :
Packets: 35219                               Packets: 224343
                                                Bad Packets: 0
Broadcast Packets: 0                         Broadcast Packets: 1
Multicast Packets: 16                       Multicast Packets: 270
ETHERNET STATISTICS (ent2) :
Packets: 70496                               Packets: 17
                                                Bad Packets: 0
Broadcast Packets: 43                       Broadcast Packets: 0
Multicast Packets: 16                       Multicast Packets: 17
ETHERNET STATISTICS (ent4) :
Packets: 288                               Packets: 43
                                                Bad Packets: 0
Broadcast Packets: 1                         Broadcast Packets: 43
Multicast Packets: 287                       Multicast Packets: 0
Invalid VLAN ID Packets: 0
Switch ID: ETHERNET0
```

---

Note that this time the ent0 adapter was used to transfer the data.

In this case, if you open multiple sessions from the Linux server to the Virtual I/O Server, the traffic should be divided between the adapters since each ftp session will be using a different port number.

Open two ftp connections to the Virtual I/O Server and check them.

First of all, reset all the statistics of the adapters:

```
$ for i in 1 2 3 4 5 6 7 8
> do
```

```
> entstat -reset ent$i
> done
```

Then check the adapter statistics as shown in Example 17-11.

*Example 17-11 entstat -all command after reset of Ethernet adapters*

---

```
$ entstat -all ent8 |grep -E "Packets:|ETHERNET"
ETHERNET STATISTICS (ent8) :
Packets: 1                               Packets: 1
                                         Bad Packets: 0
Broadcast Packets: 0                     Broadcast Packets: 0
Multicast Packets: 1                     Multicast Packets: 1
ETHERNET STATISTICS (ent6) :
Packets: 0                               Packets: 1
                                         Bad Packets: 0
Broadcast Packets: 0                     Broadcast Packets: 0
Multicast Packets: 0                     Multicast Packets: 1
ETHERNET STATISTICS (ent0) :
Packets: 0                               Packets: 1
                                         Bad Packets: 0
Broadcast Packets: 0                     Broadcast Packets: 0
Multicast Packets: 0                     Multicast Packets: 1
ETHERNET STATISTICS (ent2) :
Packets: 0                               Packets: 0
                                         Bad Packets: 0
Broadcast Packets: 0                     Broadcast Packets: 0
Multicast Packets: 0                     Multicast Packets: 0
ETHERNET STATISTICS (ent4) :
Packets: 1                               Packets: 0
                                         Bad Packets: 0
Broadcast Packets: 0                     Broadcast Packets: 0
Multicast Packets: 1                     Multicast Packets: 0
Invalid VLAN ID Packets: 0
Switch ID: ETHERNET0
```

---

On the first terminal open an ftp session from the Linux server to the Virtual I/O Server, but this time use a larger amount of data in order to give you time to open a second ftp session, as follows:

```
server1:~ # ftp 10.10.10.15
Connected to 10.10.10.15.
220 VIO_Server1 FTP server (Version 4.2 Fri Oct 17 07:20:05 CDT 2008) ready.
Name (10.10.10.15:root): padmin
331 Password required for padmin.
Password:
230-Last unsuccessful login: Thu Oct 16 20:26:56 CST 2008 on ftp from
::ffff:10.10.10.11
230-Last login: Thu Oct 16 20:27:02 CST 2008 on ftp from ::ffff:10.10.10.11
```

```

230 User padmin logged in.
Remote system type is UNIX.
Using binary mode to transfer files.
ftp> put "| dd if=/dev/zero bs=1M count=1000" /dev/zero
local: | dd if=/dev/zero bs=1M count=1000 remote: /dev/zero
229 Entering Extended Passive Mode (|||33038|)
150 Opening data connection for /dev/zero.

```

Check the statistics of the adapter as shown in Example 17-12.

*Example 17-12 entstat -all command after opening one ftp session*

---

```

$ entstat -all ent8 |grep -E "Packets:|ETHERNET"
ETHERNET STATISTICS (ent8) :
Packets: 41261                               Packets: 87496
                                             Bad Packets: 0
Broadcast Packets: 11                       Broadcast Packets: 11
Multicast Packets: 38                       Multicast Packets: 34
ETHERNET STATISTICS (ent6) :
Packets: 41241                               Packets: 87521
                                             Bad Packets: 0
Broadcast Packets: 11                       Broadcast Packets: 0
Multicast Packets: 4                       Multicast Packets: 34
ETHERNET STATISTICS (ent0) :
Packets: 41235                             Packets: 87561
                                             Bad Packets: 0
Broadcast Packets: 0                       Broadcast Packets: 0
Multicast Packets: 2                       Multicast Packets: 32
ETHERNET STATISTICS (ent2) :
Packets: 21                                 Packets: 2
                                             Bad Packets: 0
Broadcast Packets: 11                       Broadcast Packets: 0
Multicast Packets: 2                       Multicast Packets: 2
ETHERNET STATISTICS (ent4) :
Packets: 34                                 Packets: 11
                                             Bad Packets: 0
Broadcast Packets: 0                       Broadcast Packets: 11
Multicast Packets: 34                       Multicast Packets: 0
Invalid VLAN ID Packets: 0
Switch ID: ETHERNET0

```

---

The first ftp session is using the physical adapter ent0 to transfer the data. Now open a second terminal and a new ftp session to the Virtual I/O Server:

```

server1:~ # ftp 10.10.10.15
Connected to 10.10.10.15.
220 VIO_Server1 FTP server (Version 4.2 Fri Oct 17 07:20:05 CDT 2008) ready.
Name (10.10.10.15:root): padmin
331 Password required for padmin.

```

```

Password:
230-Last unsuccessful login: Thu Oct 16 20:26:56 CST 2008 on ftp from
::ffff:10.10.10.11
230-Last login: Thu Oct 16 20:29:57 CST 2008 on ftp from ::ffff:10.10.10.11
230 User padmin logged in.
Remote system type is UNIX.
Using binary mode to transfer files.
ftp> put "| dd if=/dev/zero bs=1M count=1000" /dev/null
local: | dd if=/dev/zero bs=1M count=1000 remote: /dev/null
229 Entering Extended Passive Mode (|||33041|)
150 Opening data connection for /dev/null.
1000+0 records in
1000+0 records out
1048576000 bytes (1.0 GB) copied, 154.686 seconds, 6.8 MB/s
226 Transfer complete.
1048576000 bytes sent in 02:34 (6.46 MB/s)
ftp>

```

**Note:** In the second ftp session the device /dev/null was used, since the device /dev/zero was already used by the first ftp session.

At this time both ftp transfers should have been completed. Check the adapter statistics again as shown in Example 17-13.

*Example 17-13 entstat -all command after opening two ftp session*

---

```

$ entstat -all ent8 |grep -E "Packets:|ETHERNET"
ETHERNET STATISTICS (ent8) :
Packets: 704780                               Packets: 1493888
                                                Bad Packets: 0
Broadcast Packets: 108                         Broadcast Packets: 108
Multicast Packets: 437                         Multicast Packets: 391
ETHERNET STATISTICS (ent6) :
Packets: 704389                               Packets: 1493780
                                                Bad Packets: 0
Broadcast Packets: 108                         Broadcast Packets: 0
Multicast Packets: 46                          Multicast Packets: 391
ETHERNET STATISTICS (ent0) :
Packets: 352118                               Packets: 1493757
                                                Bad Packets: 0
Broadcast Packets: 0                           Broadcast Packets: 0
Multicast Packets: 23                          Multicast Packets: 368
ETHERNET STATISTICS (ent2) :
Packets: 352271                               Packets: 23
                                                Bad Packets: 0
Broadcast Packets: 108                         Broadcast Packets: 0
Multicast Packets: 23                          Multicast Packets: 23
ETHERNET STATISTICS (ent4) :

```

Packets: 391	Packets: 108
Broadcast Packets: 0	Bad Packets: 0
Multicast Packets: 391	Broadcast Packets: 108
Invalid VLAN ID Packets: 0	Multicast Packets: 0
Switch ID: ETHERNET0	

---

Note that both adapters were used to transfer approximately the same amount of data. It shows that traffic of the ftp sessions was spread across both adapters.

This example can illustrate how network utilization can be monitored and improved with the tuning of a parameter on the Shared Ethernet Adapter. It can also be used as base for new configurations and monitoring.

## 17.2.2 Advanced SEA monitoring

On Virtual I/O Server an advanced tool **seastat** can be used to keep track of the number of packets and bytes received by and sent from each MAC address. It can monitor traffic for the Virtual I/O Server and individual virtual I/O clients. User can monitor the statistics based on a number of parameters like MAC address, IP address, hostname etc.

### How to use seastat

In order to use advanced statistics using **seastat** it has to be enabled first as shown in Example 17-14 on page 492. In this example ent5 is an SEA.

Format of seastat command is:

```
seastat -d <device_name> -c [-n | -s search_criterion=value]
```

<device\_name> is the shared adapter device whose statistics is sought.

-c is used to clear all per client SEA statistics.

-n displays name resolution on the IP addresses.

#### *Example 17-14 Enabling advanced SEA monitoring*

---

```
$ seastat -d ent5
Device ent5 has accounting disabled

$ lsdev -dev ent5 -attr
accounting disabled Enable per-client accounting of network statistics True
ctl_chan ent3 Control Channel adapter for SEA failover True
gvrp no Enable GARP VLAN Registration Protocol (GVRP) True
ha_mode auto High Availability Mode True
jumbo_frames no Enable Gigabit Ethernet Jumbo Frames True
large_receive no Enable receive TCP segment aggregation True
largesend 0 Enable Hardware Transmit TCP Resegmentation True
```



```

netaddr      0      Address to ping                               True
pvid         1      PVID to use for the SEA device               True
pvid_adapter ent2    Default virtual adapter to use for non-VLAN-tagged packets True
qos_mode     disabled N/A                                         True
real_adapter ent0    Physical adapter associated with the SEA      True
thread       1      Thread mode enabled (1) or disabled (0)     True
virt_adapters ent2   List of virtual adapters associated with the SEA (comma separated) True

$ chdev -dev ent5 -attr accounting=enabled
ent5 changed

$ lsdev -dev ent5 -attr
accounting enabled Enable per-client accounting of network statistics True
ctl_chan    ent3    Control Channel adapter for SEA failover     True
gvrp        no      Enable GARP VLAN Registration Protocol (GVRP) True
ha_mode     auto    High Availability Mode                       True
jumbo_frames no      Enable Gigabit Ethernet Jumbo Frames        True
large_receive no      Enable receive TCP segment aggregation      True
largesend   0      Enable Hardware Transmit TCP Resegmentation True
netaddr     0      Address to ping                               True
pvid        1      PVID to use for the SEA device               True
pvid_adapter ent2   Default virtual adapter to use for non-VLAN-tagged packets True
qos_mode    disabled N/A                                         True
real_adapter ent0    Physical adapter associated with the SEA      True
thread      1      Thread mode enabled (1) or disabled (0)     True
virt_adapters ent2   List of virtual adapters associated with the SEA (comma separated) True

```

---

Example 17-15 on page 493 shows SEA statistics without any search criterion. So, its displaying statistics for all the clients this Virtual I/O Server is serving to.

*Example 17-15 Sample seastat statistics*

---

```

$ seastat -d ent5

=====

Advanced Statistics for SEA
Device Name: ent5

=====
MAC: 6A:88:82:AA:9B:02
-----

VLAN: None
VLAN Priority: None

Transmit Statistics:                Receive Statistics:
-----                          -----
Packets: 7                          Packets: 2752
Bytes: 420                          Bytes: 185869

=====
MAC: 6A:88:82:AA:9B:02
-----

VLAN: None
VLAN Priority: None
IP: 9.3.5.115

Transmit Statistics:                Receive Statistics:
-----                          -----
Packets: 125                          Packets: 3260
Bytes: 117242                          Bytes: 228575

```

