

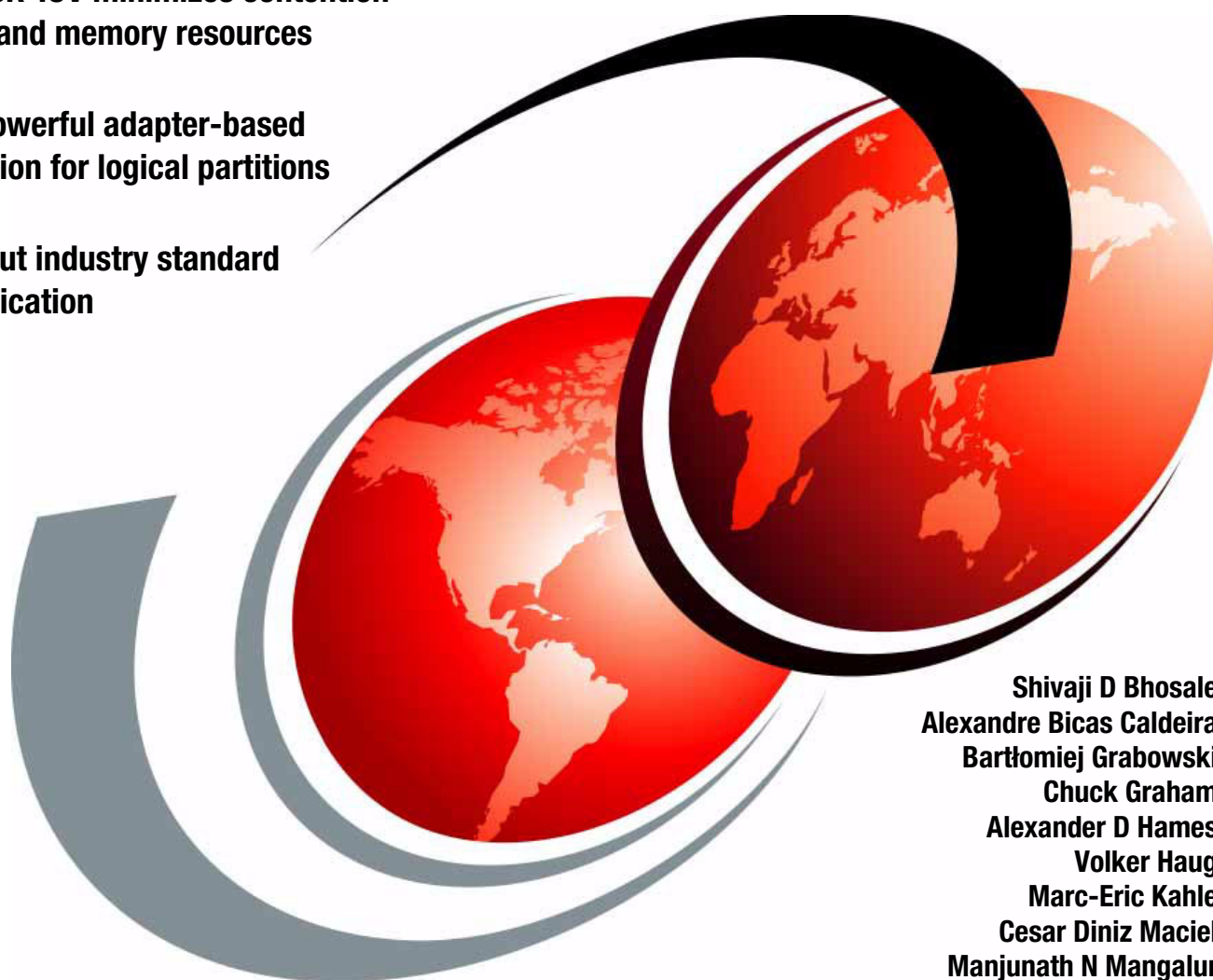
IBM Power Systems SR-IOV

Technical Overview and Introduction

See how SR-IOV minimizes contention with CPU and memory resources

Explore powerful adapter-based virtualization for logical partitions

Learn about industry standard PCI specification



Shivaji D Bhosale
Alexandre Bicas Caldeira
Bartłomiej Grabowski
Chuck Graham
Alexander D Hames
Volker Haug
Marc-Eric Kahle
Cesar Diniz Maciel
Manjunath N Mangalur
Monica Sanchez



International Technical Support Organization

IBM Power Systems SR-IOV: Technical Overview and Introduction

July 2014

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

First Edition (July 2014)

This edition applies to the IBM Power 770 (9117-MMD), IBM Power 780 (Machine type 9179-MHD) , IBM Power ESE (Machine type 8412-EAD) Power Systems servers and HMC V7R7.9.0.

© Copyright International Business Machines Corporation 2014. All rights reserved.

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

Contents

Notices	v
Trademarks	vi
Preface	vii
Authors	vii
Now you can become a published author, too!	ix
Comments welcome	ix
Stay connected to IBM Redbooks	x
Chapter 1. SR-IOV overview	1
1.1 Introduction	2
1.2 SR-IOV compared with similar virtualization technologies	2
1.2.1 Overview comparison	2
1.2.2 PowerVM Shared Ethernet Adapter versus SR-IOV	3
1.2.3 Integrated Virtual Ethernet versus SR-IOV	3
1.3 Benefits of single root I/O virtualization (SR-IOV)	3
1.3.1 Direct access I/O and performance	3
1.3.2 Adapter sharing	4
1.3.3 Adapter resource provisioning (QoS)	4
1.3.4 Flexible deployment	4
1.3.5 Reduced costs	5
1.4 Architecture overview	5
Chapter 2. Planning	7
2.1 SR-IOV hardware requirements and planning introduction	8
2.2 Hardware requirements	8
2.3 Operating system requirements	9
2.4 System management requirements	9
Chapter 3. Deployment scenarios	11
3.1 Single partition	12
3.2 Multiple partitions	13
3.3 Using SR-IOV and VIOS	13
3.4 Link aggregation	17
Chapter 4. Configuration	21
4.1 Verify prerequisites	22
4.1.1 Verify that the Power Systems server is SR-IOV Capable	22
4.1.2 Verify that the system has at least one SR-IOV capable adapter	23
4.2 Adapter operations	23
4.2.1 Switching adapter from dedicated mode to SR-IOV shared mode	23
4.2.2 Switching adapter from SR-IOV shared mode to dedicated mode	24
4.3 Physical port operations	25
4.3.1 Physical port properties: General	26
4.3.2 Physical port properties: Advanced	26
4.3.3 Physical port properties: Port Counters	28
4.4 Logical port operations	29
4.4.1 Adding a logical port during partition creation	29
4.4.2 Adding a logical port using dynamic partitioning	32

4.4.3 Editing a logical port	33
4.5 Device mapping	36
4.6 Operating system adapter configuration	37
4.6.1 AIX	37
4.6.2 IBM i	40
4.6.3 Linux	41
4.6.4 Virtual I/O Server	42
4.7 Adapter configuration backup and restore	46
4.8 Command-line interface (CLI) support	47
4.9 Miscellaneous notes	48
Chapter 5. Maintenance	49
5.1 Adapter firmware	50
5.2 Problem determination and data collection	51
5.2.1 SR-IOV Platform Dump	51
5.3 Problem recovery	56
5.4 Concurrent maintenance	59
5.4.1 Adapter concurrent and non-concurrent maintenance	60
5.5 IBM i performance metrics	64
5.6 HMC commands for SR-IOV handling	66

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 grant 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 websites are provided for convenience only and do not in any manner serve as an endorsement of those websites. The materials at those websites are not part of the materials for this IBM product and use of those websites is at your own risk.

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

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

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

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

COPYRIGHT LICENSE:

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

Trademarks

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

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

Active Memory™

AIX®

developerWorks®

Global Technology Services®

IBM®

POWER®

Power Systems™

POWER7®

POWER7+™

POWER8®

PowerHA®


PowerVM®

PureFlex®

PureSystems®

Redbooks®

Redpaper™

Redbooks (logo) ®

RS/6000®

System Storage®

Tivoli®

The following terms are trademarks of other companies:

C3, and Phyteland device are trademarks or registered trademarks of Phytel, Inc., an IBM Company.

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

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

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

Preface

This IBM® Redpaper™ publication describes the adapter-based virtualization capabilities that are being deployed in high-end IBM POWER7+™ and POWER8® processor-based servers.

Peripheral Component Interconnect Express (PCIe) single root I/O virtualization (SR-IOV) is a virtualization technology on IBM Power Systems™ servers. SR-IOV allows multiple logical partitions (LPARs) to share a PCIe adapter with little or no run time involvement of a hypervisor or other virtualization intermediary.

SR-IOV does not replace the existing virtualization capabilities that are offered as part of the IBM PowerVM® offerings. Rather, SR-IOV compliments them with additional capabilities.

This paper describes many aspects of the SR-IOV technology:

- ▶ A comparison of SR-IOV with standard virtualization technology
- ▶ Overall benefits of SR-IOV
- ▶ Architectural overview of SR-IOV
- ▶ Planning requirements
- ▶ SR-IOV deployment models that use standard I/O virtualization
- ▶ Configuring the adapter for dedicated or shared modes
- ▶ Tips for maintaining and troubleshooting your system
- ▶ Scenarios for configuring your system

This paper is directed to clients, IBM Business Partners, and system administrators who are involved with planning, deploying, configuring, and maintaining key virtualization technologies.

Authors

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

Shivaji D Bhosale is a Team Lead at Application and Infrastructure Systems Management (ISM), Systems and Technology Group, Pune, India. He has 17 years of experience in IT field. He holds a Master's degree in Computer Management. He is an IBM Certified Specialist in Cloud Computing Infrastructure Architect. Shivaji is coauthor of IBM Redbooks® publications (*WebSphere Cloudburst Appliance and PowerVM*, SG24-7806, and *Adopting IBM PureApplication System V1.0*, SG24-8113), and several IBM developerWorks® articles about IBM PureSystems®.

Alexandre Bicas Caldeira works on the IBM Power Systems Advanced Technical Support team for IBM Brazil. He holds a degree in computer science from the Universidade Estadual Paulista (UNESP). Alexandre has more than 14 years of experience working with IBM and IBM Business Partners on Power Systems hardware, IBM AIX®, and PowerVM virtualization products. He is also skilled on IBM System Storage®, IBM Tivoli® Storage Manager, IBM PureSystems, IBM System x, and VMware.

Bartłomiej Grabowski is an IBM i, and PowerVM Senior Technical Specialist in DHL IT Services in the Czech Republic. He has nine years of experience with IBM i. He holds a Bachelor's degree in Computer Science from Academy of Computer Science and

Management in Bielsko-Biala. His areas of expertise include IBM i administration, IBM PowerHA® solution, based on hardware and software replication, Power Systems hardware, and PowerVM. He is an IBM Certified Systems Expert and a coauthor of several PowerVM Redbooks.

Chuck Graham is a Senior Technical Staff Member and is the Lead Architect for PowerVM SR-IOV support on Power Systems. He joined IBM in 1982 after graduating from the University of Iowa with an Bachelor of Science degree in Electrical and Computer Engineering. He has spent the majority of his career as a Developer and Architect of physical and virtual I/O solutions for IBM computer systems and currently works in the Power Hypervisor area. Chuck is also a Master Inventor with numerous patents in the field of computer I/O.

Alexander D Hames is an I/O Development Engineer on the Systems Solutions Enablement and I/O Development team within the Systems and Technology Group in Austin, TX. He holds a degree in Electrical Engineering from West Virginia University. He has six years of experience at IBM and his areas of expertise include Fibre Channel, Ethernet, SAN storage, I/O virtualization, and Power Systems.

Volker Haug is an Open Group Certified IT Specialist within IBM Systems and Technology Group in Germany, supporting Power Systems clients and IBM Business Partners. He holds a diploma degree in Business Management from the University of Applied Studies in Stuttgart. His career includes more than 27 years of experience with Power Systems, AIX, and PowerVM virtualization; he has written several IBM Redbooks publications about Power Systems and PowerVM. Volker is a IBM POWER8 Champion and a member of the German Technical Expert Council, an affiliate of the IBM Academy of Technology.

Marc-Eric Kahle is an AIX Software Specialist at the IBM Global Technology Services® in Ehningen, Germany. He also has worked as a Power Systems Hardware Support Specialist in the IBM RS/6000®, Power Systems, and AIX fields since 1993. He has worked at IBM Germany since 1987. His area of expertise includes Power Systems hardware and he is an AIX Certified Specialist. He has participated in the development of seven other IBM Redbooks publications.

Cesar Diniz Maciel is an Executive IT Specialist with IBM in the United States. He joined IBM in 1996 in Presales Technical Support for the IBM RS/6000 family of UNIX servers in Brazil, and came to IBM United States in 2005. He is part of the Global Techline team, working on presales consulting for Latin America. He holds a degree in Electrical Engineering from Universidade Federal de Minas Gerais (UFMG) in Brazil. His areas of expertise include Power Systems, AIX, and IBM Power Virtualization. He has written extensively about Power Systems and related products. This is his eighth ITSO residency.

Manjunath N Mangalur is a Staff Software Engineer at the IBM Power Systems and Technology Lab in IBM India. He holds a degree in Information Science from Vishweshwaraiah Technological University. He has over eight years of experience with IBM and has worked with the AIX operating system, IBM Power Systems and PureFlex® systems. His areas of expertise include AIX security, kernel, Linux, PowerVM Enterprise, virtualization, Power Systems, PureFlex systems, IBM Systems Director, and reliability, availability, serviceability (RAS).

Monica Sanchez is an Advisory Software Engineer with more than 13 years of experience in AIX and Power Systems support. Her areas of expertise include AIX, HMC, and networking. She holds a degree in Computer Science from Texas A&M University and is currently part of the Power HMC Product Engineering team, providing level 2 support for the IBM Power Systems Hardware Management Console.

The project that produced this publication was managed by:

Scott Vetter
Executive Project Manager, PMP

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

Tamikia Barrow, Bill Brandmeyer, Charlie Burns, Medha D. Fox, Charles S. Graham, Alexander Hames, Samuel Karunakaran, Kris Kendall, Stephen Lutz, Michael J. Mueller, Kanisha Patel, Anitra Powell, Rajendra Patel, Vani Ramagiri, Woodrow Lemcke, Tim Schimke, Jacobo Vargas, Bob Vidrick
IBM

Now you can become a published author, too!

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

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

ibm.com/redbooks/residencies.html

Comments welcome

Your comments are important to us!

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

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

ibm.com/redbooks

- ▶ Send your comments in an email to:

redbooks@us.ibm.com

- ▶ Mail your comments to:

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

Stay connected to IBM Redbooks

- ▶ Find us on Facebook:
<http://www.facebook.com/IBMRedbooks>
- ▶ Follow us on Twitter:
<http://twitter.com/ibmredbooks>
- ▶ Look for us on LinkedIn:
<http://www.linkedin.com/groups?home=&gid=2130806>
- ▶ Explore new Redbooks publications, residencies, and workshops with the IBM Redbooks weekly newsletter:
<https://www.redbooks.ibm.com/Redbooks.nsf/subscribe?OpenForm>
- ▶ Stay current on recent Redbooks publications with RSS Feeds:
<http://www.redbooks.ibm.com/rss.html>



SR-IOV overview

Single root I/O virtualization (SR-IOV) is an extension to the PCI Express (PCIe) specification that allows multiple operating systems to simultaneously share a PCIe adapter with little or no runtime involvement from a hypervisor or other virtualization intermediary.

This chapter introduces SR-IOV architecture, and key benefits and features when deployed on IBM Power Systems.

1.1 Introduction

SR-IOV is PCI standard architecture that enables PCIe adapters to become self-virtualizing. It enables adapter consolidation, through sharing, much like logical partitioning enables server consolidation. With an adapter capable of SR-IOV, you can assign virtual *slices* of a single physical adapter to multiple partitions through logical ports; all of this is done without the need for a Virtual I/O Server (VIOS).

Initial SR-IOV deployment supports up to 48 logical ports per adapter, depending on the adapter (for a list of available SR-IOV adapters, see Figure 2-1 on page 8). You can provide additional fan-out for more partitions by assigning a logical port to a VIOS, and then using that logical port as the physical device for a Shared Ethernet Adapter (SEA). VIOS clients can then use that SEA through a traditional virtual Ethernet configuration.

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

1.2 SR-IOV compared with similar virtualization technologies

SR-IOV has several key differences, compared to other technologies such as IBM PowerVM Shared Ethernet Adapters or Integrated Virtual Ethernet (IVE). For the IBM POWER7® and POWER7+, and POWER8 processor-based systems, PowerVM SEA has been the standard method for virtualizing an Ethernet adapter. In some POWER7 processor-based systems, the planar Ethernet adapter (IVE) is another available option. The next sections compare these technologies to SR-IOV technology and to virtual Network Interface Controller (vNIC) technology which leverages SR-IOV.

1.2.1 Overview comparison

Table 1-1 lists the basic differences between the three available virtualization technologies for Ethernet adapters.

Table 1-1 Virtualization technology comparison

Technology	Live Partition Mobility	Quality of service (QoS)	Direct access perf.	Per client scalability	Link Aggregation/Etherchannel	Requires VIOS
SR-IOV	No ^a	Yes	Yes	High	Yes ^b	No
vNIC	Yes	Yes	No ^c	High	Yes ^b	Yes
SEA/vEth	Yes	No	No	Medium	Yes	Yes
IVE	No	No	Yes	Low	Yes	No

a. SR-IOV can optionally be combined with VIOS and virtual Ethernet to use higher-level virtualization functions like Live Partition Mobility (LPM). See Chapter 3, “Deployment scenarios” on page 11 for possible deployment scenarios.

b. See 3.4, “Link aggregation” on page 17 for more information

c. Requires fewer system resources when compared to SEA/vEth. See 3.3, “Using SR-IOV and VIOS” on page 13.

1.2.2 PowerVM Shared Ethernet Adapter versus SR-IOV

PowerVM Shared Ethernet Adapter (SEA) has advantages and disadvantages when compared with SR-IOV. Additionally, SR-IOV can be used in conjunction with PowerVM SEA to provide extra scalability and functionality such as Live Partition Mobility.

Generally, SR-IOV and PowerVM SEA are for two separate use cases. The PowerVM SEA technology can scale well for connectivity and provides options for advanced functions such as Live Partition Mobility. For a large scale application that requires little bandwidth per virtual adapter and no bandwidth control per virtual adapter, and particularly this case for an existing PowerVM installation, SEA has the potential to be the appropriate solution.

However, PowerVM SEA does not allow a hardware-direct connection, and the extra context switches require additional processing, and has few options for quality of service (QoS). SR-IOV can be used as a high performance solution with varying bandwidth requirements per logical port and SR-IOV allows administrators to control bandwidth allocations per logical port.

PowerVM SEA can be implemented using nearly every Ethernet adapter in the IBM Power Systems portfolio; SR-IOV is limited to certain adapters. Also, SR-IOV supports only selected Power Systems servers. PowerVM SEA is a feature of the VIOS. VIOS is not a requirement for SR-IOV. Fewer system resources are required to use SR-IOV than PowerVM SEA overall.

PowerVM vNIC provides options for advanced functions such as Live Partition Mobility with better performance and I/O efficiency when compared to PowerVM SEA. In addition, PowerVM vNIC provides users with bandwidth control (QoS) by leveraging SR-IOV logical ports as the physical interface to the network.

1.2.3 Integrated Virtual Ethernet versus SR-IOV

IVE has many similarities to SR-IOV from a functionality point of view. IVE and SR-IOV both support certain 1Gbps and 10Gbps Ethernet feature codes and provide logical ports to LPARs for Ethernet connectivity. Both can be configured without using the VIOS.

SR-IOV differs from IVE in areas such as QoS and system availability. For IVE, a logical port competes for bandwidth with all the other logical ports defined on the same physical port. This does not prevent a logical port from reaching the media speed if the other logical ports do not have an intensive I/O workload. The QoS feature on SR-IOV assigns the minimum bandwidth percentages that you want per logical port. A logical port can go above this percentage if no contention for bandwidth exists on the link.