```
=====
MAC: 6A:88:85:BF:16:02
-----

VLAN: None
VLAN Priority: None

Transmit Statistics:          Receive Statistics:
-----
Packets: 1                   Packets: 1792
Bytes: 42                     Bytes: 121443

=====
MAC: 6A:88:86:26:F1:02
-----

VLAN: None
VLAN Priority: None
IP: 9.3.5.119

Transmit Statistics:          Receive Statistics:
-----
Packets: 2                   Packets: 1573
Bytes: 190                   Bytes: 107535

=====
MAC: 6A:88:86:26:F1:02
-----

VLAN: None
VLAN Priority: None

Transmit Statistics:          Receive Statistics:
-----
Packets: 1                   Packets: 2575
Bytes: 42                     Bytes: 173561

=====
MAC: 6A:88:8D:E7:80:0D
-----

VLAN: None
VLAN Priority: None
Hostname: vios1
IP: 9.3.5.111

Transmit Statistics:          Receive Statistics:
-----
Packets: 747                 Packets: 3841
Bytes: 364199                 Bytes: 327541

=====
MAC: 6A:88:8D:E7:80:0D
-----

VLAN: None
VLAN Priority: None

Transmit Statistics:          Receive Statistics:
-----
Packets: 10                  Packets: 3166
Bytes: 600                    Bytes: 214863

=====
MAC: 6A:88:8F:36:34:02
-----
```

```

VLAN: None
VLAN Priority: None

Transmit Statistics:          Receive Statistics:
-----
Packets: 9                   Packets: 3149
Bytes: 540                   Bytes: 213843

=====
MAC: 6A:88:8F:36:34:02
-----

VLAN: None
VLAN Priority: None
IP: 9.3.5.121

Transmit Statistics:          Receive Statistics:
-----
Packets: 125                 Packets: 3256
Bytes: 117242                Bytes: 229103

=====
MAC: 6A:88:8F:ED:33:0D
-----

VLAN: None
VLAN Priority: None

Transmit Statistics:          Receive Statistics:
-----
Packets: 10                  Packets: 3189
Bytes: 600                   Bytes: 216243

=====
MAC: 6A:88:8F:ED:33:0D
-----

VLAN: None
VLAN Priority: None
IP: 9.3.5.112

Transmit Statistics:          Receive Statistics:
-----
Packets: 330                 Packets: 3641
Bytes: 194098                Bytes: 309419

=====
$

```

This command will show an entry for each pair of VLAN, VLAN priority, IP-address and MAC address. So, you will notice in Example 17-15 there are 2 entries for several MAC addresses. One entry is for MAC address while the other one is for the IP address configured over that MAC.

### Statistics based on search criterion

**seastat** can also print statistics based on some search criteria. Currently following search criterion are supported:

- ▶ MAC address (mac)

- ▶ Priority, as explained in 3.6, “DoS Hardening” on page 140
- ▶ VLAN id (vlan)
- ▶ IP address (ip)
- ▶ Hostname (host)
- ▶ Greater than bytes sent (gbs)
- ▶ Greater than bytes recv (gbr)
- ▶ Greater than packets sent (gps)
- ▶ Greater than packets recv (gpr)
- ▶ Smaller than bytes sent (sbs)
- ▶ Smaller than bytes recv (sbr)
- ▶ Smaller than packets sent (sps)
- ▶ Smaller than packets recv (spr)

In order to use a search criterion you need to specify it in the form:

`<search_criteria>=<value>`

Example 17-16 on page 496 shows statistics based on a search criterion as IP address. It is specified as “`ip=9.3.5.121`”

*Example 17-16 seastat statistics using search criterion*

---

```

$ seastat -d ent5 -n
=====
Advanced Statistics for SEA
Device Name: ent5
=====
MAC: 6A:88:8D:E7:80:0D
-----

VLAN: None
VLAN Priority: None
IP: 9.3.5.111

Transmit Statistics:          Receive Statistics:
-----
Packets: 13                  Packets: 81
Bytes: 2065                   Bytes: 5390

=====
MAC: 6A:88:8F:36:34:02
-----

VLAN: None
VLAN Priority: None
IP: 9.3.5.121

Transmit Statistics:          Receive Statistics:

```

```

-----
Packets: 1                               Packets: 23
Bytes: 130                               Bytes: 1666
-----
$
$ seastat -d ent5 -s ip=9.3.5.121
-----

Advanced Statistics for SEA
Device Name: ent5

-----
MAC: 6A:88:8F:36:34:02
-----

VLAN: None
VLAN Priority: None
IP: 9.3.5.121

Transmit Statistics:                     Receive Statistics:
-----
Packets: 115                             Packets: 8542
Bytes: 14278                             Bytes: 588424
-----

```

## 17.3 AIX client network monitoring

On the AIX virtual I/O client the **entstat** command can be used to monitor a virtual Ethernet adapter as shown in examples above. Likewise, it can also be used to monitor a physical Ethernet adapter.

## 17.4 IBM i client network monitoring

In this section we describe monitoring the network health and performance from an IBM i virtual I/O client.

### 17.4.1 Checking network health on the IBM i virtual I/O client

To have an active working TCP/IP network configuration on the IBM i virtual I/O client check for the following:

1. The TCP/IP interface is in an *active* state:

Use the **WRKTCPSTS \*IFC** command as shown in Example 17-17.

If the interface is not active, then if it is in *inactive* state start it using option **9=Start** otherwise proceed with the next step.

*Example 17-17 IBM i Work with TCP/IP Interface Status screen***Work with TCP/IP Interface Status**

System:

E101F170

Type options, press Enter.

5=Display details 8=Display associated routes 9=Start 10=End  
 12=Work with configuration status 14=Display multicast groups

Opt	Internet Address	Network Address	Line Description	Interface Status
	9.3.5.119	9.3.4.0	ETH01	Active
	127.0.0.1	127.0.0.0	*LOOPBACK	Active

Bottom

F3=Exit F9=Command line F11=Display line information F12=Cancel  
 F13=Sort by column F20=Work with IPv6 interfaces F24=More keys

2. The corresponding Ethernet line description is *active* (varied on):

Use the **WRKCFGSTS \*LIN** command as shown in Example 17-18.

If the line description is not active, then if it is *varied off* vary it on using option **1=Vary on** otherwise proceed with the next step.

*Example 17-18 IBM i Work with Configuration Status screen***Work with Configuration Status**

E101F170

11/03/08

13:42:17

Position to . . . . .

Starting characters

Type options, press Enter.

1=Vary on 2=Vary off 5=Work with job 8=Work with description  
 9=Display mode status 13=Work with APPN status...

Opt	Description	Status	-----Job-----		
	ETH01	ACTIVE			
	ETH01NET01	ACTIVE			
	ETH01TCP01	ACTIVE	QTCPWRK	QSYS	020473

```

QESLINE          VARIED OFF
QTILINE          VARIED OFF

```

Bottom

Parameters or command

====>

F3=Exit F4=Prompt F12=Cancel F23=More options F24=More keys

---

3. The corresponding virtual Ethernet adapter resource CMNxx (type 268C) is *operational*:

Use the **WRKHDWRSC \*CMN** command as shown in Example 17-19.

If the virtual Ethernet adapter resource is not operational, then if it is in *inoperational* state you can try a reset of the virtual IOP resource for recovery, if it is in *not connected* state check the IBM i and Virtual I/O Server virtual Ethernet adapter partition configuration.

*Example 17-19 IBM i Work with Communication Resources screen*

---

#### Work with Communication Resources

System:

E101F170

Type options, press Enter.

5=Work with configuration descriptions 7=Display resource detail

Opt	Resource	Type	Status	Text
	CMB06	6B03	Operational	Comm Processor
	LIN03	6B03	Operational	Comm Adapter
	CMN02	6B03	Operational	Comm Port
	CMB07	6B03	Operational	Comm Processor
	LIN01	6B03	Operational	Comm Adapter
	CMN03	6B03	Operational	Comm Port
	<b>CMB08</b>	<b>268C</b>	<b>Operational</b>	<b>Comm Processor</b>
	<b>LIN02</b>	<b>268C</b>	<b>Operational</b>	<b>LAN Adapter</b>
	<b>CMN08</b>	<b>268C</b>	<b>Operational</b>	<b>Ethernet Port</b>

Bottom

F3=Exit F5=Refresh F6=Print F12=Cancel

---

## 17.4.2 Monitoring network performance on the IBM i virtual I/O client

In the following we show examples for IBM i network monitoring using IBM Performance Tools for i5/OS.

Example 17-20 shows a *system report* for *TCP/IP Summary* we created via the following command:

```
PRTSYSRPT MBR(Q308150002) TYPE(*TCPIP)
```

### Example 17-20 IBM i System Report for TCP/IP Summary

```
System Report                               110308 16:16:5
                                           TCP/IP Summary
Page 000
Member . . . : Q308150002 Model/Serial . . : MMA/10-1F170      Main storage . . : 8192.0 MB Started . . . . : 11/03/08
15:00:0
Library . . . : QPFRDATA System name . . : E101F170          Version/Release : 6/ 1.0 Stopped . . . . : 11/03/08
16:00:0
Partition ID : 005 Feature Code . . : 5622-5622           Int Threshold . . : .00 %
Virtual Processors: 2 Processor Units : 1.00
MTU          KB          ----- Packets Received ----- KB          ----- Packets Sent -----
-----
Line Type/   Size      Received      Number      Pct Transmitted
Pct
Line Name   (bytes)  /Second      Unicast     Non-Unicast  Error  Error  /Second      Unicast
Non-Unicast Erro
-----
                    576
*LOOPBACK          0          2          0          0 .00          0          2
0 .0
ETHERNET          1,492
ETH01              190         472,183         1,768          0 .00          3         209,000
42 .0
```

Example 17-21 shows a *component report* for *TCP/IP Activity* we created via the following command:

```
PRTCPTRPT MBR(Q308150002) TYPE(*TCPIP)
```

### Example 17-21 IBM i Resource Report for Disk Utilization

```
Component Report                               11/03/08 17:17:0
                                           TCP/IP Activity
Page Member . . . : Q308150002 Model/Serial . . : MMA/10-1F170      Main storage . . : 8192.0 MB Started . . . . :
11/03/08 15:00:0
Library . . . : QPFRDATA System name . . : E101F170          Version/Release : 6/ 1.0 Stopped . . . . : 11/03/08
16:00:0
Partition ID : 005 Feature Code . . : 5622-5622           Int Threshold . . : .00 %
Virtual Processors: 2 Processor Units : 1.00
Line Type/Line Name : ETHERNET /ETH01 MTU Size: 1492 bytes
KB          ----- Packets Received ----- KB          ----- Packets Sent -----
Itv  Received      Pct Transmitted
End  /Second      Error  /Second      Unicast      Non-Unicast      Pct
-----
More...
15:08          0          11          29 .00          0          0          0 .00
15:09          0          13          35 .00          0          0          0 .00
```



15:11	0	70	26	.00	0	46	0	.00
15:12	0	354	34	.00	2	485	0	.00
15:13	0	230	28	.00	1	353	5	.00
15:14	0	26	32	.00	0	9	0	.00
15:15	<b>11,412</b>	470,644	31	<b>.00</b>	<b>176</b>	207,984	0	<b>.00</b>
15:16	0	11	27	.00	0	0	2	.00
More...								

---

These system and component reports for TCP/IP can help to get a first overview of IBM i network usage and performance by providing information about the network I/O throughput and the percentage of packet errors.

For additional information about IBM Performance Tools for i5/OS refer to *IBM eServer iSeries Performance Tools for iSeries*, SC41-5340.

Another approach for long-term network monitoring for IBM i which also allows IBM i cross-partition monitoring is using the System i Navigator's Management Central monitors function. Based on the experienced average network utilization the user may define a threshold to be notified about an unusual high network utilization to be alerted for potential network performance problems.

For further information about using System i Navigator for performance monitoring refer to *Managing OS/400 with Operations Navigator V5R1 Volume 5: Performance Management*, SG24-6565.

## 17.5 Linux for Power client network monitoring

Red Hat and Novell Linux distributors include the netstat utility in the default installation. The **netstat** command with the -p option gives you information on socket and process id. This output (-p option used) provides information on processes using the network resources. Hence the bandwidth usage can be collected by analyzing the throughput on per adapter basis from the **netstat** command execution.

Besides the netstat utility, users can also use tcpdump utility from the link shown below.

<http://sourceforge.net/projects/tcpdump>

The **tcpdump** command is a tool for network monitoring, protocol debugging and data acquisition. The tool precisely sees all the network traffic. Hence, this tool can be used to create statistical monitoring scripts. A major drawback for this tool is the size of the flat file containing text output. Refer to the distributors documentation for more information on the availability of this tool in the current Linux release.





## Third-party monitoring tools for AIX and Linux

Linux does not have all the monitoring commands that are available for AIX or PowerVM Virtualization. The `/proc/ppc64/lparcfg` special device provides significant information.

This information is used by the following third-party tools that can be used both on AIX and Linux operating systems.

Note that these tools are not IBM products; they are supported through their developers and user communities.

### 18.1 nmon utility

The **nmon** utility is a freely downloadable monitoring tool for AIX and Linux. This tool provides a text-based summary, similar to the **topas** command, of key system metrics. Both online immediate monitoring or saving the data to a file for later analysis, nmon analyzer provides an easy way of transforming the saved file data into the key important graphs.

**Note:** The `nmon` utility is included in AIX 6.1 TL2, or later releases and in the Virtual I/O Server Version 2.1, or later releases. To use the `nmon` utility at the shell prompt execute the command `topas →~` (tilde). So the information on `nmon` might be used if the system is running earlier versions of AIX or Virtual I/O Server.

As of version 11, the `nmon` command is simultaneous multithreading and partition aware. `nmon` version 12 now integrates CPU donation statistics.

The `nmon` command is available from:

<http://www-941.haw.ibm.com/collaboration/wiki/display/WikiPtype/nmon>

Extract and copy the `nmon` binary file to the partition you want to monitor, typically under `/usr/sbin/nmon`. Optionally, change and verify the `iostat` flags to continuously maintain the disk I/O history using the following commands:

```
# lsattr -E -l sys0 -a iostat
iostat false Continuously maintain DISK I/O history True
# chdev -l sys0 -a iostat=true
sys0 changed
# lsattr -E -l sys0 -a iostat
iostat true Continuously maintain DISK I/O history True
```

You may get a warning if you do not position this flag to true, but `nmon` will continue to show non-null values for disk usage.

**Important:** If you have a large number of disks (more than 200), then setting `iostat` to true will start consuming CPU time, around 2 percent. Once the measurement campaign is completed you should set the flag back to false.

Run the `nmon` command and then press `p` to show partition statistics. The result differs between the AIX and Linux systems.

### 18.1.1 nmon on AIX 6.1

On AIX 6.1 systems, `nmon` provides hundreds of statistics including nearly all that `topas` provides. This includes, but is not limited to:

- ▶ CPU
- ▶ Memory
- ▶ Paging
- ▶ Network
- ▶ Disks

- ▶ Logical volumes
- ▶ File systems
- ▶ NFS
- ▶ Async I/O
- ▶ Fibre Channel adapters
- ▶ SEA
- ▶ Kernel numbers
- ▶ Multiple page size stats
- ▶ ESS
- ▶ WLM
- ▶ WPARs
- ▶ Top processes

As you see in Figure 18-1 and Figure 18-2 on page 506, the output changes if you are running on a shared or a dedicated partition.

```

--nmon--p=Partitions--Host=DB_server--Refresh=2 secs--18:15.24
Shared-CPU-Logical-Partition
Partition: Number=8 "DB_server_mobile"
Flags: LPARed DRable SMT-bound Shared UnCapped PoolAuth Mover Not-Donating.
Summary: Entitled= 1.00 Used 0.01 ( 0.6%) 0.2% of CPUs in System
PoolCPUs= 3 Unused 2.99 0.2% of CPUs in Pool
CPU-Stats----- Capacity----- ID-Memory-----
max Phys in sys 16 Cap. Processor Min 0.50 LPAR ID Group:Pool 32776:10
Phys CPU in sys 4 Cap. Processor Max 4.00 Memory(MB) Min:Max 512:6144
Virtual Online 2 Cap. Increment 0.01 Memory(MB) Online 1024
Logical Online 4 Cap. Unallocated 0.00 Memory Region LMB 128MB min
Physical pool 3 Cap. Entitled 1.00 Time-----Seconds
SMT threads/CPU 2 -MinReqVirtualCPU 0.10 Time Dispatch Wheel 0.0100
CPU-----Min-Max Weight----- MaxDispatch Latency 0.0100
Virtual 2 4 Weight Variable 128 Time Pool Idle 2.9862
Logical 2 8 Weight Unallocated 0 Time Total Dispatch 0.0064
-----
Event= 0 --- --- SerialNo Old=--- Current=C1F170 When=---
-----
Not a Dedicated Donating LPAR

```

Figure 18-1 nmon LPAR statistics on an AIX shared partition

At the top you see the partition type (shared, uncapped, dedicated, sharable). You also find donation information at the bottom for dedicated partitions.

```

-nmon-----p=Partitions-----Host=Apps_server-----Refresh=2 secs-----18:15.11
Shared-CPU-Logical-Partition
Partition: Number=4 "Apps_server"
Flags: LPARed DRable SMT-bound Dedicated Sharable Mover Donating Not-Donating.
Summary: Entitled= 1.00 Used 0.00 ( 0.4%) 0.1% of CPUs in System
- You don't have Shared Processor Pool Utilisation Authority
CPU-Stats----- Capacity----- ID-Memory-----
max Phys in sys 16 Cap. Processor Min 1.00 LPAR ID Group:Pool 32772:65535
Phys CPU in sys 4 Cap. Processor Max 3.00 Memory(MB) Min:Max 128:8192
Virtual Online 1 Cap. Increment 1.00 Memory(MB) Online 4096
Logical Online 2 Cap. Unallocated 0.00 Memory Region LMB 128MB min
Physical pool 0 Cap. Entitled 1.00 Time-----Seconds
SMT threads/CPU 2 -MinReqVirtualCPU 1.00 Time Dispatch Wheel 0.0000
CPU-----Min-Max Weight----- MaxDispatch Latency 0.0000
Virtual 1 3 Weight Variable 0 Time Pool Idle 0.0000
Logical 1 6 Weight Unallocated 0 Time Total Dispatch 0.0036
-----
Event= 0 --- --- SerialNo Old=--- Current=COF6A0 When=---
-----
Donating-LPAR User System Wait Idle| Idle Busy | Idle Busy|
Physical-CPU 0.00 0.00 0.00 0.00| Donate 0.99 0.00 | Stolen 0.00 0.00|

```

Figure 18-2 nmon LPAR statistics on an AIX dedicated partition

## 18.1.2 nmon on Linux

The nmon tool is available for IBM Power Systems running Red Hat or Novell SUSE Linux distributions. A comprehensive usage of nmon, including the source files can be found at

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

On Linux systems, the number of PowerVM Virtualization-related metrics is restricted. nmon therefore shows fewer statistics in the logical partition section than on AIX 6.1 systems.

```

LPAR Stats
LPAR=5 SerialNumber=IBM,02101F170 Type=IBM,9117-MMA
Flags: Shared-CPU=true Capped=false
Systems CPU Pool= 400.00 Active= 4.00 Total= 16.00
LPARs CPU Min= 0.10 Entitlement= 0.80 Max= 4.00
Virtual CPU Min= 1.00 VP Now= 1.00 Max= 4.00
Memory Min= unknown Now= 256.00 Max= 2048.00
Other Weight= 128.00 UnallocWeight= 0.00 Capacity= 0.01
BoundThrds= 1.00 UnallocCapacity= 0.00 Increment
Physical CPU use= 0.016 [timebase=512000000]

```

Figure 18-3 nmon LPAR statistics report for a Linux partition

### 18.1.3 Additional nmon statistics

**nmon** can also display other kinds of statistics such as those related to disk, network, memory, adapters, and so on. Refer to the **nmon** documentation for more information about these. You can also press **h** while **nmon** is running to get a help summary, or use **nmon -h** for more information about specific options.

### 18.1.4 Recording with the nmon tool

You can record resource usage using **nmon** for subsequent analysis with the **nmon** analyzer tool, or other post-capture tools. This will work on both standard AIX and Linux partitions:

```
# nmon -f -t [-s <seconds> -c <count>]
```

The **nmon** process runs in the background and you can log off the partition if you wish. For best results, the count should not be greater than 1,500. The command creates a file with a name in the following format:

```
<hostname>_<date>_<time>.nmon
```

Once the recording process has finished, transfer the file to a machine that runs Microsoft Excel spreadsheet software to run the **nmon** analyzer tool.

You can find the **nmon\_analyser** tool at:

<http://www-941.haw.ibm.com/collaboration/wiki/display/Wikiptype/nmonanalyser>

## 18.2 Sysstat utility

Sysstat is a package that includes at least three groups of monitoring tools for Linux. This utility might be included in the Linux distributions. Users can also download the current version of this utility from

<http://pagesperso-orange.fr/sebastien.godard/features.html>

The tools included are **sadc**, **sar**, **iostat**, **sadf**, **mpstat**, **pidstat**. These tools can be used for obtaining various metrics on the host partition like **cpu** statistics, **memory**, **paging**, **swap space**, **interrupts**, **network activity**, **task switching activity**. Besides this wide array of system resource monitoring, this tool also offers the following advantages.

- ▶ output can be saved to a file for analysis
- ▶ averages can be calculated over the sampling period

- ▶ specify the duration of data collection
- ▶ support for hotplug in some environment
- ▶ support for 32 and 64 bit architectures

## 18.3 Ganglia tool

Ganglia, at:

<http://ganglia.sourceforge.net/>

is a monitoring system initially designed for large high performance computing clusters and grids. It uses very lightweight agents and may use multicast communication to save computing and network resources.

With additional metrics added, Ganglia can visualize performance data that is specific to a virtualized Power Systems environment. For Power Systems monitoring you may design your Power Systems server as a cluster and treat all the client partitions on your server as nodes in the same cluster. This way the visualization will allow you to see the summarized overall server load.

Ganglia can show all general CPU use on a server and the amount of physical CPU used by each partition in the last hour. This includes AIX, Virtual I/O Server, and Linux partitions. From these graphs, you can find out which partitions are using the shared CPUs most.

Ganglia, for example, can record shared CPU Pool for a whole week to determine which partitions are using CPU cycles and when. Some workloads can be seen as constrained but others only run for some time each day.

Adapted Ganglia packages with additional Power Systems metrics for AIX and Linux for Power and instructions for best practices are provided at:

<http://www.perzl.org/ganglia/>

Best practices can be found at:

<http://www-941.ibm.com/collaboration/wiki/display/WikiPtype/ganglia>

## 18.4 Other third party tools

You can find additional tools on the Internet. Good sources of information are:

- ▶ Virtual I/O Server Monitoring wiki:



[http://www-941.ibm.com/collaboration/wiki/display/WikiPtype/VIOS\\_Monitoring](http://www-941.ibm.com/collaboration/wiki/display/WikiPtype/VIOS_Monitoring)

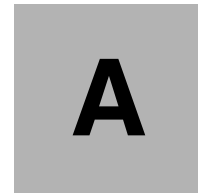
- ▶ Performance monitoring with non AIX tools wiki

<http://www-941.ibm.com/collaboration/wiki/display/WikiPtype/Performance+Other+Tools>

- ▶ Partitioning monitoring using the **lparmon** command:

<http://www.alphaworks.ibm.com/tech/lparmon>





# Sample script for disk and NIB network checking and recovery on AIX virtual clients

When using LVM mirroring between disks from two Virtual I/O Servers, a reboot of one Virtual I/O Server makes one disk *missing* and stale partitions have to be synchronized with the **varyonvg** command when the server is rebooted.

When using NIB for network redundancy, the backup does not fall back to the primary adapter until the backup adapter fails. This holds for virtual adapters and AIX V5.3-ML03 or higher even if Automatically Recover to Main Channel is set to yes due to the fact that no link-up event is received during reactivation of the path. This is because virtual Ethernet adapters are always up. If you configure NIB to do load balancing, you may want the NIB to be on the primary channel.

If the settings are correct, MPIO will not require any special attention when a Virtual I/O Server is restarted, but a failed path should be checked.

Checking and fixing these things in a system with many partitions is time consuming and prone to errors.

“Listing of the fixdualvio.ksh script” on page 514 is a sample script that you can be tailored to your needs. The script will check and fix the configuration for:

- ▶ Redundancy with dual Virtual I/O Servers and LVM mirroring
- ▶ Redundancy with dual Virtual I/O Servers, Fibre Channel SAN disks and AIX MPIO
- ▶ Network redundancy using Network Interface Backup

The script should reside on each VIO Client and if using **dsh** (distributed shell), it should also be located in the same directory on each VIO Client. Since it is local to the VIO Client, it can be customized for the individual client. It could also be executed at regular intervals using **cron**.

Distributed shell, **dsh** can be used to run the script on all required target partitions in parallel from a NIM or admin server after a Virtual I/O Server reboot.

**Notes:**

- ▶ The **dsh** command is installed by default in AIX and is part of CSM. However, use of the full function clustering offered by CSM requires a license. See the AIX documentation for **dsh** command information.
- ▶ You can use **dsh** based on **rsh**, **ssh**, or Kerberos authentication as long as **dsh** can run commands without being prompted for a password.

See Example 18-1 for information about how to run fixdualvios.ksh in parallel on partitions dbserver, appserver, and nim.

*Example 18-1 Using a script to update partitions*

---