1.3 Benefits of single root I/O virtualization (SR-IOV)

SR-IOV provides significant performance and usability benefits, as described in this section.

1.3.1 Direct access I/O and performance

The primary benefit of allocating adapter functions directly to a partition, as opposed to using a virtual intermediary (VI) like VIOS, is performance. The processing overhead involved in passing client instructions through a VI, to the adapter and back, are substantial.

With direct access I/O, SR-IOV capable adapters running in shared mode allow the operating system to directly access the slice of the adapter that has been assigned to its partition, so

there is no control or data flow through the hypervisor. From the partition perspective, the adapter appears to be physical I/O. With respect to CPU and latency, it exhibits the characteristics of physical I/O; and because the operating system is directly accessing the adapter, if the adapter has special features, like multiple queue support or receive side scaling (RSS), the partition can leverage those also, if the operating system has the capability in its device drivers.

1.3.2 Adapter sharing

The current trend of consolidating servers to reduce cost and improve efficiency is increasing the number of partitions per system, driving a requirement for more I/O adapters per system to accommodate them. SR-IOV addresses and simplifies that requirement by enabling the sharing of SR-IOV capable adapters. Because each adapter can be shared and directly accessed by up to 48 partitions, depending on the adapter, the partition to PCI slot ratio can be significantly improved without adding the overhead of a virtual intermediary.

1.3.3 Adapter resource provisioning (QoS)

Power Systems SR-IOV provides QoS controls to specify a capacity value for each logical port, improving the ability to share adapter ports effectively and efficiently. The capacity value determines the desired minimum percentage of the physical port's resources that should be applied to the logical port.

The exact resource represented by the capacity value can vary based on the physical port type and protocol. In the case of Ethernet physical ports, capacity determines the minimum percentage of the physical port's transmit bandwidth that the user desires for the logical port.

For example, consider Partitions A, B, and C, with logical ports on the same physical port. If Partition A is assigned an Ethernet logical port with a capacity value of 20%, Partitions B and C cannot use more than 80% of the physical port's transmission bandwidth unless Partition A is using less than 20%. Partition A can use more than 20% if bandwidth is available. This ensures that, although the adapter is being shared, the partitions maintain their portion of the physical port resources when needed.

1.3.4 Flexible deployment

Power Systems SR-IOV enables flexible deployment configurations, ranging from a simple, single-partition deployment, to a complex, multi-partition deployment involving VIOS partitions and VIOS clients running different operating systems.

In a single-partition deployment, the SR-IOV capable adapter in shared mode is wholly owned by a single partition, and no adapter sharing takes place. This scenario offers no practical benefit over traditional I/O adapter configuration, but the option is available.

In a more complex deployment scenario, an SR-IOV capable adapter could be shared by both VIOS and non-VIOS partitions, and the VIOS partitions could further virtualize the logical ports as shared Ethernet adapters for VIOS client partitions. This scenario leverages the benefits of direct access I/O, adapter sharing, and QoS that SR-IOV provides, and also the benefits of higher-level virtualization functions, such as Live Partition Mobility (for the VIOS clients), that VIOS can offer.

For more deployment scenarios, see Chapter 3, "Deployment scenarios" on page 11.

1.3.5 Reduced costs

SR-IOV facilitates server consolidation by reducing the number of physical adapters, cables, switch ports, and I/O slots required per system. This translates to reduced cost in terms of physical hardware required, and also reduced associated energy costs for power consumption, cooling, and floor space. You may save additional cost on CPU and memory resources, relative to a VIOS adapter sharing solution, because SR-IOV does not have the resource overhead inherent in using a virtualization intermediary to interface with the adapters.

1.4 Architecture overview

To accomplish adapter virtualization without a virtualization intermediary, the *Single Root I/O Virtualization and Sharing Specification* introduces the concepts of Physical Functions (PFs) and Virtual Functions (VFs). A Physical Function is a PCIe function that supports SR-IOV capabilities as defined in the specification. A Virtual Function is a PCIe function that is associated with a PF and is directly accessible by a system image, such as an operating system. It shares physical resources, such as an Ethernet Link, with the PF and other VFs that are associated with the same PF.

At times, this paper uses the Virtual Function and logical port interchangeably. The history of this is as follows:

Virtual Function (VF) A VF is a term used by the PCI Special Interest Group (SIG) to define a specific type of PCI function. As a PCI function, a VF has characteristics and capabilities that determine how the VF behaves on a PCI bus. The definition of a VF does not include how a VF maps to any particular parts of a PCI device (for example, network ports) other than its PCI interface.

Logical port For PowerVM, a logical port is a configurable entity that defines characteristics and capability for a portion of a physical port on a I/O device. Platform firmware uses the logical port configuration information to manage platform firmware resources and to configure an I/O device. When an SR-IOV logical port is activated, either through partition activation or through a DLPAR add operation, a VF is associated with the logical port to allow the partition to access the PCI device.

SR-IOV capable adapters can be used in two modes.

Dedicated mode This is the traditional mode, where the I/O adapter is assigned to a partition and ports are not shared. The partition owns the whole adapter and manages it from a single operating system.

SR-IOV shared mode In shared mode, the adapter is assigned to the Power Hypervisor firmware. In this mode, the adapter can be shared by multiple operating systems at the same time. Each operating system accesses its share of the adapter using a VF device driver.

In SR-IOV shared mode, adapters partition their host interface using VFs. Power Systems SR-IOV implements VFs as logical ports. Each logical port is associated with a physical port of the adapter.

Logical ports are created for a partition through the Hardware Management Console (HMC) and given a capacity, which determines the desired percentage of the physical port's bandwidth for the partition to use. (See Chapter 4, "Configuration" on page 21 for configuration details.) Each partition accesses its share of the adapter with its own VF device driver. From the partition perspective, the VF is considered a single-function, single-port adapter and treated like physical I/O. This last detail allows logical ports that are assigned to a virtualization intermediary, like VIOS, to further virtualize the adapter, extending its use to even more partitions.

Figure 1-1 shows the relationship between an SR-IOV adapter and client partitions, including a scenario in which logical ports are assigned to a virtual I/O server and then used as the physical devices of Shared Ethernet Adapters (SEA).

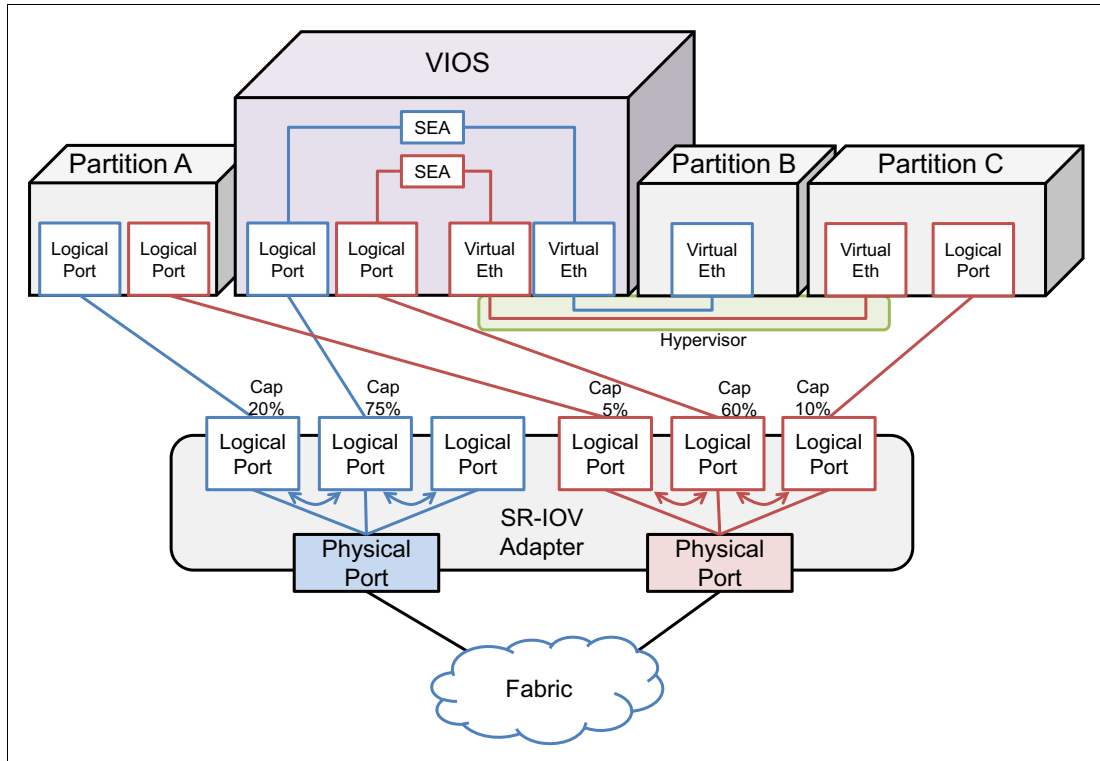


Figure 1-1 SR-IOV architecture



Planning

Implementing the SR-IOV technology requires some planning, which starts with the appropriate operating system levels, firmware level, adapter types, and adapter settings.

2.1 SR-IOV hardware requirements and planning introduction

SR-IOV adapters provide one or more external ports. The unit of virtualization is a slice of a port, also known as the Virtual Function (VF) and logical port (LP), as described in 1.4, “Architecture overview” on page 5.

SR-IOV functions present several configuration options that must be considered when planning. This section describes the server support requirement for implementing the SR-IOV functions and with IBM PowerVM.

To enable the SR-IOV adapter sharing function, a server that supports SR-IOV is required in addition to an SR-IOV capable I/O adapter. You must be aware of the following hardware considerations:

- ▶ Not all I/O slots support SR-IOV
- ▶ SR-IOV-capable I/O slots may have different capabilities
- ▶ The capabilities of SR-IOV adapters may differ
- ▶ PowerVM Standard or Enterprise Edition is required for using SR-IOV

2.2 Hardware requirements

Hardware requirements, related to SR-IOV functions, are as follows:

- ▶ Supported devices
 - Power E850 (8408-E8E)- any slot
 - Power S822 or S822L (8284-22A, 8247-22L)- 7 slots with both sockets populated C2, C3®, C5, C6, C7, C10, and C12, 4 slots if one C6, C7, C10, and C12
 - Power S824 S824L (8286-42A, 8247-42L)- 8 slots with both sockets populated C2, C3, C4, C5, C6, C7, C10, and C12, 4 slots if one C6, C7, C10, and C12
 - Power S814 or S812L (8286-41A, 8247-21L) - 4 slots C6, C7, C10, and C12
 - Power E870 (9119-MME) - Any system node slot
 - Power E880 (9119-MHE) - Any system node slot
 - PCIe Gen3 I/O expansion drawer slots C1, and C4 of the 6-slot Fan-out module
 - Power 770 (9117-MMD)
 - Power 780 (9179-MHD)
 - Power ESE (8412-EAD)

Firmware level: SR-IOV is supported from firmware level 780 on POWER7 processor-based servers and firmware level SC820_067 (FW820.10) for POWER8 processor-based servers. Check the Fix Central portal to verify the specific firmware level for your type of the machine.

<https://www.ibm.com/support/fixcentral/>

- ▶ One of the following pluggable PCIe adapters:

Table 2-1 Available SR-IOV capable I/O adapters

Adapter	Logical ports per adapter ^a	Low profile multiple OS	Full high multiple OS	Low profile Linux only	Full high Linux only
PCIe3 4-port (10 Gb FCoE and 1 GbE) SR optical fiber and RJ45	48 20/20/4/4	#EN0J ^b	#EN0H ^c	#EL38	#EL56

Adapter	Logical ports per adapter ^a	Low profile multiple OS	Full high multiple OS	Low profile Linux only	Full high Linux only
PCIe3 4-port (10 Gb FCoE and 1 GbE) SFP + copper twinax and RJ45	48 20/20/4/4	#EN0L ^b	#EN0K ^c	#EL3C	#EL57
PCIe3 4-port (10 Gb FCoE and 1 GbE) LR optical fiber and RJ45	48 20/20/4/4	#EN0N	#EN0M	N/A	N/A
PCIe3 4-port 10 GbE SR optical fiber	64 16/16/16/16	#EN16 ^d	#EN15	N/A	N/A
PCIe3 4-port 10 GbE copper twinax	64 16/16/16/16	#EN18 ^d	#EN17	N/A	N/A

a. This column provides the total logical ports per adapter followed by the number of logical ports available for each physical port.

b. SR-IOV announced February 2015 for Power E870/E880 system node now available on all POWER® servers

c. SR-IOV announced April 2014 for Power 770/780/ESE system node. With April 2015 announce, available on all POWER8 servers

d. Adapter is only available for Power E870/E880 system node, not 2U server

2.3 Operating system requirements

Minimum operating system requirements, related to SR-IOV functions, are as follows:

- ▶ VIOS
 - Virtual I/O Server Version 2.2.3.3 (2.2.3.4 with interim fix IV63331 or later for the E870 and E880)
- ▶ AIX
 - AIX 6.1 Technology Level 9 with Service Pack 2 (SP4 and APAR IV63331 or later for the E870 and E880)
 - AIX 7.1 Technology Level 3 with Service Pack 2 (SP4 and APAR IV63332 or later for the E870 and E880)
- ▶ Linux
 - SLES Linux Enterprise Server 11 SP3, or later
 - Red Hat Enterprise Linux 6.5, or later
 - Red Hat Enterprise Linux 7, or later
- ▶ IBM i
 - IBM i 7.1 TR8 (TR9 or later for the E870 and E880)
 - IBM i 7.2 (TR1 or later for the E870 and E880)

2.4 System management requirements

The Hardware Management Console (HMC) is required for the IBM Power 870 (Machine type 9119-MME), the IBM Power E880 (Machine type 9119-MHE), the IBM Power 770 (Machine type 9117-MMD), the IBM Power 780 (Machine type 9179-MHD), and the IBM Power ESE (Machine type 8412-EAD), and it is required to configure SR-IOV.

The minimum HMC code level for SR-IOV support is Version 7 Release 7.9.0 (HMC V7R7.9.0).



Deployment scenarios

In this chapter, we examine various scenarios that use SR-IOV when a solution is designed. This includes scenarios that focus on ease of configuration and also scenarios that provide advanced connectivity and integration with other platform features.

One of the benefits of SR-IOV is enabling the virtualization of the network adapters without the requirement of setting up and configuring a VIOS. This chapter highlights several configuration scenarios that use this capability.

3.1 Single partition

You can assign all ports on an adapter to a single partition, and the SR-IOV adapter behaves as a traditional dedicated adapter, with all the ports assigned to the partition. Figure 3-1 shows a representation of this scenario.

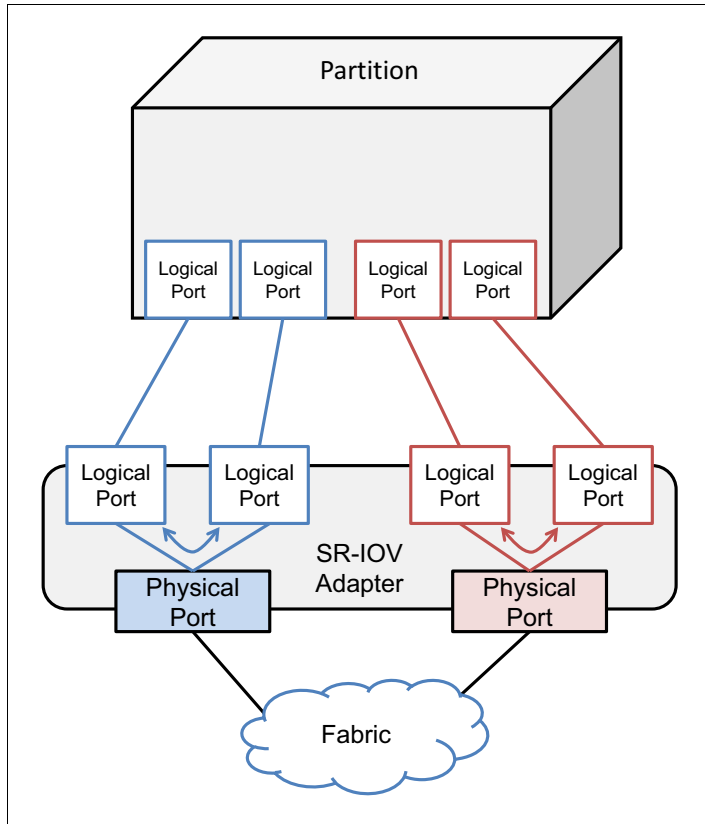


Figure 3-1 Single partition deployment

Although supported, this scenario does not offer any benefit over using a traditional adapter. The ports are dedicated to a single partition, and other partitions cannot share them.

Unsupported: This scenario is not supported by IBM i unless the system is connected to an HMC and the adapter is configured in shared mode.

3.2 Multiple partitions

When you have multiple partitions and need to connect them to the network is when SR-IOV provides the best value. Partitions can share physical ports, with the bandwidth of each port shared by the partitions according to the capacity value that is specified in the partition profile.

Figure 3-2 illustrates two partitions that share a pair of SR-IOV adapters. Each partition has one logical port associated with a physical port in two separate physical adapters. This is a good configuration for availability, because the two logical ports can be used in a high availability configuration, such as link aggregation.

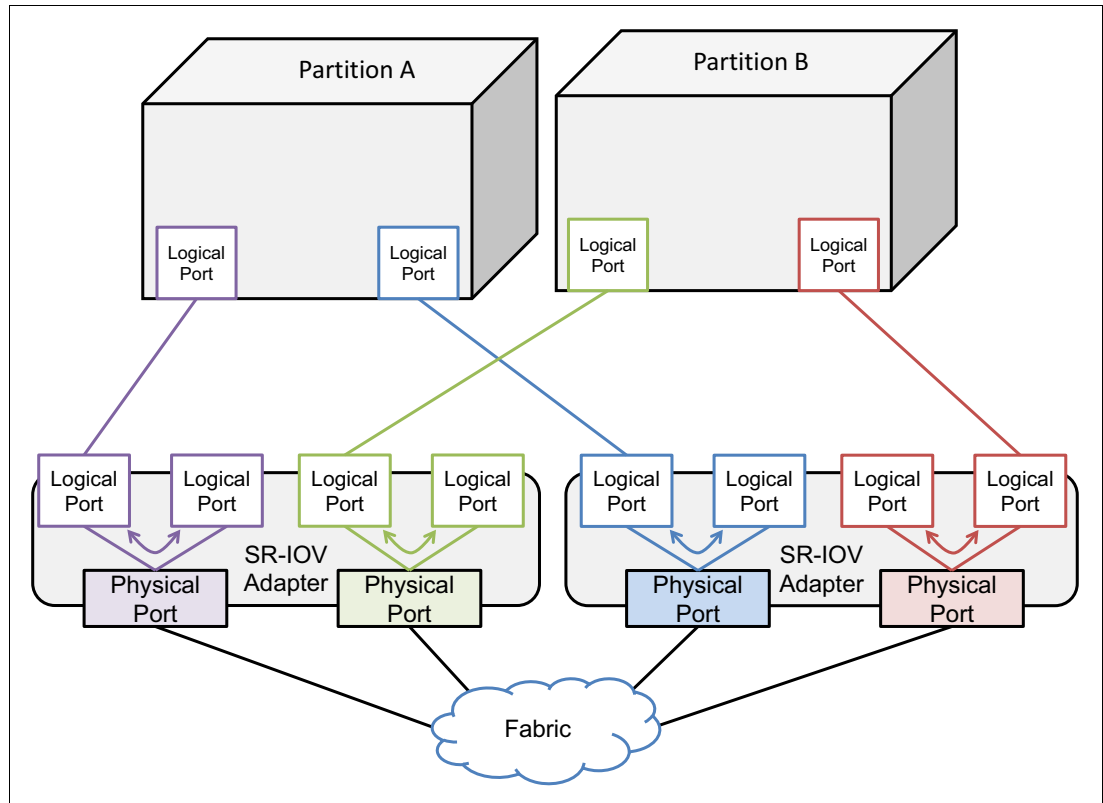


Figure 3-2 Multiple partitions accessing SR-IOV ports

3.3 Using SR-IOV and VIOS

Similar to an AIX, IBM i or Linux partition, a VIOS partition can have SR-IOV ports configured and used as regular adapters. These adapters can be part of a Shared Ethernet Adapter (SEA) configuration and bridge traffic from client partitions to the physical network. Because the client partitions configure only a Virtual Ethernet Adapter, they can continue using the advanced features of PowerVM, such as Live Partition Mobility and IBM Active Memory™ Sharing. Using SR-IOV ports in the VIOS instead of physical adapters offers the advantage of leveraging, on VIOS, the Capacity setting that is available for the SR-IOV logical ports. See “Logical port properties: General” on page 30 for more information about how to configure Capacity.