```
# dsh -n dbserver,appserver,nim /tmp/fixdualvios.ksh | dshbak >\
/tmp/fixdualvio.out
```

---

**Tips:**

- ▶ Use the **DSH\_LIST=<file listing lpars>** variable so you do not have to type in the names of the target LPARs when using **dsh**.
- ▶ Use the **DSH\_REMOTE\_CMD=/usr/bin/ssh** variable if you use **ssh** for authentication.
- ▶ The output file **/tmp/fixdualvio.out** will reside on the system running the **dsh** command.

The **dshbak** command will group the output from each server.

Example 18-2 shows how to run the script and the output listing from the sample partitions named dbserver, appserver, and nim.

*Example 18-2 Running the script and listing output*

---

```
# export DSH_REMOTE_CMD=/usr/bin/ssh
# export DSH_LIST=/root/nodes
# dsh /tmp/fixdualvios.ksh|dshbak > /tmp/fixdualvios.out
```

HOST: appserver  
-----

1 Checking if Redundancy with dual VIO Server and LVM mirroring is being used.  
Redundancy with dual VIO Server and LVM mirroring is NOT used.  
No disk has missing status in any volume group.

2 Checking if Redundancy with dual VIO Server, Fibre Channel SAN disks and AIX MPIO is being used.  
Status:  
Enabled hdisk0 vscsi0  
Enabled hdisk0 vscsi1  
**hdisk1 has vscsi0 with Failed status. Enabling path.**  
paths Changed  
New status:  
Enabled hdisk1 vscsi0  
Enabled hdisk1 vscsi1

3 Checking if Network redundancy using Network interface backup is being used.  
EtherChannel en2 is found.  
**Backup channel is being used. Switching back to primary.**  
Active channel: primary adapter

HOST: dbserver  
-----

1 Checking if Redundancy with dual VIO Server and LVM mirroring is being used.  
Redundancy with dual VIO Server and LVM mirroring is NOT used.  
No disk has missing status in any volume group.

2 Checking if Redundancy with dual VIO Server, Fibre Channel SAN disks and AIX MPIO is being used.  
**hdisk0 has vscsi0 with Failed status. Enabling path.**  
paths Changed  
New status:

```
Enabled hdisk0 vscsi0
Enabled hdisk0 vscsi1
```

```
3 Checking if Network redundancy using Network interface backup is
being used.
```

```
EtherChannel en2 is found.
```

```
Backup channel is being used. Switching back to primary.
```

```
Active channel: primary adapter
```

```
HOST: nim
```

```
-----
```

```
1 Checking if Redundancy with dual VIO Server and LVM mirroring is
being used.
```

```
Redundancy with dual VIO Server and LVM mirroring is being used.
```

```
Checking status.
```

```
No disk in rootvg has missing status.
```

```
No disk has missing status in any volume group.
```

```
2 Checking if Redundancy with dual VIO Server, Fibre Channel SAN disks
and AIX MPIO is being used.
```

```
Redundancy with dual VIO Server, Fibre Channel SAN disks and AIX MPIO
is NOT used.
```

```
3 Checking if Network redundancy using Network interface backup is
being used.
```

```
EtherChannel en2 is found.
```

```
Backup channel is being used. Switching back to primary.
```

```
Active channel: primary adapter
```

---

**Note:** The reason for the Failed status of the paths is that the `hcheck_interval` parameter had not been set on the disks yet.

This script assumes that the configuration may be using one or more of the following:

1. Redundancy with dual VIO Server and LVM mirroring
2. Redundancy with dual VIO Server, Fibre Channel SAN disks and AIX MPIO
3. Network redundancy using "Network interface backup"

### Listing of the `fixdualvio.ksh` script

```
#!/bin/ksh
#set -x
#
# This script will check and restore the dual VIO Server
```

```

# configuration for partitions served from two VIO Servers after
# one VIO Server has been unavailable.
# The script must be tailored and TESTED to your needs.
#
# Disclaimer
# IBM DOES NOT WARRANT OR REPRESENT THAT THE CODE PROVIDED IS COMPLETE OR UP-TO-DATE.
# IBM DOES NOT WARRANT, REPRESENT OR IMPLY RELIABILITY, SERVICEABILITY OR FUNCTION OF THE
# CODE. IBM IS UNDER NO OBLIGATION TO UPDATE CONTENT NOR PROVIDE FURTHER SUPPORT.

# ALL CODE IS PROVIDED "AS IS," WITH NO WARRANTIES OR GUARANTEES WHATSOEVER. IBM
# EXPRESSLY DISCLAIMS TO THE FULLEST EXTENT PERMITTED BY LAW ALL EXPRESS, IMPLIED,
# STATUTORY AND OTHER WARRANTIES, GUARANTEES, OR REPRESENTATIONS, INCLUDING, WITHOUT
# LIMITATION, THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE, AND
# NON-INFRINGEMENT OF PROPRIETARY AND INTELLECTUAL PROPERTY RIGHTS. YOU UNDERSTAND AND
# AGREE THAT YOU USE THESE MATERIALS, INFORMATION, PRODUCTS, SOFTWARE, PROGRAMS, AND
# SERVICES, AT YOUR OWN DISCRETION AND RISK AND THAT YOU WILL BE SOLELY RESPONSIBLE FOR
# ANY DAMAGES THAT MAY RESULT, INCLUDING LOSS OF DATA OR DAMAGE TO YOUR COMPUTER SYSTEM.

# IN NO EVENT WILL IBM BE LIABLE TO ANY PARTY FOR ANY DIRECT, INDIRECT, INCIDENTAL,
# SPECIAL, EXEMPLARY OR CONSEQUENTIAL DAMAGES OF ANY TYPE WHATSOEVER RELATED TO OR ARISING
# FROM USE OF THE CODE FOUND HEREIN, WITHOUT LIMITATION, ANY LOST PROFITS, BUSINESS
# INTERRUPTION, LOST SAVINGS, LOSS OF PROGRAMS OR OTHER DATA, EVEN IF IBM IS EXPRESSLY
# ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. THIS EXCLUSION AND WAIVER OF LIABILITY
# APPLIES TO ALL CAUSES OF ACTION, WHETHER BASED ON CONTRACT, WARRANTY, TORT OR ANY OTHER
# LEGAL THEORIES.
#
# Assuming that the configuration may be using one or more of:
# 1 Redundancy with dual VIO Server and LVM mirroring.
# 2 Redundancy with dual VIO Server, Fiber Channel SAN disks and
# AIX MPIO.
# 3 Network redundancy using "Network interface backup".
#
# Syntax: fixdualvio.ksh
#
#
# 1 Redundancy with dual VIO Server and LVM mirroring.
#
echo 1 Checking if "Redundancy with dual VIO Server and LVM mirroring" is being used.

# Check if / (hd4) has 2 copies
MIRROR=`lslv hd4|grep COPIES|awk '{print $2}'`
if [ $MIRROR -gt 1 ]
then
    # rootvg is most likely mirrored
    echo "Redundancy with dual VIO Server and LVM mirroring" is being used.
    echo Checking status.
    # Find disk in rootvg with missing status
    MISSING=`lsvg -p rootvg|grep missing|awk '{print $1}'`
    if [ "$MISSING" = "" ]
    then
        echo No disk in rootvg has missing status.
    else

```

```

        echo $MISSING has missing status.
#
# Restore active status and sync of rootvg
    echo Fixing rootvg.
    varyonvg rootvg
    syncvg -v rootvg

    fi
else
    echo "Redundancy with dual VIO Server and LVM mirroring" is NOT used.
fi
# Check now if ANY disk has missing status.
ANYMISSING=~lsvg -o|lsvg -ip|grep missing|awk '{print $1}'~
if [ "$ANYMISSING" = "" ]
then
    echo No disk has missing status in any volume group.
else
    echo $ANYMISSING has missing status. CHECK CAUSE!
fi

# 2 Redundancy with dual VIO Server, Fiber Channel SAN disks and
#   AIX MPIO.
echo
echo 2 Checking if "Redundancy with dual VIO Server, Fiber Channel SAN disks and AIX
MPIO" is being used.
# Check if any of the disks have more than one path (listed twice)
MPIO=~lspath | awk '{print $2}' | uniq -d`
if [ $MPIO ]
then
    for n in $MPIO
    do
        # Check if this disk has a Failed path.
        STATUS=~lspath -l $n | grep Failed | awk '{print $1}'~
        if [ $STATUS ]
        then
            ADAPTER=~lspath -l $n | grep Failed | awk '{print $3}'~
            echo $n has $ADAPTER with Failed status. Enabling path.
            chpath -s ena -l $n -p $ADAPTER
            # Check new status
            echo New status:
            lspath -l $n
        else
            echo Status:
            lspath -l $n
        fi
    done
else
    echo "Redundancy with dual VIO Server, Fiber Channel SAN disks and AIX MPIO
    "is NOT used.
fi

```



```

# 3 Network redundancy using "Network interface backup".
# Find out if this is being used and if so which interface number(s).

echo
echo 3 Checking if Network redundancy using "Network interface backup" is being used.

ECH=`lsdev -Cc adapter -s pseudo -t ibm_ech -F name | awk -F "ent" '{print $2}'`

if [ -z "$ECH" ]
then
echo No EtherChannel is defined.
else
# What is the status
  for i in $ECH
  do
    echo EtherChannel en$i is found.

    ETHCHSTATUS=`entstat -d en$i | grep Active | awk '{print $3}'`
    if [ "$ETHCHSTATUS" = "backup" ]
    then
      # switch back to primary (requires AIX5.3-ML02 or higher)
      echo Backup channel is being used. Switching back to primary.