Figure 3-3 illustrates two SR-IOV adapters shared by two VIOS, and the logical ports configured as SEAs. Each VIOS has two logical ports, each port configured on a separate physical adapter. That provides adapter redundancy for both VIOS with only two adapters. The SR-IOV adapter is transparent to the client partition because it connects to the external network through the virtual Ethernet interface and SEA.

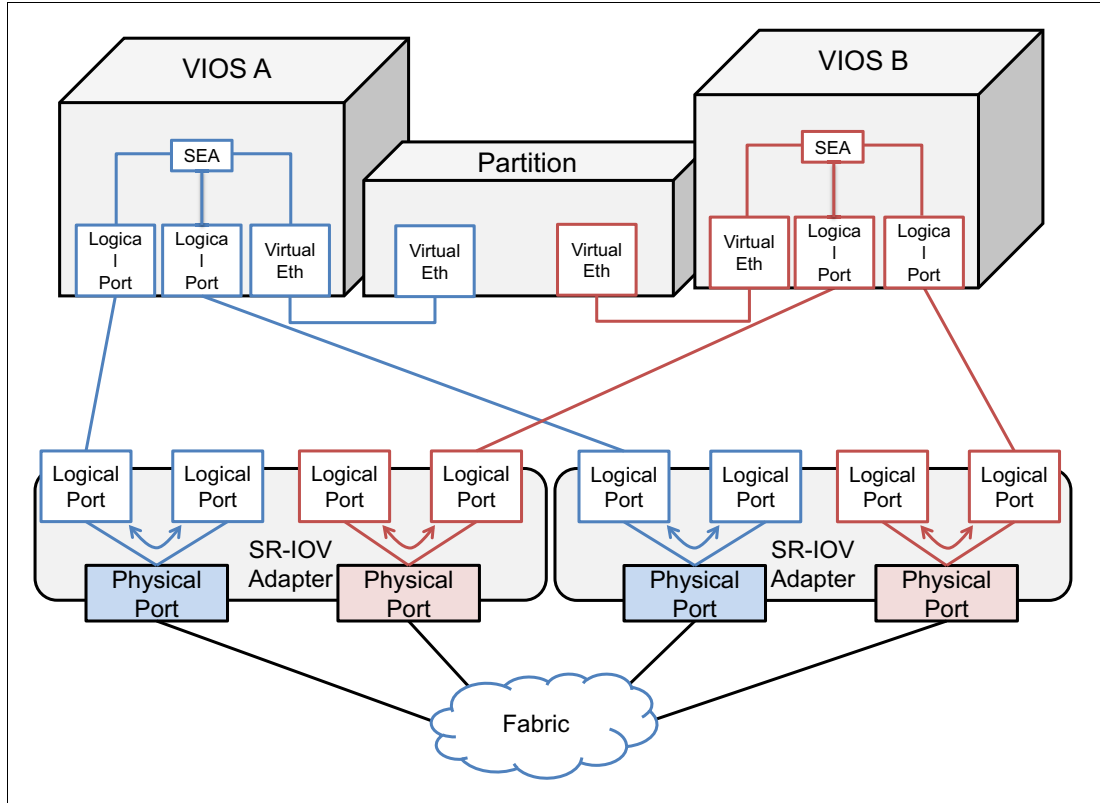


Figure 3-3 SR-IOV ports configured as SEA

SR-IOV logical ports can coexist with virtual adapters or physical dedicated adapters without restrictions. This capability enables a partition to use the SR-IOV logical port as the primary network interface, and to have a Virtual Ethernet Adapter as a backup interface. In case of an interruption in the network traffic through the logical port, it would then be routed through the backup interface. This can be achieved as follows:

- ▶ On AIX: By using the Network Interface Backup feature (Figure 3-4 on page 15).
- ▶ On IBM i: By using Virtual IP Address (VIPA).
- ▶ On Linux: You configure an active-backup bonding.

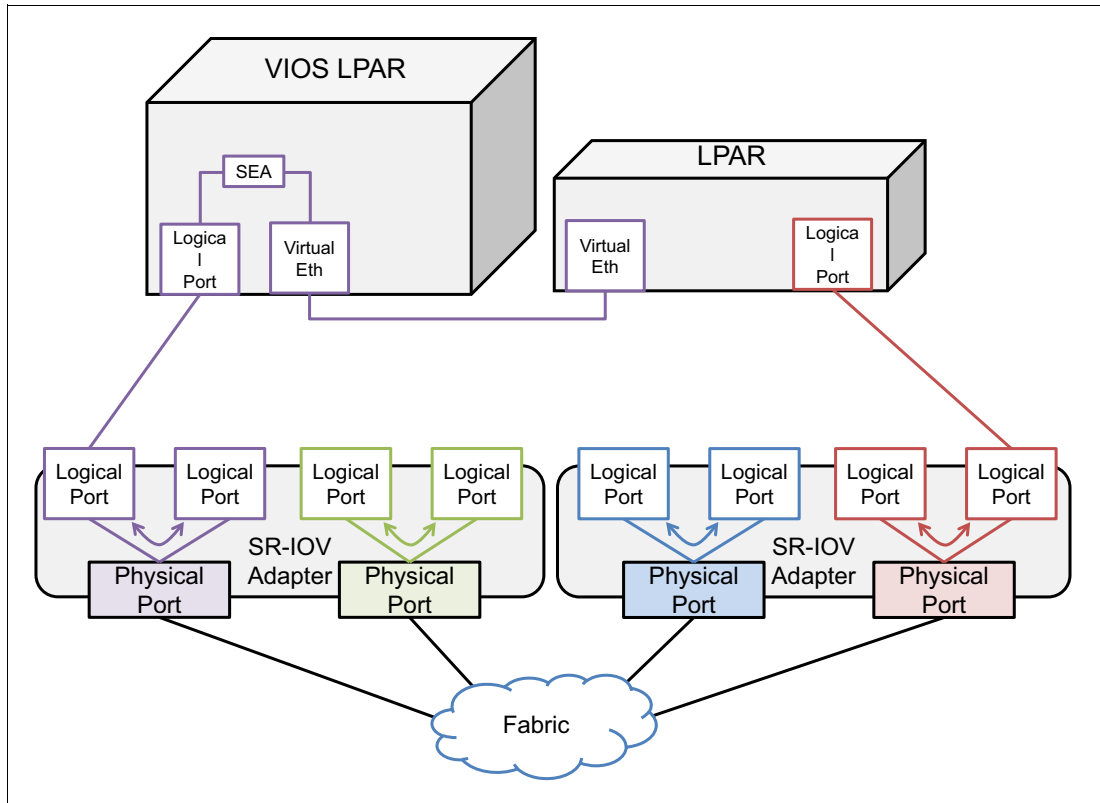


Figure 3-4 SR-IOV network interface backup

From a virtualization perspective, the SR-VIO logical ports are seen as physical adapters, therefore operations like Live Partition Mobility are not supported when an SR-IOV logical port is configured on the partition. On AIX, you can use the Network Interface Backup feature with a virtual adapter, together with a dynamic partition operation to remove the logical port from the partition in order to route the network traffic to the virtual adapter. You can then move the partition to another server by using Live Partition Mobility, and at the destination server reconfigure an SR-IOV port to the partition. Linux partitions can perform a similar operation using channel bonding to a virtual Ethernet adapter.

PowerVM vNIC combines many of the best features of SR-IOV and PowerVM SEA to provides a network solution with options for advanced functions such as Live Partition Mobility along with better performance and I/O efficiency when compared to PowerVM SEA. In addition PowerVM vNIC provides users with bandwidth control (QoS) capability by leverages SR-IOV logical ports as the physical interface to the network.

Figure 3-5 on page 16 shows a logical diagram of the vNIC structure and illustrates how the data flow is directly between LPAR memory and the SR-IOV adapter.

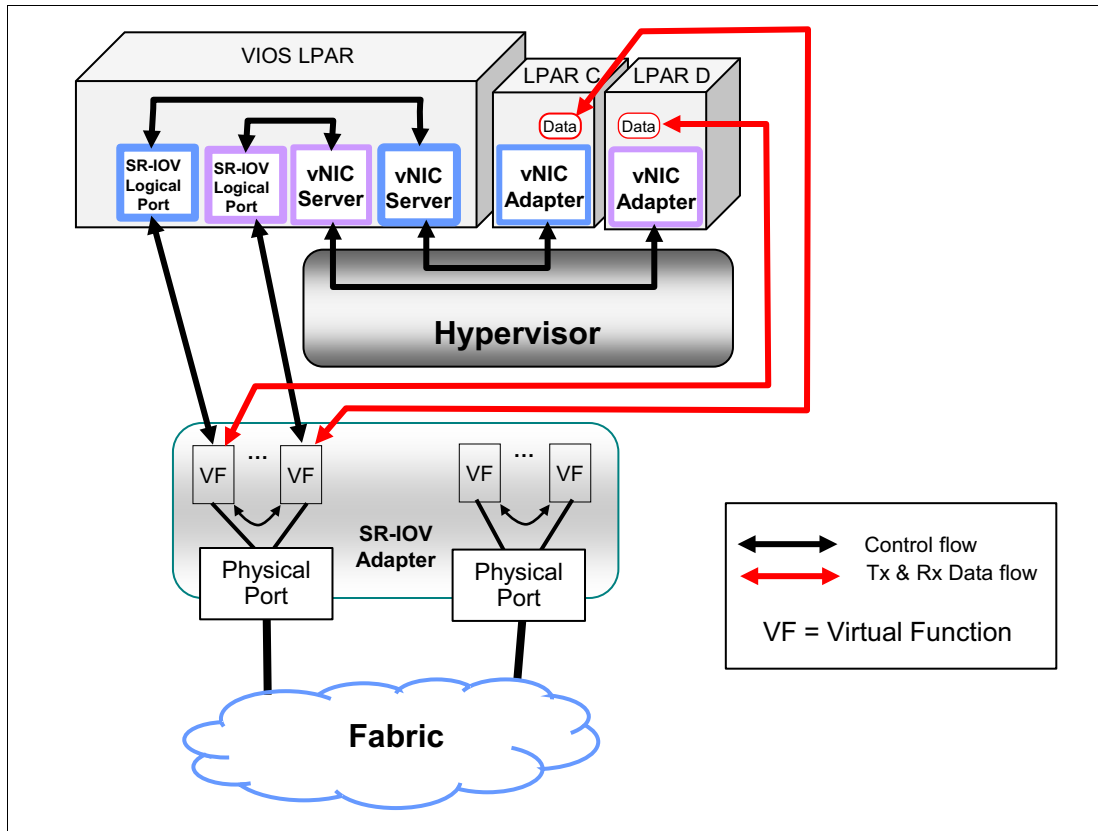


Figure 3-5 vNIC control and data flow to SR-IOV adapters

The key element of the vNIC model is a one-to-one mapping between a vNIC virtual adapter in the client LPAR and the backing SR-IOV logical port in the VIOS. With this model, packet data for transmission (similarly for receive) is moved from the client LPAR memory to the SR-IOV adapter directly without being copied to the VIOS memory. The benefits of bypassing the VIOS are reduction of the processing of a memory copy (specifically lower latency), and the reduction in the CPU and VIOS memory consumption (greater efficiency).

vNIC support can be added to a partition in a single step by adding a vNIC client virtual adapter to the partition using the management console (HMC). The management console creates all the necessary devices in the client LPAR as well as the VIOS. From the user perspective there is no additional configuration of the VIOS components for vNIC beyond configuration of the vNIC client adapter.

The minimum required PowerVM and operating system levels to support vNIC are as follows:

- ▶ PowerVM 2.2.4
 - VIOS Version 2.2.4
 - System Firmware Release 840
 - HMC Release 8 Version 8.4.0
- ▶ Operating Systems
 - AIX 7.1 TL4 or AIX 7.2
 - IBM i 7.1 TR10 or IBM i 7.2 TR3

A partition can be configured with up to six vNIC client adapters. If more than six vNIC client adapters are used in a partition, the partition will run, as there is no check to prevent the extra adapters, but certain operations such as Live Partition Mobility may fail.

3.4 Link aggregation

Link aggregation is a way to provide redundancy and additional bandwidth. You aggregate multiple adapters onto a single Ethernet interface that can be configured with TCP/IP, and the operating system distributes the traffic across the multiple adapters. If a failure occurs in one physical adapter, traffic is routed to the remaining functional adapters. For SR-IOV, you configure link aggregation between multiple SR-IOV logical ports by using the Link Aggregation Control Protocol (LACP).

To provide a valid LACP implementation the following configuration is supported:

- ▶ LACP (IEEE802.3ad, IEEE802.3ax) configured with multiple main logical ports. Those main logical ports can have only a single logical port configured per physical port.

On AIX and IBM i, only SR-IOV logical ports must be part of the link aggregation. Linux also supports link aggregation between an SR-IOV logical port and a Virtual Ethernet Adapter. With SR-IOV, using 802.3ad/802.1ax standards is possible.

On AIX, link aggregation LACP is part of the operating system since version AIX V5.1. IBM i introduced link aggregation in version i V7R1 TR3, and Linux implements link aggregation through the channel bonding driver.

If you decide to use SR-IOV logical ports in LACP configuration, remember that only one logical port per physical port can be used. Therefore, the best approach is to set the logical port capacity to 100%. This prevents users from adding a logical port to the physical port when LACP being used. Figure 3-6 shows the supported LACP configuration, with only one logical port assigned to a physical port.

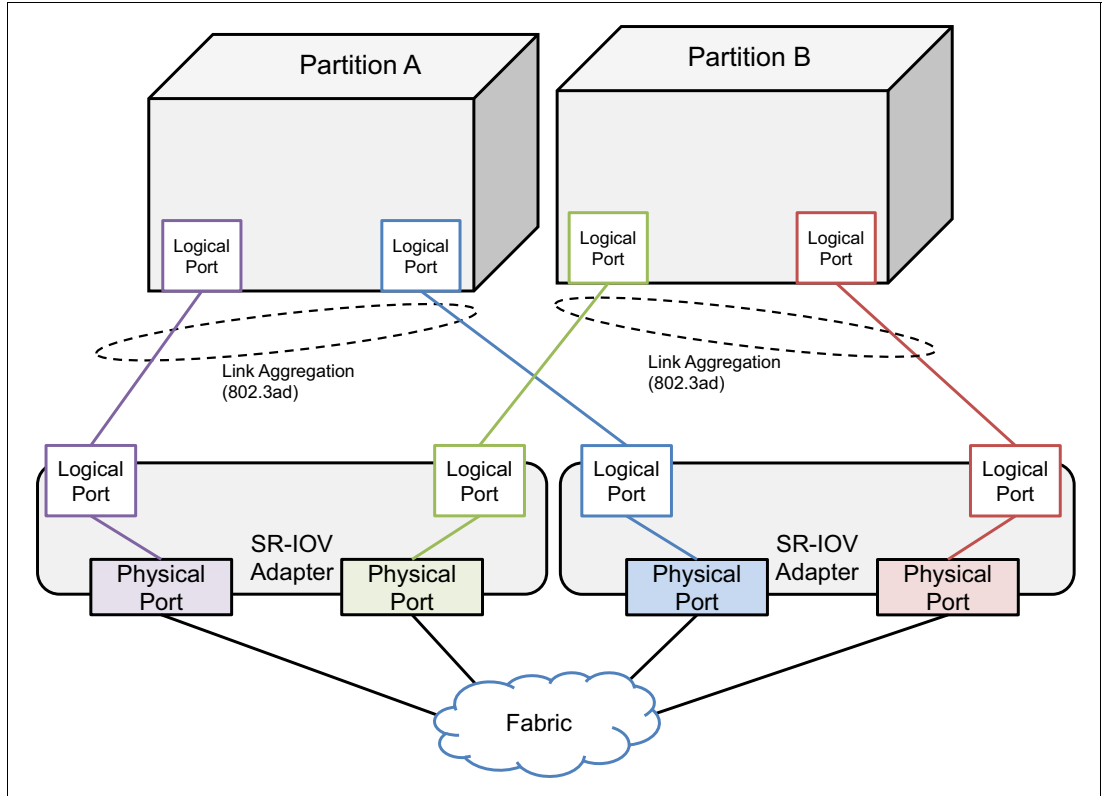


Figure 3-6 Supported LACP configuration with SR-IOV logical ports

Figure 3-7 shows an example of an invalid LACP configuration. This configuration, with more than one logical port assigned to a physical port, will not work.

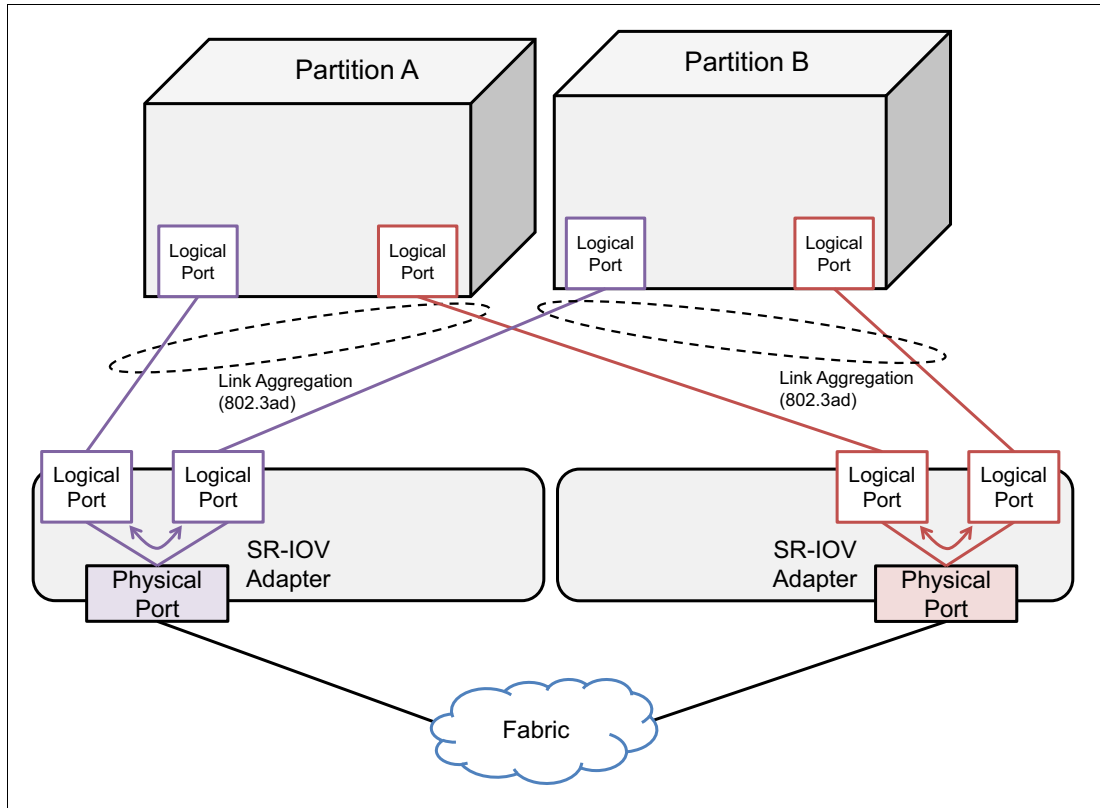


Figure 3-7 Invalid LACP configuration with many SR-IOV logical ports.

In an active-passive configuration, an SR-IOV logical port can be a primary (active) or backup (passive) adapter, or both. This can be achieved as follows:

- ▶ On AIX: By using the Network Interface Backup feature.
- ▶ On IBM i: By using Virtual IP Address (VIPA).
- ▶ On Linux: You configure an active-backup bonding.

Multiple primary SR-IOV logical ports are allowed in an LACP configuration. An SR-IOV logical port cannot be included as a primary adapter in an Etherchannel configuration with more than one primary adapter. When an SR-IOV logical port is configured in an active-passive configuration, it must be configured with the capability to detect when to fail over from the primary to the backup adapter.

- ▶ On AIX, configure a backup adapter and an IP address to **ping**.
- ▶ On IBM i with VIPA, options for detecting network failures besides link failures include Routing Information Protocol (RIP), Open Shortest Path First (OSPF) or customer monitor script.
- ▶ On Linux, use the bonding support to configure monitoring to detect network failures.



Configuration

This chapter describes various aspects of configuring an SR-IOV adapter.

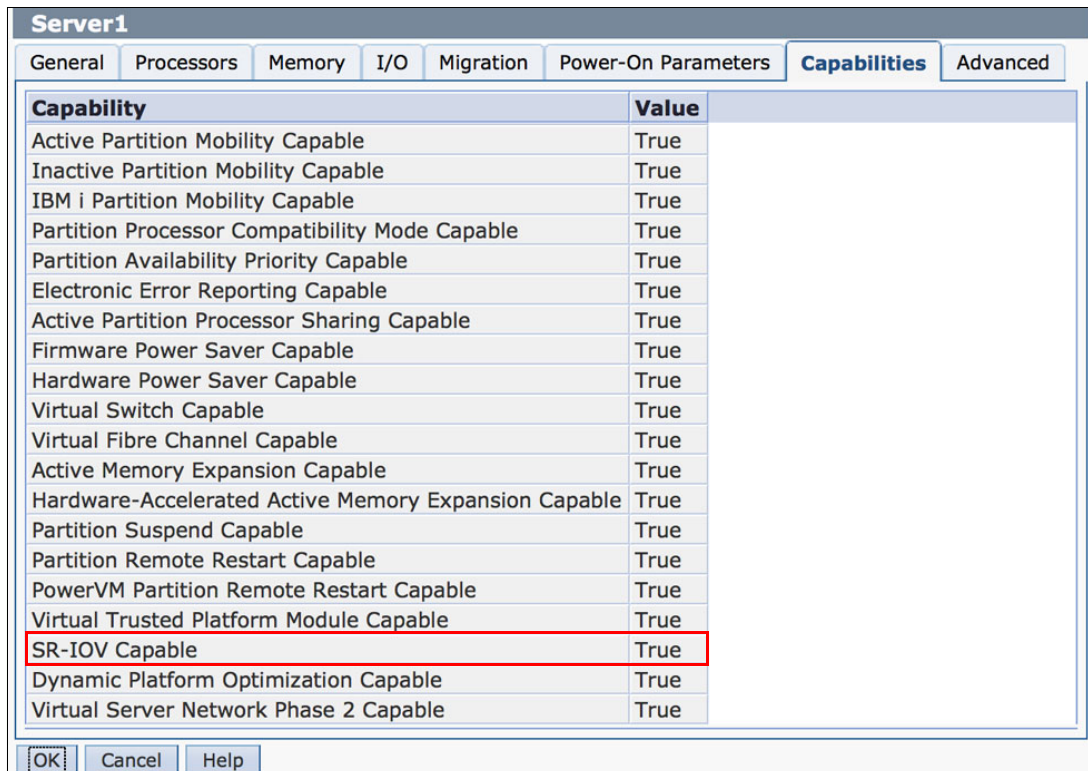
4.1 Verify prerequisites

Before configuring your SR-IOV capable adapter, verify that all prerequisites are met.

- ▶ Verify that the Power Systems server is SR-IOV Capable
- ▶ Verify that the system has at least one SR-IOV capable adapter

4.1.1 Verify that the Power Systems server is SR-IOV Capable

Log in to the Hardware Management Console and navigate to **Systems Management** → **Servers**. Select the appropriate server, and then click **Properties** from the **Tasks** menu. In the Properties panel, click the **Capabilities** tab (Figure 4-1).



Server1	
General Processors Memory I/O Migration Power-On Parameters Capabilities Advanced	
Capability	Value
Active Partition Mobility Capable	True
Inactive Partition Mobility Capable	True
IBM i Partition Mobility Capable	True
Partition Processor Compatibility Mode Capable	True
Partition Availability Priority Capable	True
Electronic Error Reporting Capable	True
Active Partition Processor Sharing Capable	True
Firmware Power Saver Capable	True
Hardware Power Saver Capable	True
Virtual Switch Capable	True
Virtual Fibre Channel Capable	True
Active Memory Expansion Capable	True
Hardware-Accelerated Active Memory Expansion Capable	True
Partition Suspend Capable	True
Partition Remote Restart Capable	True
PowerVM Partition Remote Restart Capable	True
Virtual Trusted Platform Module Capable	True
SR-IOV Capable	True
Dynamic Platform Optimization Capable	True
Virtual Server Network Phase 2 Capable	True

OK Cancel Help

Figure 4-1 Server Properties panel: Capabilities tab

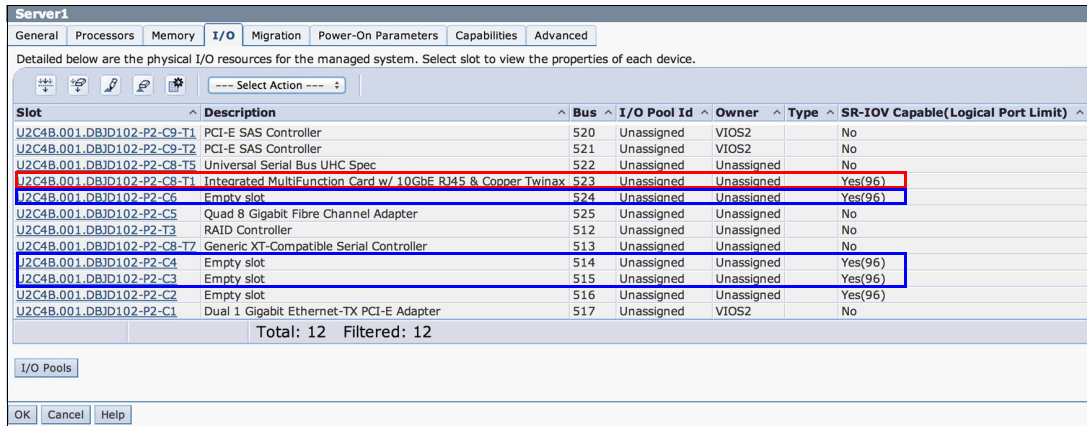
Verify that the SR-IOV Capable entry is set to True in the Value column.

If the SR-IOV Capable entry is not listed, the system most likely does not meet the hardware or firmware requirements that are described in 2.1, “SR-IOV hardware requirements and planning introduction” on page 8.

If the SR-IOV Capability is False, this means you are running PowerVM Express edition. With PowerVM Express the user can enable SR-IOV Shared mode but can create logical ports only for a single partition, and the adapter cannot be shared.

4.1.2 Verify that the system has at least one SR-IOV capable adapter

From the server Properties panel, click the **I/O** tab (Figure 4-2).



The screenshot shows the 'Server1' Properties panel with the 'I/O' tab selected. Below the tab are several icons and a 'Select Action' dropdown. The main area contains a table of I/O resources with the following columns: Slot, Description, Bus, I/O Pool Id, Owner, Type, and SR-IOV Capable(Logical Port Limit). The table lists 12 resources, including various controllers and adapters. Two rows are highlighted with blue boxes: 'U2C4B.001.DB1D102-P2-C6' (Empty slot) and 'U2C4B.001.DB1D102-P2-C3' (Empty slot), both showing 'Yes(96)' in the SR-IOV Capable column. A red box highlights the row 'U2C4B.001.DB1D102-P2-C8-T5' (Integrated MultiFunction Card w/ 10GbE R345 & Copper Twinax), which also shows 'Yes(96)'. The table footer indicates 'Total: 12 Filtered: 12'. Below the table is an 'I/O Pools' section and 'OK', 'Cancel', and 'Help' buttons.

Slot	Description	Bus	I/O Pool Id	Owner	Type	SR-IOV Capable(Logical Port Limit)
U2C4B.001.DB1D102-P2-C9-T1	PCI-E SAS Controller	520	Unassigned	VIOS2		No
U2C4B.001.DB1D102-P2-C9-T2	PCI-E SAS Controller	521	Unassigned	VIOS2		No
U2C4B.001.DB1D102-P2-C8-T5	Universal Serial Bus UHC Spec	522	Unassigned	Unassigned		No
U2C4B.001.DB1D102-P2-C8-T1	Integrated MultiFunction Card w/ 10GbE R345 & Copper Twinax	523	Unassigned	Unassigned		Yes(96)
U2C4B.001.DB1D102-P2-C6	Empty slot	524	Unassigned	Unassigned		Yes(96)
U2C4B.001.DB1D102-P2-C5	Quad 8 Gigabit Fibre Channel Adapter	525	Unassigned	Unassigned		No
U2C4B.001.DB1D102-P2-T3	RAID Controller	512	Unassigned	Unassigned		No
U2C4B.001.DB1D102-P2-C8-T7	Generic XT-Compatible Serial Controller	513	Unassigned	Unassigned		No
U2C4B.001.DB1D102-P2-C4	Empty slot	514	Unassigned	Unassigned		Yes(96)
U2C4B.001.DB1D102-P2-C3	Empty slot	515	Unassigned	Unassigned		Yes(96)
U2C4B.001.DB1D102-P2-C2	Empty slot	516	Unassigned	Unassigned		Yes(96)
U2C4B.001.DB1D102-P2-C1	Dual 1 Gigabit Ethernet-TX PCI-E Adapter	517	Unassigned	VIOS2		No

Figure 4-2 Server Properties panel: I/O tab

Verify that you have at least one adapter with a value of Yes in the SR-IOV Capable column.

The number in parentheses in the last column indicates the maximum number of logical ports that the slot supports. However, if an SR-IOV capable adapter has been installed and set to shared mode, the number indicates the maximum logical ports that the adapter supports.

If no SR-IOV capable adapters are installed, you will need to install one into an empty slot that has a value of “Yes” in the SR-IOV Capable column.

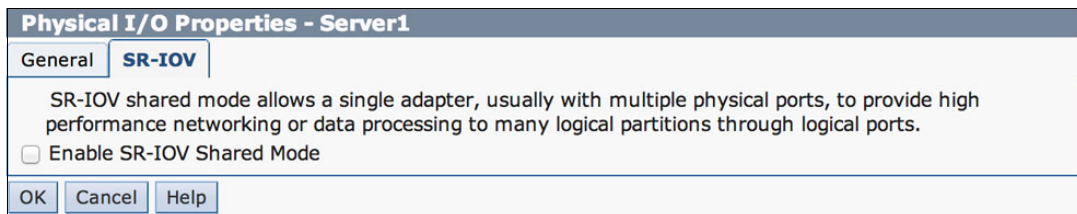
4.2 Adapter operations

When you have an SR-IOV capable system and adapter, the next step is to enable SR-IOV shared mode on the adapter to share the physical ports with different partitions. If you do not have more than one partition, or if you do not want to share the adapter, it can remain in Dedicated mode.

4.2.1 Switching adapter from dedicated mode to SR-IOV shared mode

Navigate to **Systems Management** → **Servers**. Select the appropriate server, and click **Properties** from the **Tasks** menu. In the Properties panel, click the **I/O** tab.

Click the link of an SR-IOV capable adapter in the I/O table. The Physical I/O Properties panel opens (Figure 4-3). Click the **SR-IOV** tab.



The screenshot shows the 'Physical I/O Properties - Server1' panel with the 'SR-IOV' tab selected. The panel contains a text box explaining SR-IOV shared mode: 'SR-IOV shared mode allows a single adapter, usually with multiple physical ports, to provide high performance networking or data processing to many logical partitions through logical ports.' Below this text is a checkbox labeled 'Enable SR-IOV Shared Mode', which is currently unchecked. At the bottom of the panel are 'OK', 'Cancel', and 'Help' buttons.

Figure 4-3 SR-IOV capable adapter properties: dedicated mode (shared mode not enabled)

Select the **Enable SR-IOV Shared Mode** check box, and then click **OK**. To see the change, exit the server Properties panel and then reopen it.

As Figure 4-4 shows, the adapter is now owned by the hypervisor, and the SR-IOV Capable column is updated to display the logical port limit of the adapter.

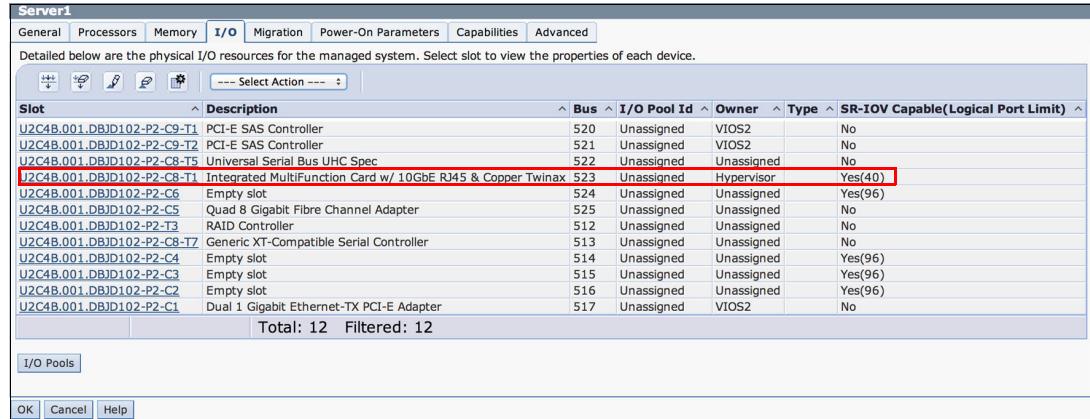


Figure 4-4 Server Properties panel: I/O tab with SR-IOV shared mode enabled

4.2.2 Switching adapter from SR-IOV shared mode to dedicated mode

To switch the adapter back to dedicated mode, remove all logical ports from their respective partitions, deselect the **Shared Mode** check box on the SR-IOV tab of the adapter properties, and then click **OK** (Figure 4-5).

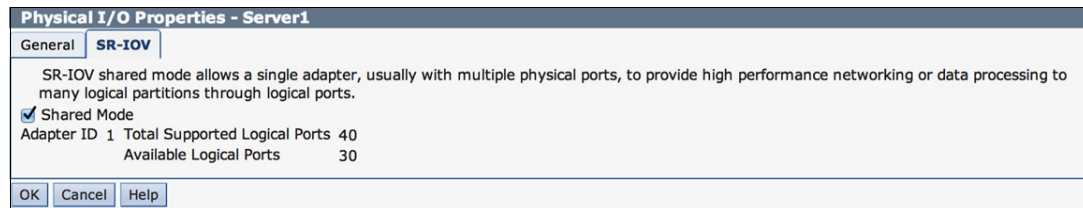


Figure 4-5 SR-IOV adapter properties: Shared Mode enabled

If active or inactive partitions are using the logical ports of the adapter, the mode change operation will fail.

For logical ports in use by active partitions, select the active partition, click **Dynamic Partitioning** → **SR-IOV Logical Ports** from the **Tasks** menu, and dynamically remove the logical ports.

If only inactive partitions are using the logical ports of the adapter when the mode switch operation is performed, the operation will fail, and the Release SR-IOV Logical Ports panel opens (Figure 4-6). Release the logical ports from the inactive partitions by clicking **Select Action** → **Select All**, and then click **OK**. After all logical ports are released, try the mode switch operation again.

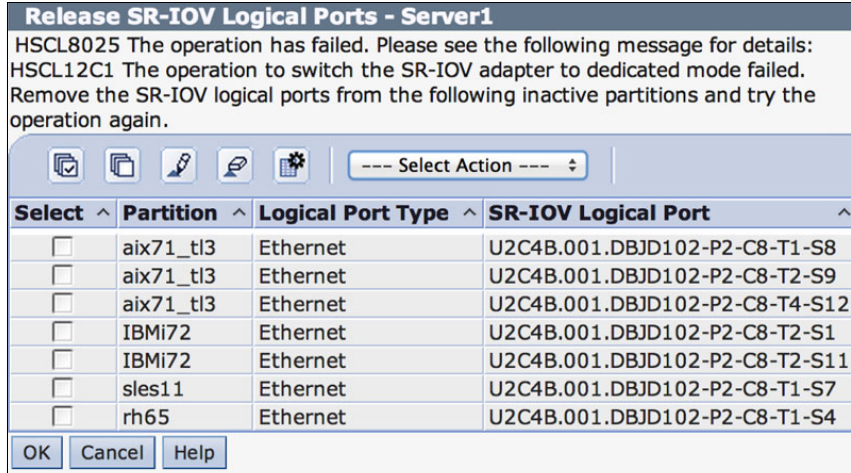


Figure 4-6 Release SR-IOV Logical Ports panel

4.3 Physical port operations

To configure the physical ports of an SR-IOV capable adapter, select the appropriate server, and then click **Properties** from the **Tasks** menu. On the I/O tab of the Properties panel, click the link for the adapter you want to configure, and then click the **SR-IOV** tab (Figure 4-7). Here you can select the link for the physical port you want to view or modify. Alternately, you can select the radio button of a physical port to see the logical ports configured on it.

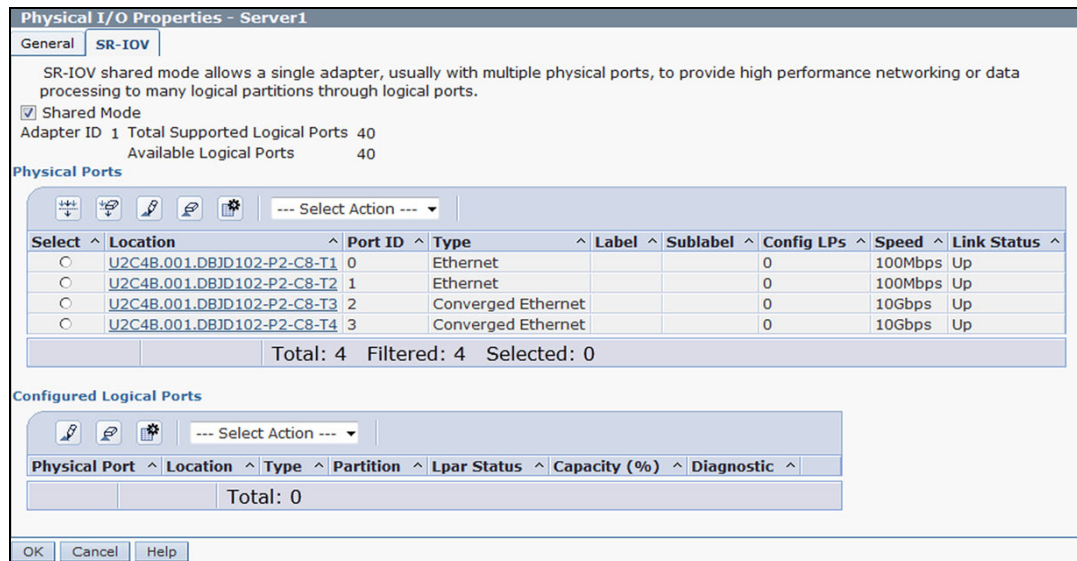


Figure 4-7 SR-IOV adapter Physical I/O Properties: SR-IOV tab

The Physical Port Properties panel has three tabs: General, Advanced, and Port Counters.

4.3.1 Physical port properties: General

The General tab (Figure 4-8) has the following properties and settings.

Label and Sublabel These port labels are for your reference so that the physical ports are easier to identify; they can be set to anything you choose. They will be displayed in their respective columns on the Physical I/O Properties panel. (Figure 4-7 on page 25)

Capacity This section shows the available capacity for this physical port. The total capacity is always 100.

Configurable Negotiated Properties

Here you can set configured properties, like port speed. However, the actual speed will ultimately be determined by the adapter and will be shown in the Negotiated column.