      /usr/lib/methods/ethchan_config -f en$i

      # Check the new status
      NEWSTATUS=`entstat -d en$i | grep Active | awk '{print $3}'`
      echo Active channel: $NEWSTATUS adapter
      #
    else
      echo Active channel: $ETHCHSTATUS adapter.
    fi
  done

fi
exit
end

```



# Abbreviations and acronyms

<b>ABI</b>	Application Binary Interface	<b>CHRP</b>	Common Hardware Reference Platform
<b>AC</b>	Alternating Current	<b>CLI</b>	Command Line Interface
<b>ACL</b>	Access Control List	<b>CLVM</b>	Concurrent LVM
<b>AFPA</b>	Adaptive Fast Path Architecture	<b>CPU</b>	Central Processing Unit
<b>AIO</b>	Asynchronous I/O	<b>CRC</b>	Cyclic Redundancy Check
<b>AIX</b>	Advanced Interactive Executive	<b>CSM</b>	Cluster Systems Management
<b>APAR</b>	Authorized Program Analysis Report	<b>CUoD</b>	Capacity Upgrade on Demand
<b>API</b>	Application Programming Interface	<b>DCM</b>	Dual Chip Module
<b>ARP</b>	Address Resolution Protocol	<b>DES</b>	Data Encryption Standard
<b>ASMI</b>	Advanced System Management Interface	<b>DGD</b>	Dead Gateway Detection
<b>BFF</b>	Backup File Format	<b>DHCP</b>	Dynamic Host Configuration Protocol
<b>BIND</b>	Berkeley Internet Name Domain	<b>DLPAR</b>	Dynamic LPAR
<b>BIST</b>	Built-In Self-Test	<b>DMA</b>	Direct Memory Access
<b>BLV</b>	Boot Logical Volume	<b>DNS</b>	Domain Naming System
<b>BOOTP</b>	Boot Protocol	<b>DRM</b>	Dynamic Reconfiguration Manager
<b>BOS</b>	Base Operating System	<b>DR</b>	Dynamic Reconfiguration
<b>BSD</b>	Berkeley Software Distribution	<b>DVD</b>	Digital Versatile Disk
<b>CA</b>	Certificate Authority	<b>EC</b>	EtherChannel
<b>CATE</b>	Certified Advanced Technical Expert	<b>ECC</b>	Error Checking and Correcting
<b>CD</b>	Compact Disk	<b>EOF</b>	End of File
<b>CDE</b>	Common Desktop Environment	<b>EPOW</b>	Environmental and Power Warning
<b>CD-R</b>	CD Recordable	<b>ERRM</b>	Event Response resource manager
<b>CD-ROM</b>	Compact Disk-Read Only Memory	<b>ESS</b>	Enterprise Storage Server®
<b>CEC</b>	Central Electronics Complex	<b>F/C</b>	Feature Code
		<b>FC</b>	Fibre Channel
		<b>FC_AL</b>	Fibre Channel Arbitrated Loop

<b>FDX</b>	Full Duplex	<b>LA</b>	Link Aggregation
<b>FLOP</b>	Floating Point Operation	<b>LACP</b>	Link Aggregation Control Protocol
<b>FRU</b>	Field Replaceable Unit	<b>LAN</b>	Local Area Network
<b>FTP</b>	File Transfer Protocol	<b>LDAP</b>	Lightweight Directory Access Protocol
<b>GDPS®</b>	Geographically Dispersed Parallel Sysplex™	<b>LED</b>	Light Emitting Diode
<b>GID</b>	Group ID	<b>LMB</b>	Logical Memory Block
<b>GPFS</b>	General Parallel File System™	<b>LPAR</b>	Logical Partition
<b>GUI</b>	Graphical User Interface	<b>LPP</b>	Licensed Program Product
<b>HACMP™</b>	High Availability Cluster Multiprocessing	<b>LUN</b>	Logical Unit Number
<b>HBA</b>	Host Bus Adapter	<b>LV</b>	Logical Volume
<b>HMC</b>	Hardware Management Console	<b>LVCB</b>	Logical Volume Control Block
<b>HTML</b>	Hypertext Markup Language	<b>LVM</b>	Logical Volume Manager
<b>HTTP</b>	Hypertext Transfer Protocol	<b>MAC</b>	Media Access Control
<b>Hz</b>	Hertz	<b>Mbps</b>	Megabits Per Second
<b>I/O</b>	Input/Output	<b>MBps</b>	Megabytes Per Second
<b>IBM</b>	International Business Machines	<b>MCM</b>	Multichip Module
<b>ID</b>	Identification	<b>ML</b>	Maintenance Level
<b>IDE</b>	Integrated Device Electronics	<b>MP</b>	Multiprocessor
<b>IEEE</b>	Institute of Electrical and Electronics Engineers	<b>MPIO</b>	Multipath I/O
<b>IP</b>	Internetwork Protocol	<b>MTU</b>	Maximum Transmission Unit
<b>IPAT</b>	IP Address Takeover	<b>NFS</b>	Network File System
<b>IPL</b>	Initial Program Load	<b>NIB</b>	Network Interface Backup
<b>IPMP</b>	IP Multipathing	<b>NIM</b>	Network Installation Management
<b>ISV</b>	Independent Software Vendor	<b>NIMOL</b>	NIM on Linux
<b>ITSO</b>	International Technical Support Organization	<b>N_PORT</b>	Node Port
<b>IVM</b>	Integrated Virtualization Manager	<b>NPIV</b>	N_Port Identifier Virtualization
<b>JFS</b>	Journaled File System	<b>NVRAM</b>	Non-Volatile Random Access Memory
<b>L1</b>	Level 1	<b>ODM</b>	Object Data Manager
<b>L2</b>	Level 2	<b>OS</b>	Operating System
<b>L3</b>	Level 3	<b>OSPF</b>	Open Shortest Path First
		<b>PCI</b>	Peripheral Component Interconnect

<b>PCI-e</b>	iPeriphera Component Interconnect Express	<b>RPM</b>	Red Hat Package Manager
<b>PIC</b>	Pool Idle Count	<b>RSA</b>	Rivet, Shamir, Adelman
<b>PID</b>	Process ID	<b>RSCT</b>	Reliable Scalable Cluster Technology
<b>PKI</b>	Public Key Infrastructure	<b>RSH</b>	Remote Shell
<b>PLM</b>	Partition Load Manager	<b>SAN</b>	Storage Area Network
<b>POST</b>	Power-On Self-test	<b>SCSI</b>	Small Computer System Interface
<b>POWER</b>	Performance Optimization with Enhanced Risc (Architecture)	<b>SDD</b>	Subsystem Device Driver
<b>PPC</b>	Physical Processor Consumption	<b>SMIT</b>	System Management Interface Tool
<b>PPFC</b>	Physical Processor Fraction Consumed	<b>SMP</b>	Symmetric Multiprocessor
<b>PTF</b>	Program Temporary Fix	<b>SMS</b>	System Management Services
<b>PTX®</b>	Performance Toolbox	<b>SMT</b>	Simultaneous Multithreading
<b>PURR</b>	Processor Utilization Resource Register	<b>SP</b>	Service Processor
<b>PV</b>	Physical Volume	<b>SPOT</b>	Shared Product Object Tree
<b>PVID</b>	Physical Volume Identifier	<b>SRC</b>	System Resource Controller
<b>PVID</b>	Port Virtual LAN Identifier	<b>SRN</b>	Service Request Number
<b>QoS</b>	Quality of Service	<b>SSA</b>	Serial Storage Architecture
<b>RAID</b>	Redundant Array of Independent Disks	<b>SSH</b>	Secure Shell
<b>RAM</b>	Random Access Memory	<b>SSL</b>	Secure Socket Layer
<b>RAS</b>	Reliability, Availability, and Serviceability	<b>SUID</b>	Set User ID
<b>RCP</b>	Remote Copy	<b>SVC</b>	SAN Virtualization Controller
<b>RDAC</b>	Redundant Disk Array Controller	<b>TCP/IP</b>	Transmission Control Protocol/Internet Protocol
<b>RIO</b>	Remote I/O	<b>TSA</b>	Tivoli System Automation
<b>RIP</b>	Routing Information Protocol	<b>UDF</b>	Universal Disk Format
<b>RISC</b>	Reduced Instruction-Set Computer	<b>UDID</b>	Universal Disk Identification
<b>RMC</b>	Resource Monitoring and Control	<b>VIPA</b>	Virtual IP Address
<b>RPC</b>	Remote Procedure Call	<b>VG</b>	Volume Group
<b>RPL</b>	Remote Program Loader	<b>VGDA</b>	Volume Group Descriptor Area
		<b>VGSA</b>	Volume Group Status Area
		<b>VLAN</b>	Virtual Local Area Network
		<b>VP</b>	Virtual Processor
		<b>VPD</b>	Vital Product Data

<b>VPN</b>	Virtual Private Network
<b>VRRP</b>	Virtual Router Redundancy Protocol
<b>VSD</b>	Virtual Shared Disk
<b>WLM</b>	Workload Manager
<b>WWN</b>	Worldwide Name
<b>WWPN</b>	Worldwide Port Name

# Related publications

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

## IBM Redbooks

For information about ordering these publications, see “How to get Redbooks” on page 526. Note that some of the documents referenced here may be available in softcopy only.

- ▶ *AIX 5L Practical Performance Tools and Tuning Guide*, SG24-6478
- ▶ *A Practical Guide for Resource Monitoring and Control (RMC)*, SG24-6615
- ▶ *Effective System Management Using the IBM Hardware Management Console for pSeries*, SG24-7038
- ▶ *Hardware Management Console V7 Handbook*, SG24-7491
- ▶ *i5/OS on eServer p5 Models A Guide to Planning, Implementation, and Operation*, SG24-8001
- ▶ *IBM AIX Version 6.1 Differences Guide*, SG24-7559
- ▶ *IBM AIX Continuous Availability Features*, REDP-4367
- ▶ *IBM Director on System p5*, REDP-4219
- ▶ *IBM System i and System p System Planning and Deployment: Simplifying Logical Partitioning*, SG24-7487
- ▶ *IBM i and Midrange External Storage*, SG24-7668
- ▶ *IBM System p Advanced POWER Virtualization Best Practices*, REDP-4194
- ▶ *IBM System p Live Partition Mobility*, SG24-7460
- ▶ *Implementing an IBM/Brocade SAN with 8 Gbps Directors and Switches*, SG24-6116
- ▶ *Implementing High Availability Cluster Multi-Processing (HACMP) Cookbook*, SG24-6769
- ▶ *Integrated Virtual Ethernet Adapter Technical Overview and Introduction*, REDP-4340
- ▶ *Integrated Virtualization Manager on IBM System p5*, REDP-4061
- ▶ *Introduction to pSeries Provisioning*, SG24-6389

- ▶ *Introduction to Workload Partition Management in IBM AIX Version 6.1*, SG24-7431
- ▶ *Linux Applications on pSeries*, SG24-6033
- ▶ *Managing AIX Server Farms*, SG24-6606
- ▶ *NIM from A to Z in AIX 5L*, SG24-7296
- ▶ *Partitioning Implementations for IBM eServer p5 Servers*, SG24-7039
- ▶ *PowerVM Virtualization on IBM System p: Introduction and Configuration*, SG24-7940
- ▶ *Virtualizing an Infrastructure with System p and Linux*, SG24-7499

## Other publications

These publications are also relevant as further information sources:

- ▶ The following types of documentation are located on the Internet at:  
<http://www.ibm.com/systems/p/support/index.html>
  - User guides
  - System management guides
  - Application programmer guides
  - All commands reference volumes
  - Files reference
  - Technical reference volumes used by application programmers
- ▶ Detailed documentation about the PowerVM feature and the Virtual I/O Server is at:  
<https://www14.software.ibm.com/webapp/set2/sas/f/vios/documentation/home.html>

## Online resources

These Web sites are also relevant as further information sources:

- ▶ IBM System p Virtualization—The most complete virtualization offering for UNIX and Linux:  
<http://www-01.ibm.com/cgi-bin/common/ssi/ssialias?infotype=an&subtype=ca&htmlfid=897/ENUS207-269&appname=usn&language=enus>



- ▶ HMC interaction script:  
<http://www.the-welters.com/professional/scripts/hmcMenu.txt>
- ▶ IBM Redbooks:  
<http://www.redbooks.ibm.com/>
- ▶ BM Systems information center—Power Systems Virtual I/O Server and Integrated Virtualization Manager commands:  
<http://publib.boulder.ibm.com/infocenter/systems/scope/hw/topic/iphcg/iphcg.pdf>
- ▶ IBM System Planning Tool:  
<http://www-304.ibm.com/jct01004c/systems/support/tools/systemplanningtool>
- ▶ IBM wikis:  
<http://www.ibm.com/developerworks/wikis/dashboard.action>
  - AIX Wiki - Performance Other Tools:  
<http://www.ibm.com/developerworks/wikis/display/WikiPtype/Performance+Other+Tools>
  - nmon analyzer tool:  
<http://www.ibm.com/developerworks/wikis/display/Wikiptype/nmonanalyzer>
- ▶ Virtual I/O Server monitoring wiki:  
[http://www.ibm.com/developerworks/wikis/display/WikiPtype/VIOS\\_Monitoring](http://www.ibm.com/developerworks/wikis/display/WikiPtype/VIOS_Monitoring)
- ▶ The nmon tool:  
<http://www.ibm.com/developerworks/wikis/display/WikiPtype/nmon>
- ▶ IBM System p and AIX Information Center:  
<http://publib16.boulder.ibm.com/pseries/index.htm>  
Virtual I/O Server 2.1 release notes:  
[http://publib.boulder.ibm.com/infocenter/systems/scope/aix/index.jsp?topic=/com.ibm.aix.resources/61relnotes.htm&tocNode=int\\_188639](http://publib.boulder.ibm.com/infocenter/systems/scope/aix/index.jsp?topic=/com.ibm.aix.resources/61relnotes.htm&tocNode=int_188639)
- ▶ pSeries and AIX information center—Installing and configuring the system for Kerberos integrated login using KRB5:  
[http://publib.boulder.ibm.com/infocenter/pseries/v5r3/index.jsp?topic=/com.ibm.aix.security/doc/security/kerberos\\_auth\\_only\\_load\\_module.htm](http://publib.boulder.ibm.com/infocenter/pseries/v5r3/index.jsp?topic=/com.ibm.aix.security/doc/security/kerberos_auth_only_load_module.htm)

- ▶ Advanced Power Virtualization:  
<http://www-03.ibm.com/systems/power/software/virtualization/index.html>
- ▶ Tivoli software information center:  
<http://publib.boulder.ibm.com/tividd/td/IdentityManager5.0.html>
- ▶ EnergyScale for IBM POWER6 microprocessor-based systems:  
<http://www.research.ibm.com/journal/rd/516/mccreary.html>
- ▶ System power management support in the IBM POWER6 microprocessor:  
<http://www.research.ibm.com/journal/rd/516/floyd.html>
- ▶ IBM i 6.1 Information Center:  
<http://publib.boulder.ibm.com/infocenter/systems/scope/i5os/index.jsp>
- ▶ IBM i Software Knowledge Base: Cross-Referencing (Device Mapping) IBM i Disks with VIOS Disks with IVM  
[http://www-912.ibm.com/s\\_dir/slkbases.nsf/1ac66549a21402188625680b0002037e/23f1e308b41e40a486257408005aea5b?OpenDocument&Highlight=2,481468986](http://www-912.ibm.com/s_dir/slkbases.nsf/1ac66549a21402188625680b0002037e/23f1e308b41e40a486257408005aea5b?OpenDocument&Highlight=2,481468986)

## How to get Redbooks

You can search for, view, or download Redbooks, Redpapers, Technotes, draft publications and Additional materials, as well as order hardcopy Redbooks, at this Web site:

[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

- Active energy manager 359
- adapter
  - adding dynamically 248, 251, 254, 256, 265, 269
  - create 267
- ADDTCPIFC command 118
- Advanced System Management Interface, ASMI 296
- AIX 422
  - automating the health checking and recovery 465
  - checking network health 229
  - crontab file 432
  - enabling jumbo frames 132
  - failed path 463
  - largesend option 135
  - LVM mirroring environment 463
  - maximum segment size, MSS 128
  - monitoring storage performance 465
  - MPIO environment 462
  - network address change 117
  - partitions allocations 397
  - restore 217
  - resynchronize LVM mirroring 236
  - stale partitions 464
  - storage monitoring 461
  - virtual Ethernet tracing 120
  - xmwlmd daemon 441

## B

- backup
  - additional information 202
  - client logical partition 190
  - DVD 194
  - external device configuration 189
  - HMC 189
  - IVM 190
  - mksysb 196, 198
  - nim\_resources.tar 196–197
  - scheduling 192
  - SEA 201
  - tape 193

- VIOS 191
  - Virtual I/O Server operating system 193
- backupios command 173, 191, 194–195, 197–198, 219, 318
- Barrier Synchronization Register 297
- bkprofdata command 190
- bosboot command 46

## C

- cfgassist command 147
- cfgdev command 22, 49, 262
- cfgmgr command 22–23, 262
- cfgnamesrv command 197, 202
- CFGPFRCOL command 468
- cfgsvc command 371
- CFGTCP command 118
- chdev command 55, 132, 238, 319
- CHGLINETH command 133
- CHGTCPA command 131
- CHGTCPDMN command 118
- chsp command 49
- chtcpip command 116
- chuser command 167
- cleargcl command 168
- Cluster Systems Management 360
- command installios 208
- commands
  - AIX
    - bosboot 46
    - cfgdev 238
    - cfgmgr 22–23
    - chdev 132, 319
    - cron 512
    - dsh 23, 363, 480, 512
    - dshbak 23, 512
    - entstat 480, 497
    - errpt 227
    - ethchan\_config 480
    - fget\_config 302
    - ifconfig 136
    - iostat 465, 504
    - ldapsearch 160, 162
    - lparmon 509

- lparstat 397, 415, 424, 429, 431
- lsattr 302, 462
- lsattr -El mem0 254
- lscfg 22, 120
- lslpp command 161
- lspath 232, 462
- lsvg 42, 227, 233, 464
- mirrorvg 46
- mktcpip 117
- mkvdev 22
- mpio\_get\_config 88
- mpstat 412, 414–415, 434
- netstat 130
- nmon 425
- no 128
- pmtu 131
- reducevg 44
- rmdev 260
- rsh 23
- sar 412, 414, 431, 442
- ssh 23
- syncvg 236
- tcptr 141
- topas 422, 435
- topasout 435, 441
- topasrec 8
- unmirrorvg 43
- varyonvg 511
- vmstat 428
- xmwlms 435
- HMC
  - defsysplanres 313
  - installios 207
  - lshmc 295
  - lslic 295
  - mkauthkeys 295, 372
  - mksysplan 317
  - OS\_install 313
  - ssh 323
- IBM i
  - ADDTCPIFC 118
  - CFGPFRCOL 468
  - CFGTCP 118
  - CHGLINETH 133
  - CHGTCPA 131
  - CHGTCPDMN 118
  - CRTPFRTA 455
  - ENDPFRCOL 455, 468
  - ENDTCPIFC 118, 133
  - PRTACTRPT 444
  - PRTCPTTRPT 444, 455, 500
  - PRTRSCRPT 469
  - PRTSYSRPT 444, 469, 500
  - RMVTCPIFC 118
  - STRPFRCOL 455, 468
  - STRPFRT 444, 455, 469
  - STRSST 42, 233, 466
  - STRTCPIFC 118, 133
  - VRYCFG 19, 133
  - WRKCFGSTS 19, 133
  - WRKDSKSTS 467
  - WRKHDWRSC 16, 122, 499
  - WRKSYSACT 442
  - WRKSYSSTS 453, 466
- Linux
  - lparcfg 281–282, 398
  - lscfg 277
  - lsvio 278
  - lsvpd 278
  - mdadm 237
  - meminfo 284
  - mpstat 434
  - multipath 227, 232
  - nmon 425
  - update-lsvpd-db 277
- NIM
  - installios 209
- other
  - definehwdev 362
  - gsk7ikm 150
  - mkldap 161
  - mksecldap 156
  - mksysplan 218
  - scp 26
  - updatehwdev 362
- SAN switch
  - portCfgNPIVPort 64, 75
  - portcfgshow 64, 75
  - version 63
  - zonestow 71, 85
- TSM
  - dsmc 204–205, 214–215
- VIOS
  - backupios 173, 191, 194–195, 197–198, 219, 318
  - bkprofddata 190
  - cfgassist 147
  - cfgdev 22, 49, 262

cfgnamesrv 197, 202  
 cfigsvc 371  
 chdev 55, 238  
 chsp 49  
 chtcpip 116  
 chuser 167  
 cleargcl 168  
 crontab 191–192, 325  
 deactivatevg 44  
 diagmenu 42, 288  
 dsmc 204, 214  
 enstat 485  
 entstat 202, 404, 477, 483  
 errlog 478  
 exportvg 44  
 fcstat 403  
 fget\_config 31, 403  
 genfilt 147  
 hostmap 197, 203  
 hostname 404  
 installios 173, 207–208  
 ioslevel 25, 186, 236, 295, 402  
 loadopt 27  
 lscfg 404  
 lsdev 22, 126, 220, 259, 403  
 lsfailedlogin 168  
 lsfware 402  
 lsgcl 168–169, 402  
 lsiparinfo 402  
 lslv 46, 403  
 lsmap 35, 39, 70, 83, 211, 220, 296, 319, 403  
 lsnetvc 404  
 lsports 69, 83  
 lspath 236, 403  
 lspv 220, 403  
 lsrep 25–26, 403  
 lssdd 31  
 lsslot 259  
 lssp 26, 403  
 lssvc 371, 402  
 lssw 402  
 lstcpip 404  
 lsuser 167, 404  
 lsvg 220, 403  
 lsvopt 403  
 mkbdsp 50  
 mkkrb5clnt 165  
 mklv 45  
 mknfsexp 197  
 mkrep 26  
 mksp 50  
 mktcpip 116, 226, 483  
 mkuser 161, 167  
 mkvdev 22, 46, 223, 226, 318, 482–483  
 mkgv 45  
 mkvopt 26  
 mount 197  
 mpio\_get\_config 403, 460  
 mpstat 412, 414, 424  
 netstat 203, 219, 224, 404  
 nmon 425  
 oem\_platform\_level 402  
 oem\_setup\_env 31, 168  
 optimizenet 203, 404  
 pcmpath 403  
 replphyvol 45  
 restorevgstruct 216  
 rmbdsp 48  
 rmdev 260  
 rmuser 167  
 savevgstruct 200, 216  
 seastat 492, 495  
 showmount 404  
 shutdown 168, 228  
 snmp\_info 404  
 startnetvc 215  
 startsvc 382  
 stopnetvc 144  
 stopsvc 382  
 su 167  
 svmon 402  
 sysstat 402  
 topas 402, 416, 419–420, 424, 452  
 topasout 435  
 topasrec 8  
 traceroute 404  
 unloadopt 28  
 updateios 228  
 vasistat 402  
 vfcmap 61, 70  
 viosecure 140, 146, 203  
 viostat 403, 460  
 vmstat 402, 428  
 wkldout 402  
 CPU metrics 411  
 CPU monitoring  
   AIX 422

- cross-partition 416
- donated processors 418
- IBM i 442
- multiple shared processor pools, MSPP 420
- system-wide tools 414
- variable processor frequency 415, 430
- CPU utilization
  - IBM i 442
  - report generation 435
- cron command 512
- crontab command 191–192, 325
- CRTPFRDTA command 455
- CSM 360

## D

- deactivatevg command 44
- dedicated processor 408
- definehwdev command 362
- defsysplanres command 313
- diagmenu command 42, 288
- disk reserve policy 302
- distributed shell 465
- DLPAR
  - adapters
    - add adapters dynamically 256, 265
    - adding dynamically 258
    - move adapters 258
  - cfgdev command 262
  - HMC 256, 265
  - Linux 279
  - memory
    - add memory dynamically in AIX 251
    - add memory dynamically in IBM i 251
    - add memory dynamically in Linux 284
    - removing memory dynamically 254
  - operations on AIX and IBM i 248
  - processors
    - adding and removing processors 248
    - adding CPU 248, 280
  - rmdev command 260
  - virtual adapters remove 269
  - virtual processors change 250
- DLPAR cfgmgr 262
- DoS Hardening 149
- dsh command 23, 363, 480, 512
- DSH\_LIST variable 23, 512
- DSH\_REMOTE\_CMD variable 23, 512
- dshbak command 23, 512

- dsmc command 204–205, 214–215

## E

- ENDPFRCOL command 455, 468
- ENDTCPIFC command 118, 133
- EnergyScale 359, 415
  - variable processor frequency 415, 430
- entitlement 408
  - computation 413
  - consumption 408
- entstat command 202, 404, 477, 480, 483, 485
- entstat commands 497
- errlog command 478
- errpt command 227
- ethchan\_config command 480
- EtherChannel 479
- Etherchannel 482
- Ethernet
  - maximum transfer unit, MTU size 127
- Ethernet adapter
  - replacing 287
  - subnet mask 225
- expansion pack
  - VIOS 164
- exportvg command 44

## F

- fcstat command 403
- fget\_config command 31, 302, 403
- Fibre Channel adapter
  - replace 92, 290
- firewall
  - disable 146
  - rules 146
  - setup 144
- firmware 295
- fixdualvio.ksh script 514
- fixdualvios.sh script 512

## G

- Ganglia 508
- genfilt command 147
- Global Shared Processor Pool 406, 409
- GPFS cluster 55
- gsk7ikm command 150

**H**

Hardware Management Console, HMC 207, 360  
 allow performance information collection 425  
 backup 189  
 hardware information 391  
 mksysplan 218  
 monitoring 390  
 naming conventions 31  
 processor sharing option 418  