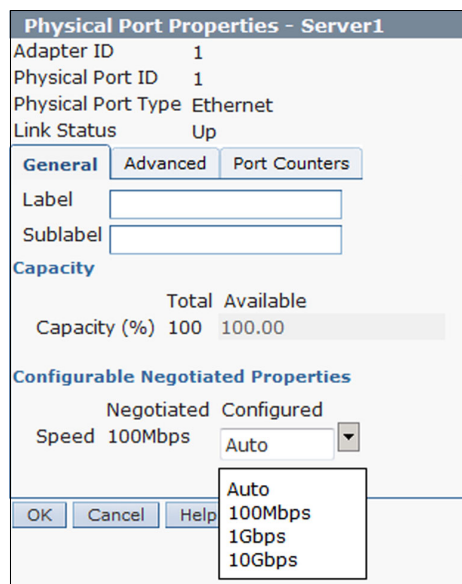


Figure 4-8 Physical Port Properties panel: General tab

4.3.2 Physical port properties: Advanced

The Advanced tab (Figure 4-9 on page 27) is where you configure the flow control, MTU size, port switch mode, and maximum number of logical ports permitted for that physical port.

The Advanced tab also shows the maximum number of diagnostic logical ports and promiscuous logical ports, and how many of each are configured.

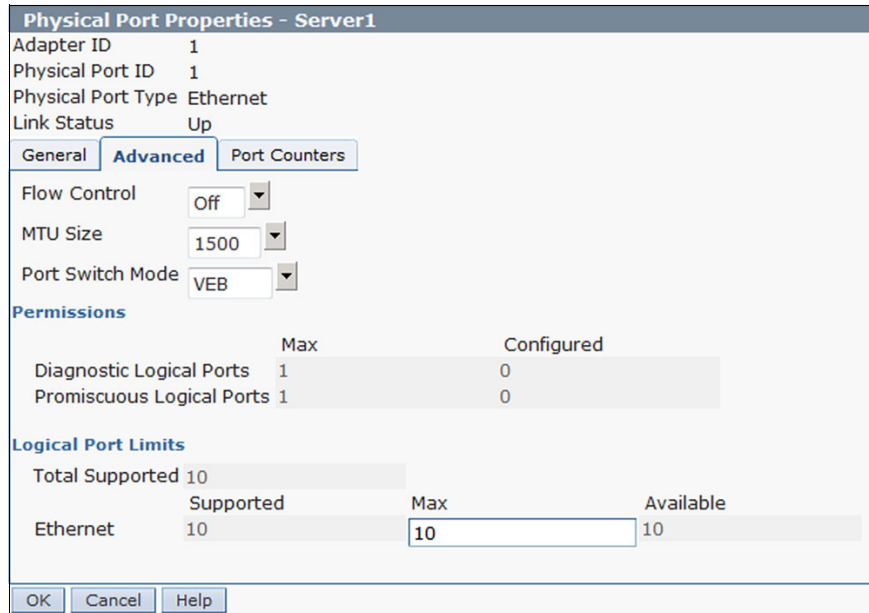


Figure 4-9 Physical Port Properties panel: Advanced tab

The Advanced tab has following properties and settings:

- Flow Control** This read-only property specifies the *priority* flow control of the physical port. If the physical port does not support priority flow control, this field is not displayed. The HMC has no control over this value.
- Flow Control** This drop-down list specifies the state of the flow control capability of the physical port. The flow control setting is effective only when the state of the priority Flow Control is set to Off or not displayed.
- MTU Size** This drop-down list has supported maximum transmission unit (MTU) size, retrieved from the hypervisor. MTU size indicates the largest possible unit of data that can be sent in a single frame.
- Port Switch Mode** The two modes available are Virtual Ethernet Bridge (VEB) and Virtual Ethernet Port Aggregator (VEPA):
- VEB** This is the default port switch mode. Bridging between logical ports (VFs) on the same physical port is done by the adapter. Logical port to logical port traffic is not exposed to an external switch, resulting in lower latency for this type of traffic.
 - VEPA** Bridging between logical ports on the same physical port is done by an external switch. Logical port to logical port traffic flows out of the physical port and is reflected back to the physical port by the external switch. The external switch must support Reflective Relay (also known as hairpin turn) and is manually enabled for the switch port. This configuration allows an external switch to inspect all network traffic, for example by a firewall.

Note: Port switch mode (VEB/VEPA) can be changed only when no logical ports are configured on the physical port.

Permissions This specifies the maximum and configured number of logical ports that are in the diagnostic or promiscuous mode.

Logical Port Limits This specifies the total, maximum, and available number of configured logical ports that are supported by the system firmware for the selected physical port.

Total Supported

The total number of logical ports supported for all protocol types on the physical port.

Supported

The number of logical ports of the specified type (for example, Ethernet) supported by the physical port. The physical port may have more logical port limits based on the logical port type.

Max

The maximum number of logical ports, of that type, that can be allocated to logical partitions. The value is either a default value determined by the hypervisor or a user-specified value that is no higher than the value in the respective “Supported” column.

Note: The logical port limits Max value can be changed only when no logical ports are configured on the adapter, including those on other physical ports.

4.3.3 Physical port properties: Port Counters

The Port Counters tab (Figure 4-10) shows statistics for the physical port. These values are collected from the hypervisor each time the Physical Port Properties panel is opened or if you click **Refresh**. Also, you can click **Clear Statistics** to reset the counter values for the physical port.

Physical Port Properties - Server1	
Adapter ID	1
Physical Port ID	0
Physical Port Type	Ethernet
Link Status	Up
<input type="button" value="General"/> <input type="button" value="Advanced"/> <input type="button" value="Port Counters"/>	
Name	Value
Transmit Packets	736875
Transmit Unicast Packets	736031
Transmit Multicast Packets	788
Transmit Broadcast Packets	56
Transmit Bytes	1100506358
Transmit Dropped	0
Transmit Errors	0
Transmit Pause Off Frames	0
Transmit Pause On Frames	0
Transmit Lost Carrier	0
Transmit Underruns	0
Transmit Lost CTS	0
Transmit Defers	0
Transmit Single Collisions	0
Transmit Multiple Collisions	0
Transmit Excess Collisions	0
Transmit Late Collisions	0
Received Packets	13057463
Received Unicast Packets	671305
Received Multicast Packets	3280678
<input type="button" value="Refresh"/> <input type="button" value="Clear Statistics"/>	
<input type="button" value="OK"/> <input type="button" value="Cancel"/> <input type="button" value="Help"/>	

Figure 4-10 Physical Port Properties panel: Port Counters

4.4 Logical port operations

You can create an SR-IOV logical port through the SR-IOV Logical Ports panel during partition creation or by adding one to an existing partition profile.

4.4.1 Adding a logical port during partition creation

In the Create Lpar Wizard window (Figure 4-11), if you have at least one SR-IOV capable adapter in shared mode, you see a panel named SR-IOV Logical Ports.

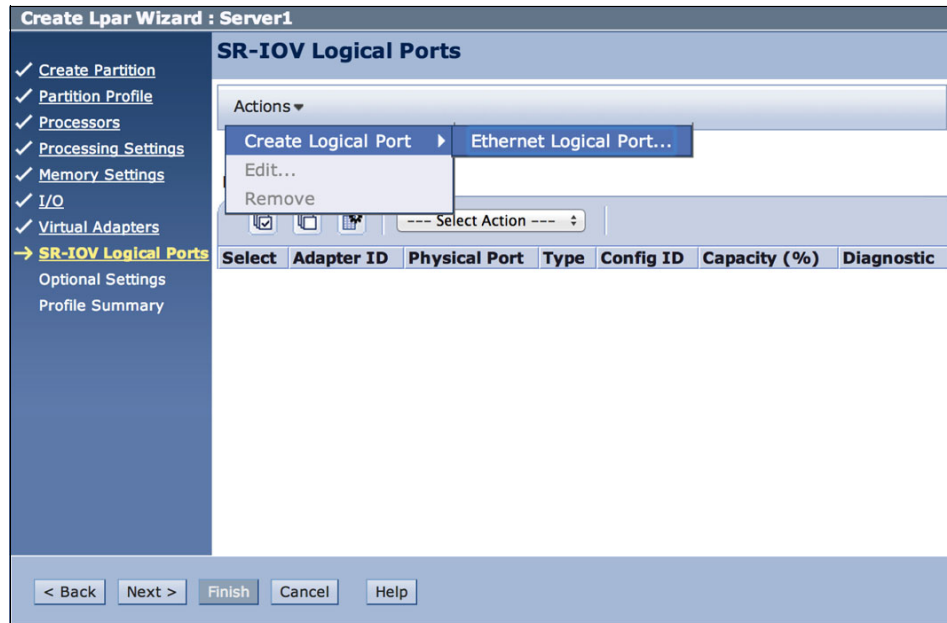


Figure 4-11 Create Lpar Wizard: SR-IOV Logical Ports

Click **Actions** → **Create Logical Port** and then select the type of logical port you want to configure. The HMC displays a panel that lists all available SR-IOV physical ports of that type in the system (Figure 4-12).

Select	Adapter Id	Physical Port	Label	Sublabel	Speed	Active LPs	Available LPs	Link Status
<input checked="" type="radio"/>	1	U2C4B.001.DBJD102-P2-C8-T1			100Mbps	5	5	Up
<input type="radio"/>	1	U2C4B.001.DBJD102-P2-C8-T2			100Mbps	3	7	Up
<input type="radio"/>	1	U2C4B.001.DBJD102-P2-C8-T3			10Gbps	1	9	Up
<input type="radio"/>	1	U2C4B.001.DBJD102-P2-C8-T4			10Gbps	1	9	Up

Figure 4-12 Add Ethernet Logical Port panel

Select the radio button of a physical port that has at least one available logical port (LP), then click **OK**, which then takes you to the **General** tab of the Logical Port Properties panel.

Logical port properties: General

On the Logical Port Properties panel (Figure 4-13), define the capacity for this logical port and whether the logical port should operate in diagnostic mode, promiscuous mode, or both modes.

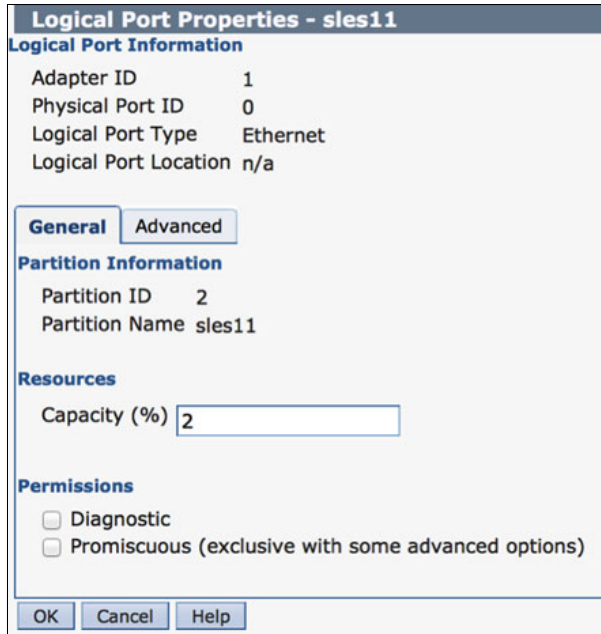


Figure 4-13 Logical Port Properties panel: General tab

The General tab has following properties and settings:

Capacity

The percentage of the physical port's resources that should be allocated to this logical port. The value must be a multiple of the default value, which is 2.0%, or the HMC returns an error.

One of the physical port resources that the Capacity setting determines is the logical port's *desired* minimum bandwidth as a percentage of the physical ports Negotiated bandwidth. The value selected for the Capacity setting is used as the percentage of the physical port's bandwidth to assign to the logical port as its *desired* minimum bandwidth.

A port does not have its bandwidth capped by the Capacity setting if there is additional bandwidth that is not currently being used. Capacity settings do not apply to received traffic, only to transmitted traffic. Any unused bandwidth, whether unassigned or unused, on a physical port will be shared equally among all logical ports.

Diagnostic mode

Enables additional diagnostics. This should be set only when you are running diagnostics on the adapter, because it can disrupt the adapter I/O traffic. Diagnostic mode can be enabled only if no other logical ports are allocated on the same physical port.

Promiscuous mode

This is primarily used when the logical port will be further virtualized. For example, if the logical port will be assigned to a Virtual I/O Server for use in a Shared Ethernet Adapter, then enabling promiscuous mode is required.

Promiscuous mode can be enabled on only one logical port per physical port.

The promiscuous mode logical port receives frames whose destination MAC addresses do not match the addresses of any other logical ports on the same physical port (traffic from logical ports on other physical ports cannot be *sniffed*).

Figure 4-14 shows the error message when performing a dynamic addition of a logical port that exceeds the maximum capacity.

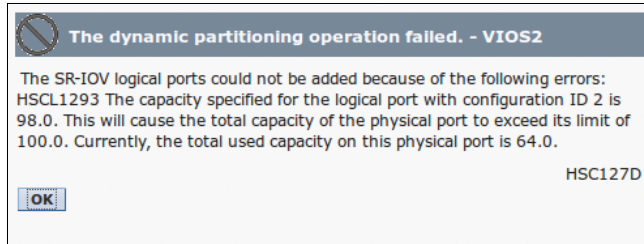


Figure 4-14 Error message when dynamically adding a logical port, exceeding capacity

Logical port properties: Advanced

Figure 4-15 shows the Logical Port Properties Advanced tab settings.

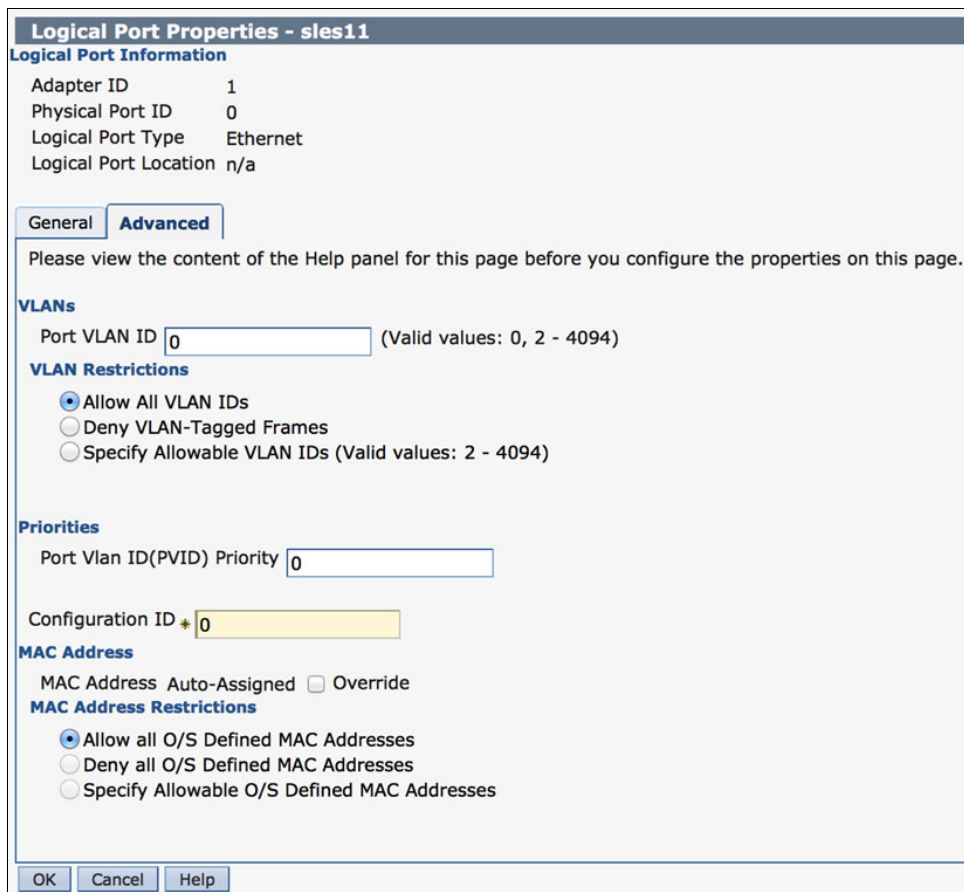


Figure 4-15 Logical Port Properties panel: Advanced tab

The Advanced tab additional properties, of the logical port, that you can modify.

Port VLAN ID Valid values are 0 and 2 - 4094.

VLAN Restrictions When promiscuous mode is enabled, the **Allow All VLAN IDs** setting is the only option available. Otherwise, you can choose whether to allow all, deny all, or allow a specified range of VLAN IDs.

Port Vlan ID (PVID) Priority
Valid values are 0 - 7.

Configuration ID This is similar to a virtual slot ID for virtual Ethernet and should typically be kept at its default value, which is assigned by the HMC.

MAC Address By default, the MAC address is auto-assigned by the HMC. To define a specific MAC address, select the **Override** check box, which enables an additional field where you can enter your MAC address.

MAC Address Restrictions
When promiscuous mode is enabled, **Allow all O/S Defined MAC Addresses** is the only option available. Otherwise, you can choose whether to allow all, deny all, or allow specific operating system defined MAC addresses.

After you set the attributes of your logical port and click **OK**, you are returned to the Create LPAR Wizard, where you can create more logical ports or complete the profile creation.

Note: Logical ports that are defined during profile creation are not configured on the adapter, and no validation of resource availability or conflict is done, until the profile is activated. During activation, HMC performs validation checks to verify resource availability. If any one of the logical port resources is not available, the activation fails. When the profile has been activated, select the logical partition and then navigate to **Properties** → **SR-IOV Logical Ports** in the **Tasks** menu to see the SR-IOV logical ports.

4.4.2 Adding a logical port using dynamic partitioning

You can also use dynamic logical partitioning (DLPAR) to add an SR-IOV logical port to a running partition.

Note: For AIX, Linux, and VIOS partitions, the DLPAR add and remove operations require a working Resource Monitoring and Control (RMC) connection between the partition and the HMC.

Select a running logical partition and navigate to **Dynamic partitioning** → **SR-IOV Logical Ports** in the **Tasks** menu to reach the SR-IOV Logical Ports panel (Figure 4-16).

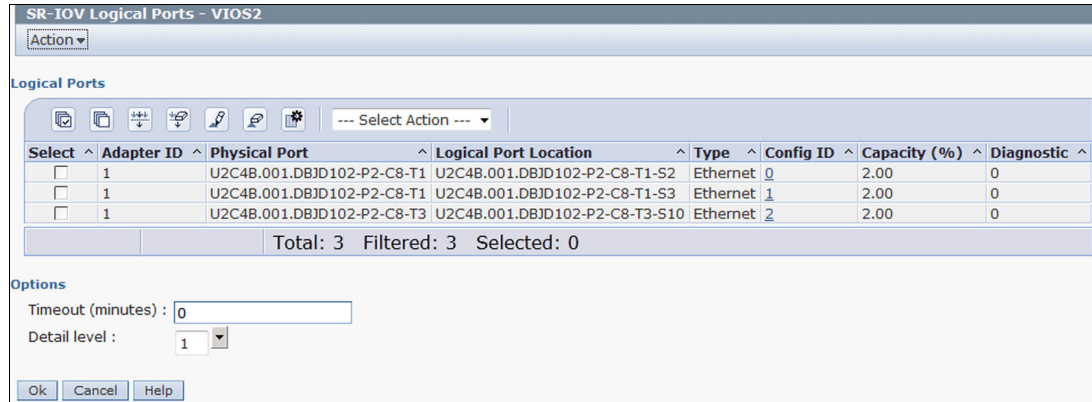


Figure 4-16 Dynamic partitioning: SR-IOV Logical Ports panel

From the SR-IOV Logical Ports panel, select **Action** → **Add Logical Port** and choose the type of logical port you want to add. Next, you are prompted to select the physical port from which you want to create the logical port. Click the appropriate radio button, and click **OK**.

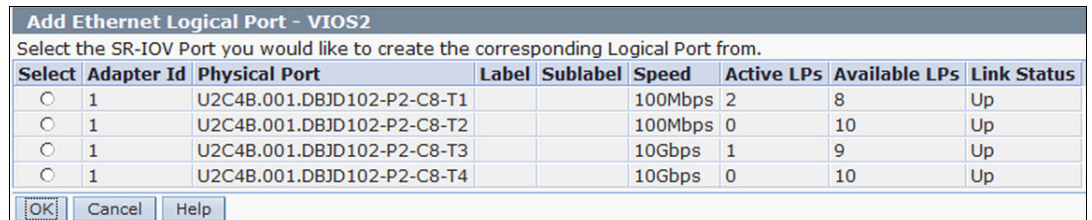


Figure 4-17 Add Ethernet Logical Port panel: select physical port

The Logical Port Properties panel opens (described in section 4.4.1, “Adding a logical port during partition creation” on page 29). Complete the configuration details from there.