 restore 206  
 shell scripting 394  
 virtual device slot numbers 32  
 virtual network monitoring 393  
 HBA and Virtual I/O Server failover 90  
 hcheck\_interval parameter 319  
 HMC  
 backup 189  
 System Plan 219  
 virtual network management 120  
 VLAN tag 124  
 hmcMenu script 394  
 Host bus adapter failover 88  
 hostmap command 197, 203  
 hostname command 404  
 Hot Plug PCI adapter 271  
 Hot Plug Task 289

**I**

IBM Director 359–360  
 Active energy manager 359  
 IBM i 454  
 6B22 devices 36  
 ASP threshold 466  
 change IP address 117  
 checking network health 230  
 checking storage health 466  
 client virtual Ethernet adapter 125  
 Collection Services 455, 468  
 component report for component interval activity 444  
 component report for storage pool activity 455  
 component report for TCP/IP activity 500  
 configuration tracing 36  
 CPU performance guideline 445  
 cross-partition memory monitoring 456  
 cross-partition monitoring 448, 471  
 cross-partition network monitoring 501  
 disk response time 470

disk service time 470  
 disk unit details 37  
 disk unit missing from the configuration 467  
 disk wait time 470  
 dispatched CPU time 448  
 display disk configuration 466  
 display disk configuration capacity 466  
 display disk configuration status 233, 466  
 enable jumbo frames 133  
 Ethernet line description 498  
 independent ASP 466  
 investigate performance data 456  
 job run priority 445  
 line description 133  
 load source unit 38  
 Management Central monitors 448, 471, 501  
 maximum frame size 134  
 memory monitoring 453  
 mirror resynchronization 236  
 mirroring across two VIOS 467  
 monitoring storage performance 467  
 MTU size 129  
 network health checking 497  
 network performance monitoring 500  
 not connected disk unit 467  
 overflow into the system ASP 466  
 packet errors 501  
 page fault 456, 470  
 performance rule for page faults 454  
 Performance Tools for i5/OS 455, 468–469, 500  
 QAPMDISK database file 468  
 reset IOP 16  
 resource report for disk utilization 469  
 restore 217  
 resume mirrored protection 467  
 sector conversion 46  
 SST 42  
 storage monitoring 466  
 storage pool 454  
 suspended disk unit 467  
 suspended mirrored disk units 236  
 system report for disk utilization 469  
 system report for resource utilization expansion 445  
 system report for TCP/IP summary 500  
 System Service Tools, SST 42  
 Systems Director Navigator for i5/OS 447, 456, 470

- TCP/IP interface status 497
- uncapped partition 443
- used storage capacity 466
- virtual Ethernet adapter resource 499
- virtual Ethernet tracing 122
- virtual IOP 16, 499
- virtual IP address, VIPA 478
- virtual optical device 16
- VSCSI adapter slots 38
- waits overview 447
- work with communication resources 499
- work with configuration status 498
- work with disk unit recovery 467
- work with disk units 466
- work with system activity 443
- Work with TCP/IP Interface Status 134
- work with TCP/IP interface status 498
- IBM Installation Toolkit for Linux for Power 276
- IBM Performance Tools for i5/OS 444, 455, 468–469, 500
- IBM Systems Director Navigator for i5/OS 447, 456, 470
- IBM Tivoli Monitoring agent 370
- IBM Tivoli Monitoring, ITM 370
- ICMP 130
- IEEE 802.3ad mode 482
- ifconfig command 136
- installios command 173, 207, 209
- INSTALLIOS\_PRIVATE\_IF variable 173
- Integrated Virtualization Manager, IVM
  - backup 190
  - monitoring 395
- interim fixes 227, 235
- Internet Control Message Protocol, ICMP 130
- ioslevel command 25, 186, 236, 295, 402
- iostat command 465, 504
- ITM
  - agent configuration 371
  - CPU utilization 377
  - data collection 371
  - health 371
  - network adapter utilization 379
  - network mappings 375
  - performance 371
  - system storage information 378
  - topology 370
- ITUAM, Tivoli Usage and Accounting Manager 381
- IVM
  - backup 190
- J**
  - jumbo frames 127, 132, 134
- K**
  - kerberos 164
- L**
  - largesend option for TCP 135
  - ldapsearch command 160, 162
  - librtas tool 273
  - licensing
    - PowerVM
      - Enterprise edition 295, 303
      - Live Partition Mobility 295, 303
  - Link aggregation 482
  - Link Aggregation Control Protocol, LACP 485
  - Linux
    - add memory dynamically 284
    - add processors dynamically 279–280
    - additional packages 279
    - bounding device configuration 481
    - check mirror sync status 237
    - cross partition monitoring 453
    - disk re-scan 287
    - DLPAR 279
    - IBM Installation Toolkit 275
    - librtas tool 273
    - Linux for Power 272
    - lparcfg 282
    - messages 281
    - monitoring 503
    - partition allocations 398
    - removal of processors 283
    - re-scan disk 287
    - RSCT 272
    - Service & Productivity tools 273
      - download 276
      - virtual processor 281
    - Live Application Mobility 302
    - Live Partition Mobility 52, 293–294
      - migration 298
      - requirements 294
      - validation 298
    - loadopt command 27
    - logical processor 408
    - logical CPU 408
    - logical memory block 296
    - logical processor 408



- utilization 414
- loose mode QoS 138
- lparcfg command 281–282, 398
- lparmon command 509
- lparstat command 397, 415, 424, 429, 431
- lsattr command 254, 302, 462
- lscfg command 22, 120, 277, 404
- lsdev command 22, 126, 220, 259, 403
- lsfailedlogin command 168
- lsfware command 402
- lsgcl command 168–169, 402
- lshmc command 295
- lslic command 295
- lsparinfo command 402
- lspp command 161
- lsiv command 46, 403
- lsmapi command 35, 39, 70, 83, 211, 220, 296, 319, 403
- lsmapi commands 220
- lsnetvc command 404
- lsnports command 69, 83
- lspath command 232, 236, 403, 462
- lspv command 220, 403
- lsrep command 25–26, 403
- lssdd command 31
- lsslot command 259
- lssp command 26, 403
- lssvc command 371, 402
- lssw command 402
- lstcpip command 404
- lsuser command 167, 404
- lsvg command 42, 220, 227, 233, 403, 464
- lsvio command 278
- lsvopt command 403
- lsvpd command 278

## M

- MAC address 208
- mdadm command 237
- meminfo command 284
- memory
  - cross-partition monitoring 452
  - huge pages 297
  - monitoring 451
    - cross-partiton 452
    - for IBM i 453
- mirrorios command 168
- mirrorvg command 46

- mkauthkeys command 295, 372
- mkbdspl command 50
- mkkrb5clnt command 165
- mkldap command 161
- mkiv command 45
- mknfsexp command 197
- mkrep command 26
- mkseclap command 156
- mksp command 50
- mksysb command 196, 198
- mksysplan command 218, 317
- mktcpip command 116, 226, 483
- mkuser command 161, 167
- mkvdev command 22, 46, 223, 226, 318, 482–483
- mkvg command 45
- mkvopt command 26
- monitoring tools 366
- mount command 197
- mover service partition 296, 301
- MPIO
  - checking storage health 232
  - healthcheck 232
  - healthcheck interval 462
  - healthcheck mode 462
- mpio\_get\_config command 88, 403, 460
- mpstat command 412, 414–415, 424, 434
- MTU discovery 127
- MTU size 119, 127
- multipath command 227, 232
- Multiple Shared Processor Pools, MSPP 244
  - default 421
  - definition 409
  - maximum 409
  - monitoring 420
  - reserve 409

## N

- N\_Port ID Virtualization, NPIV 21, 294
- netstat command 130, 203, 219, 224, 404, 203
- Network Interface Backup, NIB
  - backup adapter 480
  - testing 478
- network monitoring
  - AIX 497
  - IBM i 497
  - VIOS 476
- network security 144

NFS 196, 207  
 NIB 319  
 NIM 199, 209  
   create SPOT resource 209  
   mksysb resource 209  
 nim\_resources.tar  
   backup 196–197  
   restore 207  
 NIMOL 208  
 nmon analyser 441, 503  
 nmon command 425  
 nmon tool 504  
   recording 507  
 no command 128  
 no\_reserve parameter 302, 318  
 NPIV 56  
   configuring  
     existing AIX LPAR 74  
     new AIX LPAR 62  
   considerations 91  
   heterogeneous configuration 107  
   introduction 56  
   migration 93  
     physical to NPIV 94  
     virtual SCSI to NPIV 94  
   requirements 59

**O**

oem\_platform\_level command 402  
 oem\_setup\_env command 31, 168  
 optimizenet command 203, 404  
 OS\_install command 313

**P**

page faults 454  
 parameter  
   hcheck\_interval 319  
   no\_reserve 302, 318  
   reserve\_policy 302  
   single\_path 318  
 partitions  
   allocations 397  
   processor sharing 418  
   properties 391  
 Path MTU discovery 127, 129, 135  
 pcmpath command 403  
 performance measurements 411  
 pmtu command 131

portCfgNPIVPort command 64, 75  
 portcfgshow command 64, 75  
 POWER6 terminologies 410  
 processing unit 407–408  
 Processor Utilization of Resources Register, PURR  
 411  
 processors adding 248  
 processors removing 283  
 PRTACTRPT command 444  
 PRTCPTTRPT command 444, 455, 500  
 PRTRSCRPT command 469  
 PRTSYSRPT command 444, 469, 500  
 PURR 411, 431  
   metrics 412

**Q**

QoS 137

**R**

rebuild VIOS 218  
 Redbooks Web site 526  
   Contact us xxx  
 reducevg command 44  
 redundant error path reporting 297  
 Remote Command Execution 323  
 remote login 144  
 replphyvol command 45  
 report generation 435  
 reserve\_policy parameter 302  
 resource monitoring  
   system-wide 366  
   tools for AIX 366  
   tools for IBM i 366  
   tools for Linux 366  
   tools for VIOS 366  
 restore  
   additional data 206  
   HMC 206  
   nim\_resources.tar 207  
   SEA 217  
   tape 207  
   to a different partition 213  
   user defined virtual devices 216  
   VIOS 206  
   VIOS with NIM 209  
 restore VIOS from DVD 206  
 restore VIOS from remote file 207  
 restore VIOS from tape 207

- restorevgstruct command 216
- rmbdsp command 48
- rmdev command 22, 260
- rmuser command 167
- RMVTCPIFC command 118
- RSCT daemons 272
- rsh command 23
- Rsi.hosts file 420
- runqueue 429
  
- S**
- SAN 216
- sar command 412, 414, 431, 442
- save HMC profiles 319
- savevgstruct command 200, 216
- Scaled Processor Utilization of Resources Register 414
- SCP 26
- SCSI configuration
  - rebuild 221
- SCSI reservation 54
- seastat command 492, 495
- Secure Copy, SCP 26
- secure shell, ssh 144, 371
- security
  - kerberos 164
  - LDAP
    - gsk7ikm 150
    - ldapsearch 160, 162
    - mkldap 161
    - mksecldap 156
- shared CPU 408
- Shared Ethernet Adapter, SEA 318
  - advanced SEA monitoring 492
  - attributes 483
  - backup 201
  - checking 235
  - create 483
  - delay in failover 478
  - Ethernet statistics 484
  - kernel threads 126
  - loose mode 138
  - mapping physical to virtual 224
  - monitoring
    - hash\_mode 482
    - Network monitoring testing scenario 482
  - Quality of service 137
  - restore 217
  - SEA threading on VIOS 126
  - statistics based on search criterion 495
  - strict mode 138
  - switching active 238
  - testing SEA failover 477
  - verify primary adapter 478
- shared optical device 16
- shared partitions
  - capped 408
  - uncapped 408
- shared processor 408
- Shared Processor Pool 244, 406
- showmount command 404
- shutdown command 168, 228
- simultaneous multithreading 408
- simultaneous multithreading, SMT 408, 412, 424
- single\_path parameter 318
- snmp\_info command 404
- SPOT 209
- SPT, System Planning Tool 31, 33
- SPURR 414
- ssh command 23, 323
- SSH keys 318
- stale partitions 464
- startnetsh command 163, 215
- startsv command 382
- stolen processor cycles 418
- stopnetsh command 144
- stopsvc command 382
- Strict mode QoS 138
- STRPFRCOL command 455, 468
- STRPFRT command 444, 455, 469
- STRSST command 42, 233, 466
- STRTCPIFC command 118, 133
- su command 167
- svmon command 402
- syncvg command 236
- sysstat command 402
- System i Navigator 448
- System Planning Tool
  - Deploy System Plan 311
  - Export the System Plan 317
- System Planning Tool, SPT 31, 33, 218, 305
  
- T**
- tape
  - backup 193
- tcptr command 141

Tivoli 369–370  
 Tivoli Enterprise Portal, TEP 370  
 Tivoli Monitoring 370  
 Tivoli Storage Manager, TSM 380  
     restore 214  
 Tivoli Usage and Accounting Manager, ITUAM 381,  
 383, 386  
 topas 419  
     allow performance information collection 425  
     logical partition display 423  
     logical processors view 424  
     processor subsection display 424  
     real time consumption 422  
     SMIT interface 435  
 topas command 402, 416, 419–420, 422, 424, 435,  
 452  
 topasout command 435, 441  
 topasrec command 8  
 TotalStorage Productivity Center, TPC 369  
 traceroute command 404  
 TSM  
     agent configuration 380

## U

unloadopt command 28  
 unmirrorios command 168  
 unmirrorvg command 43  
 update VIOS 226  
 updatehwdev command 362  
 updateios command 228  
 update-lsvpd-db command 277  
 updating Virtual I/O Server 226

## V

variable processor frequency 415, 430  
 varyonvg command 511  
 vasistat command 402  
 version command 63  
 vfcmap command 61, 70  
 VIOS  
     backup 191  
     backup and restore methods 192  
     backup disk structures 200  
     backup linking information 201  
     backup scheduling 192  
     backup strategy 188  
     backup to DVD-RAM 194  
     backup to remote file 196

backup via TSM 203  
 change VLAN 118  
 checking network health 230  
 checking storage health 460  
 commit updates 228  
 create an ISO image 27  
 Denial of Service, DoS 140  
 Development engineer user 167  
 entstat 483  
 error logging 239  
 Ethernet statistics 484  
 expansion pack 164  
 file backed devices 25  
 firewall 144  
 firewall disable 146  
 fix pack 172  
 installation 172  
 installation DVD 172  
 interim fixes 235  
 link aggregation 481  
 list interim fixes 227  
 managing users  
     creating users 167  
     global command log (gcl) 169  
     read-only account 169  
     service representative (SR) account 168  
     system administrator account 167  
 migration DVD 172  
 migration from an HMC 173  
 migration from DVD 174  
 migration to version 2.1 172  
 monitoring commands 401  
 monitoring the Virtual I/O Server 476  
 Network  
     Maximum transfer unit 127  
     Path MTU discovery 129  
     TCP checksum offload 135  
     using jumbo frames 132  
 network address change 116  
 network mapping 118  
 network monitoring 481  
 NPIV 21  
 packet count 486  
 rebuild network configuration 224  
 rebuild the Virtual I/O Server 218  
 replace a Ethernet adapter 287  
 replace a Fibre Channel adapter 290  
 replacing a disk drive 41  
 reserve policy 232

- restore 206
  - restore from DVD 206
  - restore from NIM 209
  - restore from remote file 207
  - restore from tape 207
  - restore to a different partition 213
  - restore user defined virtual devices 216
  - restore with TSM 214
  - schedule with crontab 192
  - Security Hardening Rules 148
  - Service representative user 167
  - SMS menu 175
  - storage security zones 51
  - subnet mask 225
  - supported tape drives 21
  - System administrator user 167
  - topology 391
  - tracing a virtual storage configuration 35
  - unconfigure a device 288
  - update dual VIOS 228
  - updating 226
    - dual server 228
    - single server 226
  - virtual device slot numbers 120
  - virtual Ethernet
    - path MTU discovery 135
  - Virtual I/O Server as a LDAP client 149
  - virtual media repository 25
  - virtual optical device 27
  - virtual optical media 25
  - virtual optical media disk 26
  - virtual target devices 223
  - viosecure command 140, 146, 203
  - viostat command 403, 460
  - virtual adapters remove 269
  - Virtual Asynchronous Services Interface, VASI 296
  - virtual CPU 408
  - virtual Ethernet
    - Integrated Virtualization Manager, IVM 396
    - introduction 477
    - performance 135
    - slot numbers 202
  - virtual Ethernet adapter rebuild 224
  - virtual Fibre Channel adapter 57
    - managing 60
      - HMC-managed system 60
      - IVM-managed system 61
  - virtual IVE network 317
  - virtual media repository 25
  - virtual optical device 27
    - load media 27
    - unload media 28
  - virtual optical devices 15
  - virtual optical media 25
  - virtual processor 408
    - change 250
    - definition 407
    - monitoring 405
    - spare capacity 413
    - terminology 406
  - virtual SCSI
    - disk mapping options 12
    - Integrated Virtualization Manager, IVM 396
    - Linux SCSI re-scan 286
    - mapping of LUNs 29
    - number of devices per adapter 34
    - physical volumes 12
    - shared optical device 16
    - slot number 53
    - slot numbers 202
    - topology 391
    - tracing a virtual storage configuration 35
    - virtual optical devices 15
    - virtual optical media 25
    - virtual tape 20
  - virtual SCSI disk 12
  - virtual storage
    - monitoring 459
  - virtual tape 20
    - unconfigure or use in VIOS 24
  - virtual target devices 223
  - VLAN 217
    - Changing IP addresses or VLAN 116
  - VLAN tag 124
  - vmstat command 402, 428
  - VRFCFG command 19, 133
- W**
- waitqueue 423
  - wiki 508
  - wkldout command 402
  - workload group 297
  - world wide port name, WWPN 56
  - WPAR 302
  - WRKCFGSTS command 19, 133
  - WRKDSKSTS command 467
  - WRKHDWRSC command 16, 122, 499

WRKSYSACT command 442  
WRKSYSSTS command 453, 466

**X**

xmwlrm command 435

**Z**

zoneshow command 71, 85

To determine the spine width of a book, you divide the paper PPI into the number of pages in the book. An example is a 250 page book using Plainfield opaque 50# smooth which has a PPI of 526. Divided 250 by 526 which equals a spine width of .4752". In this case, you would use the .5" spine. Now select the Spine width for the book and hide the others: **Special>Conditional Text>Show/Hide>SpineSize(-->Hide:)>Set** . Move the changed Conditional text settings to all files in your book by opening the book file with the spine:fm still open and **File>Import>Formats** the Conditional Text Settings (ONLY!) to the book files.  
Draft Document for Review November 7, 2008 4:31 pm

**7590spine.fm 539**

## PowerVM Virtualization on Power Systems: Managing and

(0.5" spine)  
0.475" <-> 0.875"  
250 <-> 459 pages



To determine the spine width of a book, you divide the paper PPI into the number of pages in the book. An example is a 250 page book using Plainfield opaque 50# smooth which has a PPI of 526. Divided 250 by 526 which equals a spine width of .4752". In this case, you would use the .5" spine. Now select the Spine width for the book and hide the others: **Special>Conditional Text>Show/Hide>SpineSize(->Hide:)}>Set** . Move the changed Conditional text settings to all files in your book by opening the book file with the spine:fm still open and **File>Import>Formats** the Conditional Text Settings (ONLY!) to the book files.  
Draft Document for Review November 7, 2008 4:31 pm

**7590spine.fm 540**





# PowerVM Virtualization on IBM Power Systems (Volume 2):



**Covers AIX, IBM i and Linux for Power virtual I/O clients**

**A collection of managing and monitoring best practices focused on Virtualization**

**Includes the Virtual I/O Server 2.1 enhancements**

PowerVM virtualization technology is a combination of hardware and software that supports and manages the virtual environments on POWER5, POWER5+ and POWER6-based systems. It is a major tool to help simplify and optimize your IT infrastructure.

Available on IBM Power System, and IBM BladeCenter JS12 and SJ22 servers as optional Editions and supported by the AIX, IBM I, and Linux for Power operating systems, this set of comprehensive systems technologies and services is designed to enable you to aggregate and manage resources using a consolidated, logical view. The key benefits of deploying PowerVM virtualization and IBM Power Systems are as follows:

- ▶ Cut energy costs through server consolidation
- ▶ Reduce the cost of existing infrastructure
- ▶ Manage growth, complexity, and risk on your infrastructure

To achieve this goal, PowerVM virtualization provides the following technologies:

- ▶ Virtual Ethernet
- ▶ Shared Ethernet Adapter
- ▶ Virtual SCSI and Fibre Channel
- ▶ Micro-Partitioning technology

Additionally, these new technologies are available on POWER6 systems:

- ▶ Multiple Shared-Processor Pools
- ▶ Optional PowerVM Live Partition Mobility

This publication is an extension of *PowerVM Virtualization on System p: Introduction and Configuration*, SG24-7940. It provides an organized view of best practices for managing and monitoring your PowerVM

## 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)