When the DLPAR add operation is complete on the HMC, you might need to act on the partition to make the operating system aware of the new device.

- ▶ For an AIX partition, run **cfgmgr**.
- ▶ On the VIOS, use the **cfgdev** command.
- ▶ IBM i and Linux dynamically reconfigure and require no additional action.

Note: Remember to save all DLPAR operations to the HMC partition profile if you want the changes to remain for future activations.

4.4.3 Editing a logical port

To view or edit the properties of an existing logical port, use one of the following paths to reach the Logical Port Properties panel.

- ▶ Select the partition, then from the **Tasks** menu, click **Dynamic partitioning** → **SR-IOV Logical Ports**, then **Action** → **Edit Logical Port**. Click the hotlink of the logical port you want to edit.
- ▶ Select the partition, then from the **Tasks** menu, click **Properties**, then click the **SR-IOV Logical Ports** tab. Click the hotlink of the logical port you want to edit.

- Select the server, then from the **Tasks** menu, click **Properties**, and then click the **I/O** tab. Next, click the link for the appropriate SR-IOV adapter, and click the **SR-IOV** tab. Click the radio button by the appropriate physical port to see the associated logical ports (Figure 4-18). Then click the link of the location for the logical port that you want to edit.

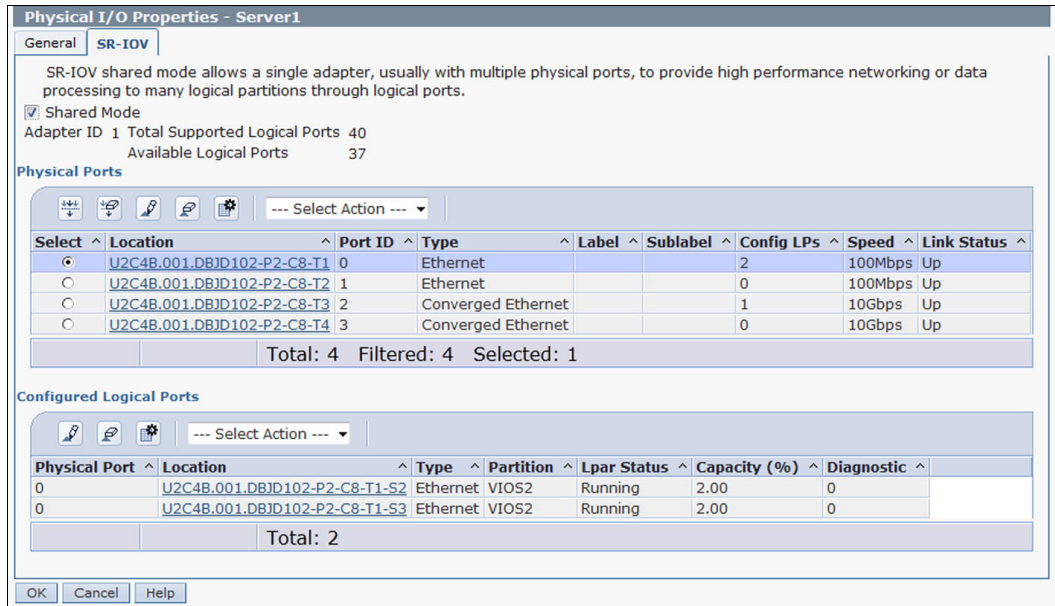


Figure 4-18 Server Physical I/O Properties panel: Configured Logical Ports

The Logical Port Properties panel will be similar to the panel in Figure 4-13 on page 30, but some fields might be read-only. Figure 4-19 shows that Diagnostics can still be enabled or disabled, but you can no longer modify the Capacity value.

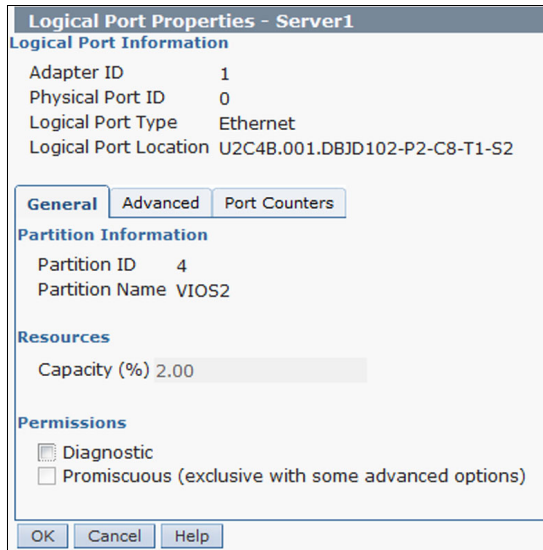


Figure 4-19 Logical Port Properties: General tab from a running partition

Some properties on the Advanced tab (Figure 4-20 on page 35) have limitations also.

- VLAN Restrictions and MAC Address Restrictions cannot be changed.
- If the VLANs list is non-empty, you can add to the list, but you cannot remove VLAN IDs from the list. The same rule applies to “MAC Addresses”

- ▶ If the VLAN is non-empty, you cannot change Port VLAN ID from zero to a non-zero value or from a non-zero value to zero, but you can change it from non-zero to another non-zero value.

The screenshot shows the 'Logical Port Properties - Server1' dialog box with the 'Advanced' tab selected. The 'Logical Port Information' section displays: Adapter ID 1, Physical Port ID 0, Logical Port Type Ethernet, and Logical Port Location U2C4B.001.DBJD102-P2-C8-T1-S2. The 'VLANs' section has a 'Port VLAN ID' field set to 0 with a note '(Valid values: 0, 2 - 4094)'. The 'VLAN Restrictions' section has three radio buttons: 'Allow All VLAN IDs' (selected), 'Deny VLAN-Tagged Frames', and 'Specify Allowable VLAN IDs (Valid values: 2 - 4094)'. The 'Priorities' section has a 'Port Vlan ID(PVID) Priority' field set to 0. The 'Configuration ID #' field is also set to 0. The 'MAC Address' section shows 'MAC Address 2e840a550200'. The 'MAC Address Restrictions' section has three radio buttons: 'Allow all O/S Defined MAC Addresses' (selected), 'Deny all O/S Defined MAC Addresses', and 'Specify Allowable O/S Defined MAC Addresses'. At the bottom are 'OK', 'Cancel', and 'Help' buttons.

Figure 4-20 Logical Port Properties: Advanced tab from a running partition

4.5 Device mapping

The **SR-IOV End-to-End Mapping** task has been added under the **Hardware Information** → **Adapters** menu at the server level (Figure 4-21).

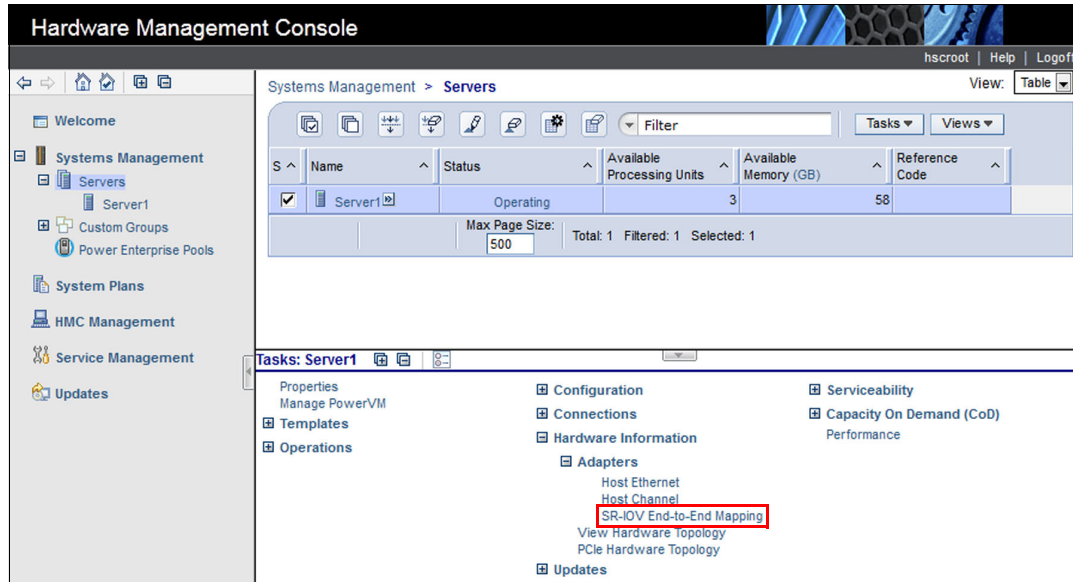


Figure 4-21 SR-IOV End-to-End Mapping menu option

This task launches a panel that lists the SR-IOV physical ports of the system (Figure 4-22). Select the radio button of a physical port to view the device mapping between the configured logical ports and the operating system devices.

SR-IOV Device Mappings - Server1							
Select	Physical Port	Type	Port ID	Configured LPs	Available LPs	Speed	Link Status
<input type="radio"/>	U2C4B.001.DBJD102-P2-C8-T1	Ethernet	0	5	5	100Mbps	Up
<input checked="" type="radio"/>	U2C4B.001.DBJD102-P2-C8-T2	Ethernet	1	3	7	100Mbps	Up
<input type="radio"/>	U2C4B.001.DBJD102-P2-C8-T3	Converged Ethernet	2	1	9	10Gbps	Up
<input type="radio"/>	U2C4B.001.DBJD102-P2-C8-T4	Converged Ethernet	3	1	9	10Gbps	Up

Logical Partition	Location	Device Name
IBMi72	U2C4B.001.DBJD102-P2-C8-T2-S1	CMN03
aix71_t13	U2C4B.001.DBJD102-P2-C8-T2-S9	ent4
IBMi72	U2C4B.001.DBJD102-P2-C8-T2-S11	CMN05

Close Help

Figure 4-22 SR-IOV Device Mappings

If the owner partition is running AIX or Linux, an RMC connection is required to see the operating system device names. Without RMC, Device Name will show Unknown.

End-to-end mapping is similar in IBM i but does not require an RMC connection.

4.6 Operating system adapter configuration

From an operating system perspective, logical ports appear as regular Ethernet adapters. Each operating system has device drivers to configure the logical ports and present them to the applications as Ethernet devices.

4.6.1 AIX

On AIX, the logical port is seen as a physical Ethernet device, as shown in Example 4-1. The VF string at the end of the description of the adapter indicates that it is a logical port (Virtual Function).

Example 4-1 SR-IOV port configuration on AIX

```
# lsdev -Ccadapter

ent0 Available 00-00 10GbE SFP+ Cu 4-port Integrated Multifunction CNA VF
(df1028e214103904)
```

The **lscfg** command lists the physical adapter characteristics, as shown in Example 4-2. The Hardware Location Code attribute shows the location of the physical adapter, and also the physical port that the logical port is attached to.

Example 4-2 lscfg output showing port information

```
# lscfg -v1 ent0
ent0 U2C4B.001.DBJD102-P2-C8-T4-S12 10GbE SFP+ Cu 4-port Integrated
Multifunction CNA VF (df1028e214103904)
```

```
Ethernet Adapter:
Network Address.....2E8409671802
ROM Level.(alterable).....0.0.9999.19068
Hardware Location Code.....U2C4B.001.DBJD102-P2-C8-T4-S12
```

The **lsattr** command lists the attributes of the adapter. Example 4-3 shows some of the attributes from the command output.

Example 4-3 Some attributes of the SR-IOV logical port

```
# lsattr -El ent0
...

jumbo_frames no Request jumbo frames True
jumbo_size 9000 Requested jumbo frame size True
large_receive yes Request Rx TCP segment aggregation True
large_send yes Request Tx TCP segment offload True
...
media_speed 10000_Full_Duplex Requested Media speed True
...
```

The media speed shows the speed that the physical port is configured to use. If your adapter is set to auto negotiation, `media_speed` in the **lsattr** output indicates `Auto_Negotiation`. The virtual port has its speed based on the physical port speed, and the capacity as configured through the HMC.

By default, the HMC configures the MTU size for the adapter, and whether to use jumbo frames. You can change the configuration by using the `chdev` command (Example 4-4).

Example 4-4 Enabling jumbo frames on AIX

```
# lsattr -El ent0 |grep jumbo
jumbo_frames    no                Request jumbo frames      True
jumbo_size       9000                Requested jumbo frame size True
# chdev -l ent0 -a jumbo_frames=yes
ent0 changed
# lsattr -El ent0 |grep jumbo
jumbo_frames    yes                Request jumbo frames      True
jumbo_size       9000                Requested jumbo frame size True
```

The `entstat` command shows more information about the adapter. Example 4-5. highlights some information from the command output that is related to the logical and physical port configuration.

Example 4-5 The entstat command showing SR-IOV port information

```
# entstat ent0
-----
ETHERNET STATISTICS (ent0) :
Device Type: 10GbE SFP+ CU Integrated Multifunction CNA 10GbE GX++ Gen2 Converged
Network Adapter (df1028e214103904)
Hardware Address: 2e:84:09:67:18:02
Elapsed Time: 0 days 0 hours 0 minutes 1 seconds
...

General Statistics:
-----
No mbuf Errors: 0
Adapter Reset Count: 4
Adapter Data Rate: 20000
Driver Flags: Up Broadcast Running
                Simplex Promiscuous 64BitSupport
                ChecksumOffload LargeSend DataRateSet
                IPV6_LSO IPV6_CSO LARGE_RECEIVE
                VIRTUAL_PORT PHYS_LINK_UP
10GbE SFP+ CU Integrated Multifunction CNA 10GbE GX++ Gen2 Converged Network
Adapter
-----
Device ID: df1028e214103904
Version: 1
Device State: Open
Physical Port Link Status: Up
Logical Port Link Status: Up
Physical Port Speed: 10 Gbps Full Duplex
...
Physical Port Promiscuous Mode Changeable: No
Physical Port Promiscuous Mode: Disabled
Logical Port Promiscuous Mode: Enabled
Physical Port All Multicast Mode Changeable: Yes
Physical Port All Multicast Mode: Disabled
Logical Port All Multicast Mode: Disabled
...
```

```

Physical Port MTU Changeable: No
Physical Port MTU: 1514
Logical Port MTU: 1514
Jumbo Frames: Disabled
...
Adapter Logical Port Counters:
  Total Number of Packets Sent: 222
  Number of Unicast Packets Sent: 190
  Number of Multicast Packets Sent: 28
  Number of Broadcast Packets Sent: 4
  Total Bytes Sent: 25845
...
DCBX Status: Enabled
  TLV Exchange Complete
  TC State [0x00000009] Enabled
  PFC State [0x00000009] Enabled
  QCN State [0x00000000] Disabled
  QCN State [0x00000009] Enabled
  Priority Flow Control: Enabled
  -----
    PFC 0: Off
    PFC 1: Off
    PFC 2: Off
    PFC 3: On
    PFC 4: Off
    PFC 5: Off
    PFC 6: Off
    PFC 7: Off
    Traffic Class  Bandwidth  Priority
    -----
    0                50%      0 1 2 4 5 6 7
    1                50%      3
    2                0%
    3                0%
    4                0%
    5                0%
    6                0%
    7                0%
    Application    Priority
    -----
    0x8906        3
MAC ACL Status: Disabled
VLAN ACL Status: Disabled
VF Minimum Bandwidth: 28%
VF Maximum Bandwidth: 100%
Controller Version: 00000B00

```

The the information from the **entstat** command, you can calculate the percentage of the physical port's bandwidth that is designated to the logical port.

- ▶ VF Minimum Bandwidth is the capacity assigned to the logical port. The example shows this value as 28%.
- ▶ Physical Port Speed is the actual speed that the port is running. The example shows this as 10 Gbps.

Therefore, the bandwidth assigned to this port is 2.8 Gbps.

Along with AIX commands, more information about SR-IOV port configuration and statistics are available on the HMC.

4.6.2 IBM i

On the IBM i, the logical ports can be identified by the specific codes. As shown in the Example 4-6, the ports are reported with 2C4C type.

Example 4-6 SR-IOV logical port highlighted on IBM i

```

Work with Communication Resources
System: C102EC8P

Type options, press Enter.
  5=Work with configuration descriptions  7=Display resource detail

Opt Resource      Type Status      Text
   CMB02          6B03 Operational Comm Processor
     LIN02          6B03 Operational Comm Adapter
       CMN01          6B03 Operational Comm Port
   CMB03          6B03 Operational Comm Processor
     LIN01          6B03 Operational Comm Adapter
       CMN02          6B03 Operational Comm Port
   CMB04          268C Operational Combined function IOP
     LIN03          6B26 Operational Comm Adapter
   CMB05          2C4C Operational Comm Processor
     LIN04          2C4C Operational Comm Adapter
       CMN03          2C4C Operational Ethernet Port

Bottom

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

```

The resource type (CCIN) depends from the type of the physical SR-IOV adapter being used. The Table 4-1 lists CCIN codes for specific SR-IOV adapters.

Table 4-1 Resource type codes for SR-IOV adapters

Feature code	Resource type (CCIN)	Description
EN0H	2B93	PCIe2 LP 4-port (10 Gb FCoE & 1 GbE) SRIOV SR&RJ45
EN0K	2CC1	PCIe2 4-port (10 Gb FCoE & 1 GbE) SFP+Copper&RJ45
EN10	2C4C	Integrated Multifunction Card w/ 10 GbE RJ45 & Copper Twinax
EN11	2C4D	Integrated Multifunction Card w/ 10 GbE RJ45 & SR Optical

4.6.3 Linux

The `lspci` command lists the adapters that are installed on a Linux system (Example 4-7).

Example 4-7 The `lspci` command

```
linux:/mnt # lspci
0000:01:00.0 Ethernet controller: Emulex Corporation Device e228 (rev 10)
0001:01:00.0 Ethernet controller: Emulex Corporation Device e228 (rev 10)
```

The `ifconfig` command can also be used to list the network interfaces on a system. Example 4-8 shows the configured interface with the TCP/IP configuration.

Example 4-8 The `ifconfig` command

```
linux:~ # ifconfig eth1
eth1      Link encap:Ethernet  HWaddr 2E:84:0F:7D:4A:01
          inet addr:10.1.1.10  Bcast:10.1.1.255  Mask:255.255.255.0
          inet6 addr: fe80::2c84:fff:fe7d:4a01/64  Scope:Link
          UP BROADCAST RUNNING MULTICAST  MTU:1500  Metric:1
          RX packets:0 errors:0 dropped:0 overruns:0 frame:0
          TX packets:5 errors:0 dropped:0 overruns:0 carrier:0
          collisions:0 txqueuelen:1000
          RX bytes:0 (0.0 b)  TX bytes:398 (398.0 b)
```

The `ethtool` command shows information about an Ethernet device on Linux. It calculates the adapter speed, based on the physical port speed and capacity that is configured at the HMC, as Example 4-9 shows.

Example 4-9 The `ethtool` command

```
linux:/mnt # ethtool eth1
Settings for eth1:
    Supported ports: [ ]
    Supported link modes:
    Supports auto-negotiation: No
    Advertised link modes:  Not reported
    Advertised pause frame use: No
    Advertised auto-negotiation: No
    Speed: 3710Mb/s
    Duplex: Full
    Port: Other
    PHYAD: 3
    Transceiver: Unknown!
    Auto-negotiation: off
    Supports Wake-on: d
    Wake-on: d
    Current message level: 0xffffffffa1 (-95)
                          drv ifup tx_err tx_queued intr tx_done rx_status
pktdata hw wol 0xffff8000
    Link detected: yes
```

The SR-IOV ports in Linux are supported by the `be2et` device driver when the adapter is in either shared or dedicated mode.

4.6.4 Virtual I/O Server

To activate a VIOS with SR-IOV adapter logical ports, you must assign the adapter logical ports to the VIOS partition by using a dynamic logical partition (DLPAR) operation at run time. Alternatively, you can update the partition profile by adding SR-IOV ports and then activating the VIOS partition.

The following steps add SR-IOV adapter ports to a partition profile. SR-IOV adapter ports can also be added while you create a new profile.

1. Open the partition profile for modification and navigate to the **SR-IOV logical ports** tab in the Logical Partition Profile properties panel (Figure 4-23).

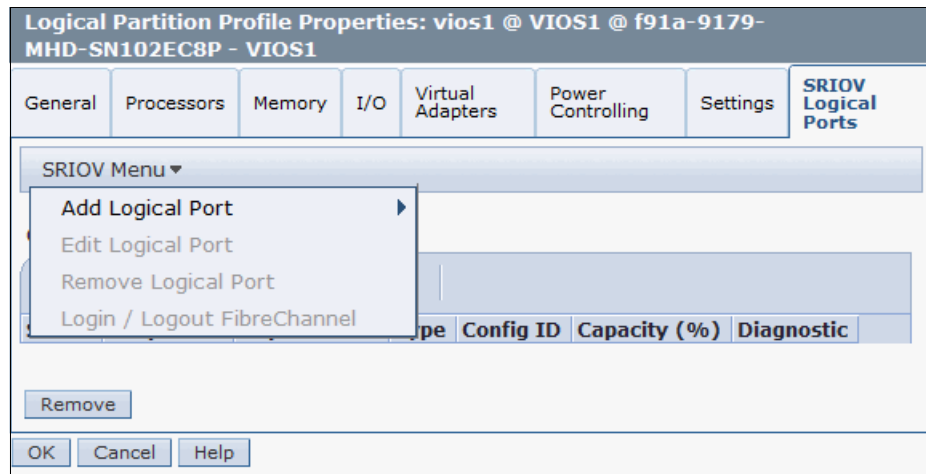


Figure 4-23 Virtual I/O partition profile

2. Click the **SR-IOV menu** and select **Add Logical Port** (Figure 4-24).

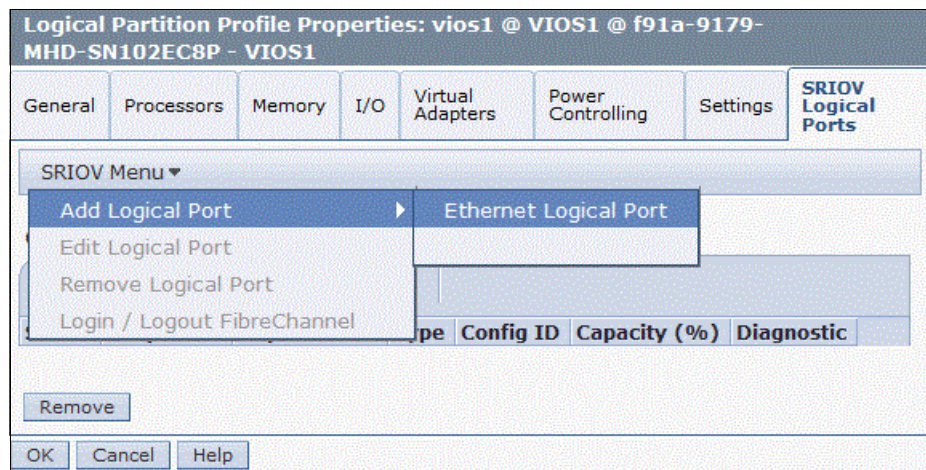


Figure 4-24 Add logical port

3. Select the type of adapter to assign to the partition. For this scenario, we select **Ethernet Logical Port**.

Note: Only the supported logical port types on your system are listed for selection of the adapter type.

4. A new panel for the physical port that supports the logical port protocol to be created opens (Figure 4-25). Select a physical port from the Add Ethernet Logical Port panel.

Add Ethernet Logical Port - VIOS1						
Select the SRIOV Port you would like to create the corresponding Logical Port from.						
Select	Adapter Id	Physical Port	Speed	Active LPs	Available LPs	Link Status
<input checked="" type="radio"/>	1	U2C4B.001.DBJD102-P2-C8-T1	10000	2	8	Up
<input type="radio"/>	1	U2C4B.001.DBJD102-P2-C8-T2	10000	0	10	Up
<input type="radio"/>	1	U2C4B.001.DBJD102-P2-C8-T3	10000	0	10	Up
<input type="radio"/>	1	U2C4B.001.DBJD102-P2-C8-T4	10000	0	10	Up

OK Cancel Help

Figure 4-25 Physical port of SR-IOV adapter

With the physical port selected, the HMC opens a panel (Figure 4-26) where you can view and modify the properties of the logical port that is assigned to the partition profile.

LogicalPortPropertiesBean - VIOS1

Logical Port Information

Adapter ID: 1
 Physical Port: 0
 Type: Ethernet

General | **Advanced**

Partition Information

Partition ID: 3
 Partition Name: VIOS1

Resources

Capacity (%):

Permissions

Diagnostic
 Promiscuous (exclusive with some advanced options)

OK Cancel Help

Figure 4-26 Logical port properties

Some of the properties are as follows,

Capacity

The percentage of the physical port’s resources that should be allocated to this logical port. The value must be a multiple of the default value, which is 2.0%, or the HMC will return an error.

One of the physical port resources that the Capacity setting determines is the logical port’s *desired* minimum bandwidth as a percentage of the physical ports Negotiated bandwidth. The value selected for the Capacity setting is used as the percentage of the physical port’s bandwidth to assign to the logical port as its desired minimum bandwidth.

A port does not have its bandwidth capped by the Capacity setting if additional bandwidth is not currently being used. Capacity settings do not apply to received traffic, only to transmitted traffic. Any unused

bandwidth, whether unassigned or unused on a physical port will be shared equally among all logical ports.

Diagnostic

With the adapter set, select this check box to run the adapter in diagnostic mode. Only one logical port per physical port is allowed to have diagnostic mode set at any one time.

Promiscuous

With the adapter set, select this check box so that the adapter allows the partition to enable unicast promiscuous mode. Promiscuous mode should be selected if the logical port will be a physical device for SEA (that is, the user wants to use SEA to further virtualize the logical port).

5. The next tab in the logical port properties panel shows advanced properties. The user can set required properties on the logical port, as shown in Figure 4-27.

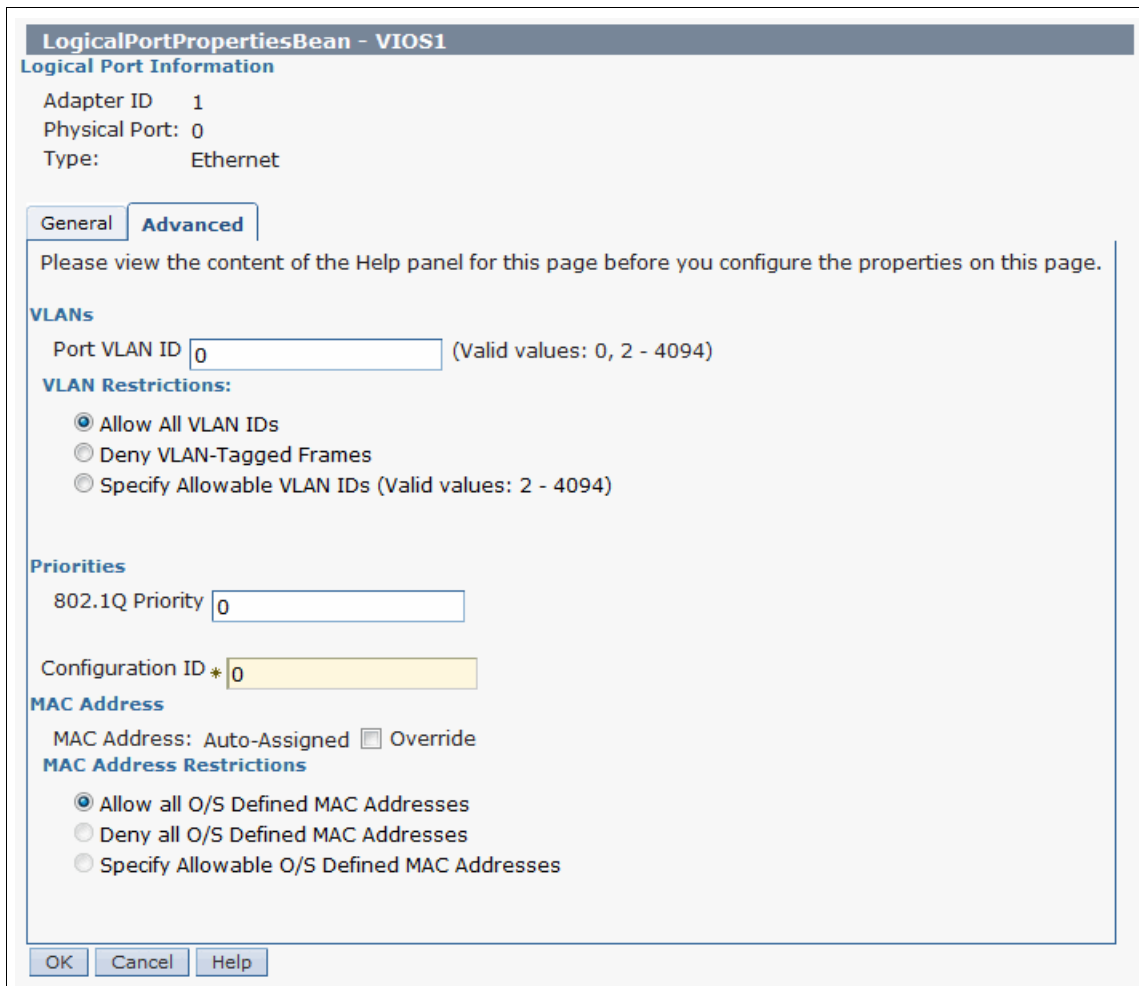


Figure 4-27 Advanced properties

- With the port added to the profile, as shown in Figure 4-28, reconfigure the partition using the `cfgmgr` command if the DLPAR operation is used, or activate the partition with the new profile, as shown in Figure 4-29.

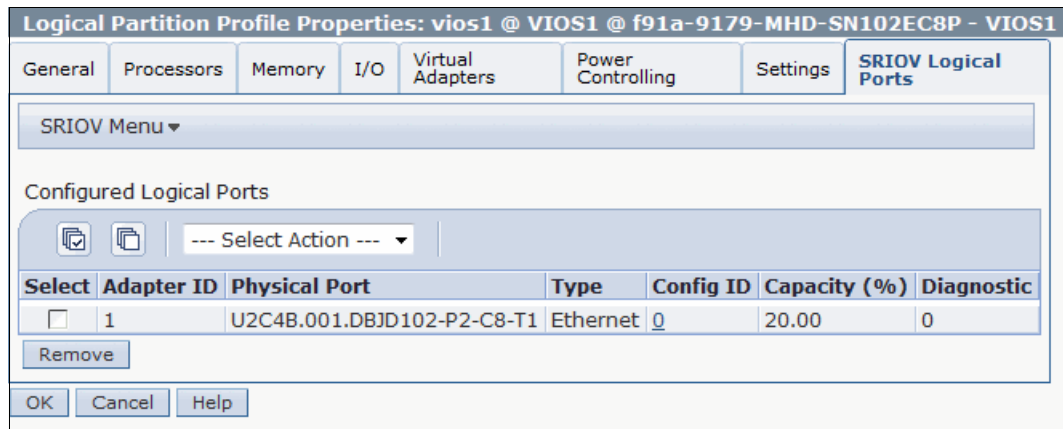


Figure 4-28 Logical port added to profile.

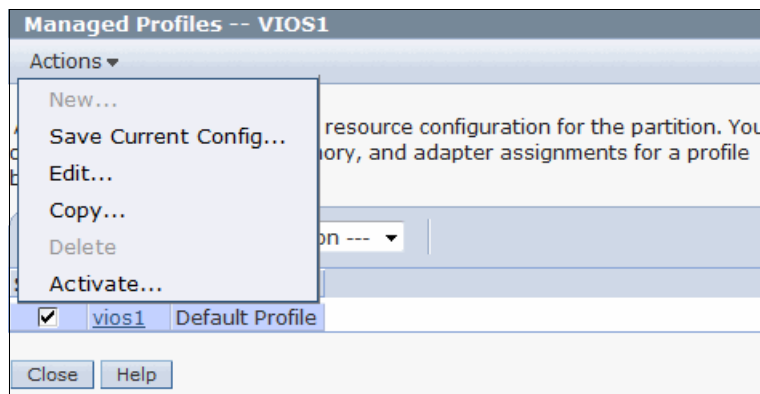


Figure 4-29 Partition activation

- With the VIOS activated, the SR-IOV adapters are now available to the operating system. You can verify the location codes from a command line, as shown in Example 4-10.

Example 4-10 Verifying location codes

```
(0) padmin @ f91e: /home/padmin
$ lsdev |grep VF
ent2                Available    10GBaseT 4-port Integrated Multifunction CNA VF
(df1028e214103b04)
```

```
(0) padmin @ f91e: /home/padmin
$ lsdev -dev ent2 -vpd
ent2                U2C4B.001.DBJD102-P2-C8-T1-S3 10GBaseT 4-port Integrated
Multifunction CNA VF (df1028e214103b04)
```

```
Ethernet Adapter:
Network Address.....2E84006E1900
ROM Level.(alterable).....0.0.9999.19068
Hardware Location Code.....U2C4B.001.DBJD102-P2-C8-T1-S3
```

Name: ethernet
Node: ethernet@0
Device Type: network
Physical Location: U2C4B.001.DBJD102-P2-C8-T1-S3

4.7 Adapter configuration backup and restore

When a change is made to an SR-IOV adapter, for example a mode change or physical port properties change, the new configuration is automatically backed up to the default backup file on the HMC's hard drive. You can also manually back up the current configuration to a specific backup file by navigating to **Configuration** → **Manage Partition Data** → **Backup** under the **Tasks** menu at the server level, and then specifying a file name when prompted.

To restore configuration from a backup file, navigate to **Configuration** → **Manage Partition Data** → **Restore** and then select the appropriate file to restore.

Note: Restoring the partition data restores *all* partition data, not only the adapter configuration.

The HMC attempts to restore the adapter configuration, but be aware of certain restrictions because the HMC cannot switch the adapter mode and restore the physical port properties in one operation.

- ▶ If the adapter is in dedicated mode when the restore takes place, and the backup file has the adapter in shared mode, the HMC restores only the *adapter* configuration (that is, switch the adapter to shared mode).

To avoid this, you can either switch the adapter to shared mode before restoring partition data or restore the partition data twice (first to switch the mode, then to restore physical port properties).

- ▶ If the adapter is in shared mode when the restore takes place, and the backup file has the adapter in dedicated mode, the HMC makes no changes to the adapter.
- ▶ If the adapter is in dedicated mode and the backup file has it in dedicated mode, the HMC makes no changes to the adapter.
- ▶ If the adapter is in shared mode, and the backup file has it in shared mode, the HMC restores the physical port settings.
- ▶ The HMC ignores all failures that are related to restoring the adapter or physical port properties.

4.8 Command-line interface (CLI) support

The HMC CLI supports the same functions as the GUI, using the `chhwres`, `lshwres`, `chsyscfg`, `mksyscfg`, and `lssyscfg` commands. See the man pages for full syntax and examples of the SR-IOV commands.

Two extra SR-IOV functions are available from the CLI only:

- Force delete (`hwdbg` command): This function requires the `hmcpe role` and forces the unconfiguring of logical ports on the adapter and switching the adapter to dedicated mode. This command (Example 4-11) requires the owner partitions to be shut down. Otherwise, it returns a list of active owner partitions.

Attention: This command removes all physical port properties for the adapter. If you want to later switch the adapter back to shared mode, you must manually reconfigure the physical and logical ports or restore partition data from a backup file.

Example 4-11 Using `hwdbg` to force delete SR-IOV configuration of an adapter

```
$ lshwres -r sriov -m Server1 --rsubtype adapter
adapter_id=1,slot_id=2101020b,adapter_max_logical_ports=40,config_state=sriov,functional_state=1,logical_ports=40,phys_loc=U2C4B.001.DBJD102-P2-C8-T1,phys_ports=4,sriov_status=running,alternate_config=0
```

```
$ hwdbg -m Server1 -r sriov -o r -a "slot_id=2101020b"
```

- Reallocate adapter: If an adapter is replaced with a new adapter having the same capabilities, and the new adapter is plugged into the same slot as the original, the hypervisor will automatically associate the old adapter's configuration with the new adapter. However, if the new adapter is plugged in to a different slot, the `chhwres` command (Example 4-12) is needed to associate the original adapter configuration with the new adapter.

Note: The system must be in standby mode to ensure partitions are powered off.

Example 4-12 Using `chhwres` to reallocate an adapter to a new slot

```
$ chhwres -m Server1 -r sriov --rsubtype adapter -o m -a \
"slot_id=2101020b,target_slot_id=21010208"
```

4.9 Miscellaneous notes

Consider the following information:

- ▶ When a system is in Manufacturing Default Configuration (MDC) mode, all adapters are in dedicated mode. Switching an adapter to shared mode will also switch the system out of MDC mode.
- ▶ Partitions with SR-IOV logical ports cannot be migrated, suspended, or remotely restarted. You must use DLPAR to remove the logical ports before you perform such tasks on the partition.
- ▶ Shared mode adapters and configured SR-IOV logical ports are not included in I/O Registry (IOR) data collection and sysplans.
- ▶ Activating full system resource profiles fails if any adapter is in shared mode.
- ▶ System profile validation or activation fails if SR-IOV logical port conflicts occur across partition profiles.
- ▶ The HMC currently limits the number of logical ports to 1024 per system because of save area space constraints.



Maintenance

This chapter describes how to maintain and troubleshoot your system. The SR-IOV adapter generates SRC codes and errorlog entries that can help you determine the cause of an error. Maintenance can be performed concurrently or non-concurrently. Specific rules must be followed to avoid any adapter problems after the replacement. Some command examples that can be used to maintain the SR-IOV adapter are also shown in this section.

5.1 Adapter firmware

When the SR-IOV adapter is used as a dedicated adapter, the adapter microcode is updated just like any other PCI adapter. Microcode updates are available from the IBM Fix Central web page:

<http://www.ibm.com/support/fixcentral/>

As soon as the available adapters are transitioned to SR-IOV mode, the *system firmware* updates the *adapter firmware* to the current built-in adapter firmware level. That is independent from the level that is currently installed on the adapter. The system firmware will always install that current available level.

Note: System firmware updates the adapter firmware when the adapter is used in SR-IOV mode, regardless of what current level is installed on the adapter.

The update takes approximately 5 minutes and stay in an *initializing* state during that time. When new system firmware service packs are installed, they might contain new SR-IOV adapter firmware. If the logical ports are in use, the adapter firmware update remains deferred for the SR-IOV adapters. To ensure all SR-IOV enabled adapters are updated, the be sure to reboot the server.

The current adapter firmware level can be obtained from the command line only for specific operating systems. The `lscfg -v1` command (Example 5-1) shows the adapter firmware on AIX.

Example 5-1 Adapter firmware query in AIX

```
# lscfg -v1 ent3
ent3          U2C4B.001.DBJD102-P2-C8-T1-S2  10GBaseT 4-port Integrated
Multifunction CNA VF (df1028e214103b04)

Ethernet Adapter:
Network Address.....2E840A550200
ROM Level.(alterable).....0.0.9999.19068
Hardware Location Code.....U2C4B.001.DBJD102-P2-C8-T1-S2
```

On Linux, the adapter firmware can be listed by the `ethtool -i` command (Example 5-2)

Example 5-2 Adapter firmware in Linux

```
[root@localhost net]# ethtool -i eth1
driver: be2net
version: 4.6.62.0r
firmware-version: 0.0.9999.19068
bus-info: 0000:01:00.0
supports-statistics: yes
supports-test: yes
supports-eeprom-access: no
supports-register-dump: no
supports-priv-flags: no
```

Note: Neither the HMC nor IBM i provide a method to check the adapter firmware.

5.2 Problem determination and data collection

On Power Systems, every effort is made to isolate problems to a single partition. For example, if a partition device driver provides a bad direct memory access (DMA) address to the adapter, thus causing an enhanced error handling (EEH) event, it affects only the logical port (VF) that is associated with the device driver, and the device driver should recover from the event. Several of these types of errors are isolated to the individual logical port.

Also, some errors affect the entire adapter. For example, if the adapter firmware detects an internal error, recovery of the whole adapter might be necessary. In this case, the hypervisor detects the error, in addition to all of the adapter logical ports. If the hypervisor can successfully recover the adapter, the logical ports go through their EEH recovery steps and recover. If the adapter cannot be recovered (for example, if the adapter is broken), then all logical ports cannot recover either, impacting the partitions that are sharing all of the adapter logical ports.

To debug possible problems, multiple files and data can be collected from the system, such as these items:

- ▶ PE debug information from the HMC.

Log in as the hscpe user and issue the **pedbg** command. Appropriate flags are provided by the support function.

- ▶ Nondisruptive platform resource dump to gather SR-IOV debug data.
- ▶ Collect any LPADump that occurred around the time of the error or create a new LPADump.

An LPADump can be initiated by the hypervisor, the operating system, the adapter itself or a user, where the System Reference Code indicates the source:

- A2D03004: User-initiated
- A2D03010: Hidden LPAR, OS or adapter initiated
- B2D3004, B2ppF00F, B2ppF011, B2ppF012, B400F104: Hypervisor initiated

5.2.1 SR-IOV Platform Dump

If an SR-IOV Platform Dump is needed, several ways are available:

- ▶ HMC command line
- ▶ ASMI menu
- ▶ HMC GUI

SR-IOV command line from HMC

Log in to the HMC as hscroot first and then use the following syntax to initiate the dump:

```
startdump -m MANAGEDSYSTEM -t resource -r 'restart sriov Uxxx.001.xxxxxxx-Px-Cx'
```

To get the MANAGEDSYSTEM name, use the **lssyscfg -r sys -F name** command to have the Managed System names displayed. The location code **Uxxx.001.xxxxxxx-Px-Cx** can be obtained from the HMC, AIX or IBM i. For IBM i use the System Service Tools (SST), as shown in Example 5-5 on page 53.

Figure 5-1 shows the task that allows an administrator to locate the relations between physical and logical ports in the HMC. Figure 5-2 shows what logical ports on which LPARs are assigned to the specific physical ports.

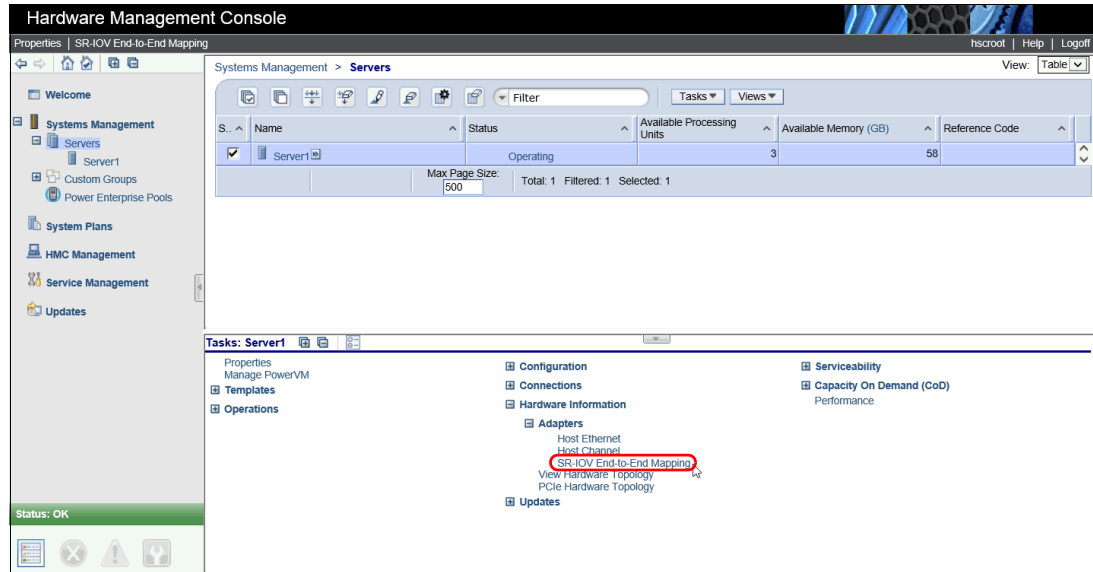


Figure 5-1 This option in the HMC allows to find relation between the logical and physical ports

SR-IOV Device Mappings - Server1							
Select	Physical Port	Type	Port ID	Configured LPs	Available LPs	Speed	Link Status
<input checked="" type="radio"/>	U2C4B.001.DBJD102-P2-C8-T1	Ethernet	0	6	4	100Mbps	Up
<input type="radio"/>	U2C4B.001.DBJD102-P2-C8-T2	Ethernet	1	2	8	100Mbps	Up
<input type="radio"/>	U2C4B.001.DBJD102-P2-C8-T3	Converged Ethernet	2	1	9	10Gbps	Up
<input type="radio"/>	U2C4B.001.DBJD102-P2-C8-T4	Converged Ethernet	3	1	9	10Gbps	Up
Logical Partition	Location	Device Name					
VIOS2	U2C4B.001.DBJD102-P2-C8-T1-S2	ent3					
VIOS2	U2C4B.001.DBJD102-P2-C8-T1-S3	ent4					
rh65	U2C4B.001.DBJD102-P2-C8-T1-S4	Unknown					
sles11	U2C4B.001.DBJD102-P2-C8-T1-S7	Unknown					
aix71_t13	U2C4B.001.DBJD102-P2-C8-T1-S8	ent3					
IBMi72	U2C4B.001.DBJD102-P2-C8-T1-S13	CMN06					

Figure 5-2 Location of the logical port from the HMC

Note: The Device Name will not be displayed if the partition is not running or it does not have a Resource Monitoring Control (RMC) in the case of AIX and Linux.

For an AIX LPAR, use the `lscfg -v1 entX` command, as shown on Example 5-3.

Example 5-3 lscfg command output

```
# lscfg -v1 ent3
ent3          U2C4B.001.DBJD102-P2-C8-T1-S2  10GBaseT 4-port Integrated
Multifunction CNA VF (df1028e214103b04)
```

```
Ethernet Adapter:
Network Address.....2E840A550200
ROM Level.(alterable).....0.0.9999.19068
Hardware Location Code.....U2C4B.001.DBJD102-P2-C8-T1-S2
```


For the IBM i, start SST, and go to **Start a service tool** → **Hardware Service Manager** → **Locate Resource By Resource Name**, and type a communication resource name, for instance: CMN06 > Display detail. It shows the physical location code, as in Example 5-4.

Example 5-4 System Service Tools location of the logical SR-IOV port

Communication Hardware Resource Detail

```

Description . . . . . : Virtual Comm Port
Type-model . . . . . : 2C4C-006
Status . . . . . : Operational
Serial number . . . . . : YL10JH326041
Part number . . . . . : 74Y3832
Resource name . . . . . : CMN06
Physical location . . . . . : U2C4B.001.DBJD102-P2-C8-T1-S13
PCI bus . . . . . :
  System bus . . . . . : 3591
  System board . . . . . : 0
  System card . . . . . : 0
Communications . . . . . :
  I/O bus . . . . . : 14
  Adapter . . . . . : 1
  Port . . . . . : 0
  Channel . . . . . :
Bridging capable . . . . . : No
  
```

Bottom

```

F3=Exit      F5=Refresh      F6=Print
F9=Change detail      F12=Cancel
  
```

With the **restart** command-line option, the dump will be disruptive. See Example 5-5 for the complete command syntax.

Example 5-5 HMC command line

```

startdump -m f91a-9179-MHD-SN102EC8P -t resource -r 'restart sriov
U2C4B.001.DBJD102-P2-C8'
  
```

When the **restart** option is used to perform a disruptive SR-IOV dump, the adapter is restarted. An adapter dump will be included also.

All logical ports on the adapter will enter EEH recovery while the dump and reboot are occurring. The logical ports will recover after the dump is complete.

SR-IOV dump with ASMI

Another way to initiate the SR-IOV Platform dump is to access the Advanced System Management Interface (ASMI) of the system and use the Resource Dump function (Figure 5-3).

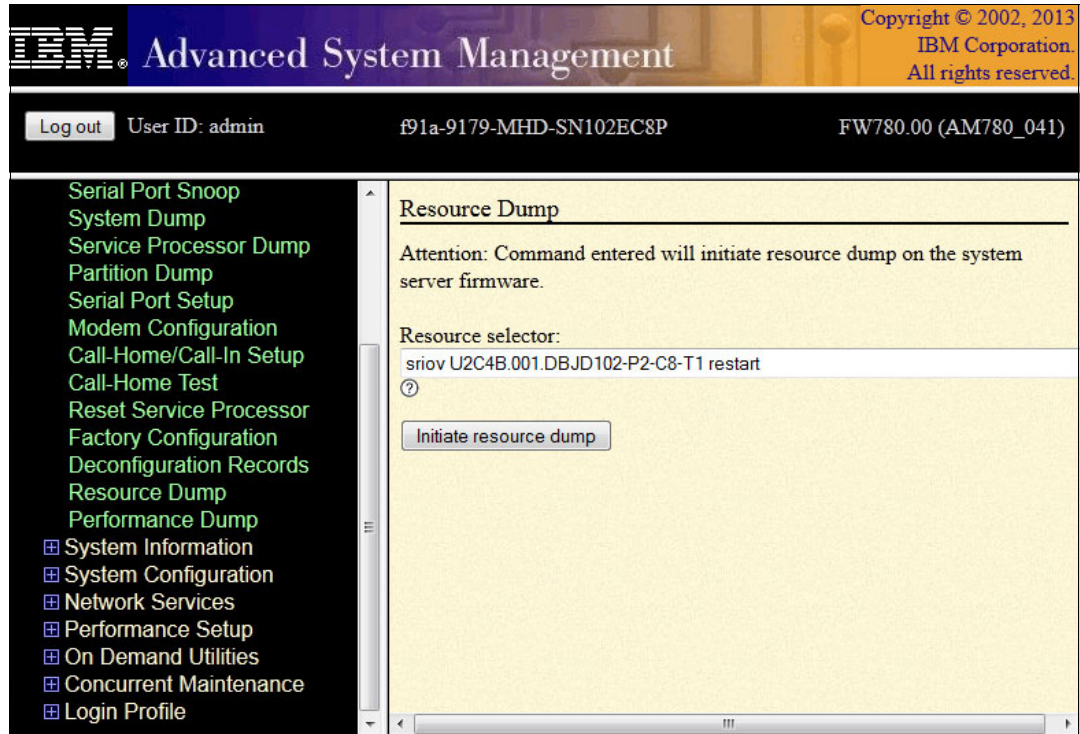


Figure 5-3 Initiate SR-IOV dump from ASMI

SR-IOV dump from HMC GUI

The easiest way to initialize the SR-IOV dump is to use the HMC GUI. Select **Service Management** → **Manage Dumps** and then select **Actions** → **Initiate Resource Dump**.

Figure 5-4 shows how to access the Resource Dump menu in the HMC GUI.

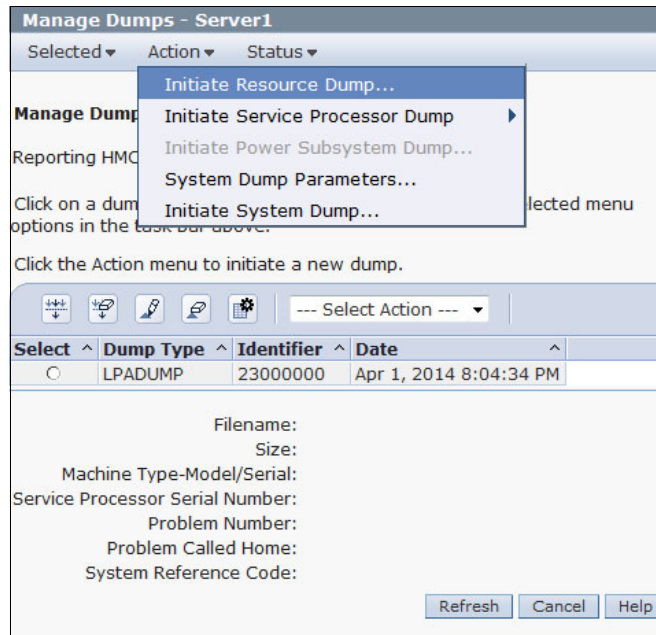


Figure 5-4 Initiating Resource Dump from HMC GUI for SR-IOV Adapter

When you select **Initiate Resource Dump**, you must specify the resource selector by using the following syntax:

```
sriov <adapter_location_code> [restart]
```

The restart is optional again. See Figure 5-5 for the HMC GUI window.

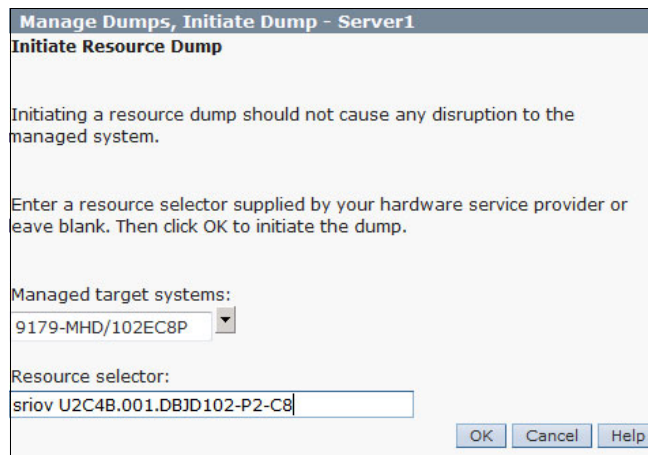


Figure 5-5 initiating the SR-IOV Adapter dump

When done, the dump is off-loaded to the attached HMC and can be managed from the HMC. This applies to all ways of gathering the dump. The file can then be downloaded to media (when available), copied to a remote system, called home, or simply deleted. See Figure 5-6

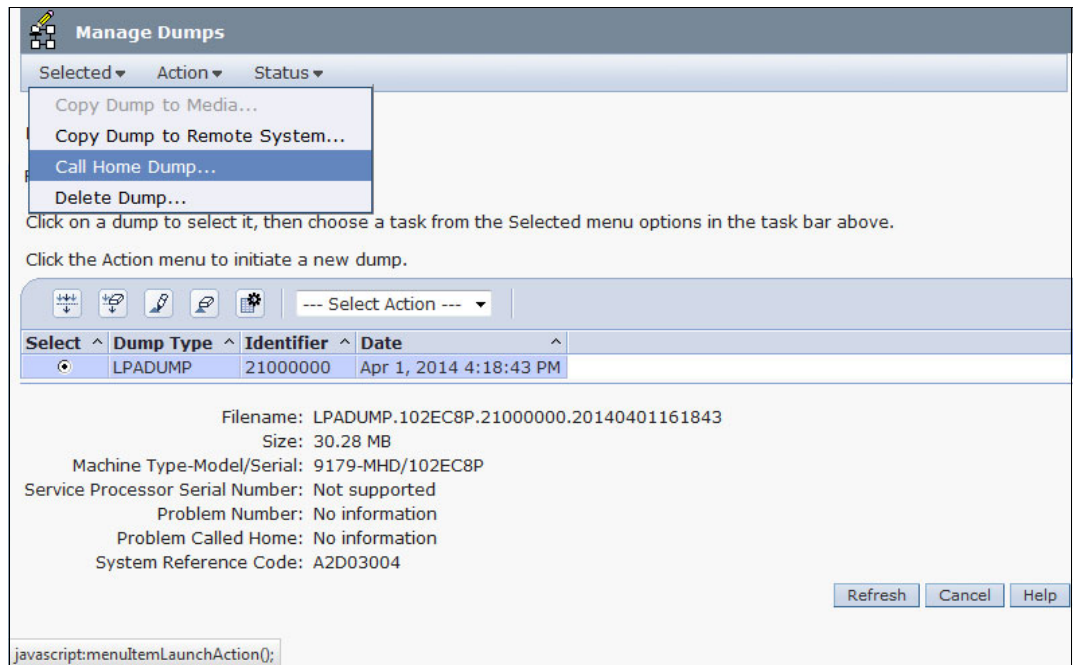


Figure 5-6 Manage Dumps HMC GUI

When the dump data has been transferred to the appropriate support team, the analysis helps to find a cause for the observed problems with the SR-IOV Adapter.

5.3 Problem recovery

Platform hardware or PCIe bus errors may result in logical ports entering a failed state. If a logical port is failed, the first item to investigate is the operating system (OS) error logs. If the logical port error is caused by a larger error that effects the entire adapter, then the OS error log will contain a Platform Log ID (PLID) that can be used to correlate the logical port error to a hypervisor error log.

When errors are encountered for the SR-IOV adapter, there are the common areas to look for details. Error logging will be done in the OS Error log, HMC Serviceable Events, and Service Processor Errorlogs, which can be accessed from the ASM Interface. Adapter errors have the system reference code (SRCs) in the B4xxxxxx range. These error logs contain a location code that can be used to determine which adapter the errors are related to. If the errors are serviceable, they may either call out the adapter or the system firmware. If other parts of the hypervisor encounter an error with the adapter or with configuration, they may also log errors.

Error codes for SR-IOV adapters

Different SRCs are available for SR-IOV issues. Some possible errors and their possible causes are shown in this section.

- ▶ B400xxxx range

For these SR-IOV errors, follow the FRU callouts in the details. Depending on the analysis, the next level of support should be involved.

- ▶ B400FF05

B400FF05 will be logged anytime an EEH event that affects the adapter is hit. The error is also logged when a manual adapter dump is forced if the restart option is used such that the adapter dump is collected and the adapter must be rebooted.

- ▶ B2006002

Firmware detected a change of the physical SR-IOV Adapter. The configuration for the original adapter is no longer valid. The adapter and all the virtual resources are not available. The SR-IOV adapter might have been replaced. It must be verified if that was done. If the old adapter was not defective and it was replaced for test purposes only, try to install the old adapter again. If the adapter is still the same, collect an SR-IOV adapter dump and contact your next level of support.

- ▶ B200600E

One or more SR-IOV adapters used by the partition did not become functional in a reasonable amount of time. The IPL of the partition continued but one or more logical ports may not be available.

- ▶ B200600F

One or more logical ports used by the partition were not configured in a reasonable amount of time. The partition IPL continued but one or more of the listed logical ports may not be available.

- ▶ B200F011

A timeout occurred waiting for a response to a message sent to the hidden LPAR that owns the adapter. An LPADUMP has been created.

- ▶ B2009004, B2009008, B200900C, B2009010, B2009014, B200910C, B2009110

An SR-IOV configuration error has occurred. Verify whether the configuration for the SR-IOV adapter is valid and that the adapter is functional. If yes, contact your next level of support.

The complete list of SRCs is available in the IBM Knowledge Center. HSCLxxxx HMC codes are also present that give detailed information of possible problems with the SR-IOV Adapter.

For more information, see the following web page:

<http://pic.dhe.ibm.com/infocenter/powersys/v3r1m5/index.jsp>

Verifying the adapter status on the HMC

You can see the SR-IOV adapter status in the Physical I/O Properties window. When a system is functional, the SR-IOV properties are displayed. If the system is not functional, then you might see the Initializing, Failed, or Missing status, depending on the adapter status that is returned by the hypervisor. For example, if the SR-IOV card stopped working and is not powering on, then the status displays as Missing. To access the SR-IOV properties window select **Systems Management** → **Servers**. Then select your managed system, and select **Properties** → **I/O**. Select the SR-IOV Adapter, then select the **SR-IOV** tab. Figure 5-7 shows the SR-IOV adapter properties window.

SR-IOV Device Mappings - Server1							
Select	Physical Port	Type	Port ID	Configured LPs	Available LPs	Speed	Link Status
<input checked="" type="radio"/>	U2C4B.001.DBJD102-P2-C8-T1	Ethernet	0	6	4	100Mbps	Up
<input type="radio"/>	U2C4B.001.DBJD102-P2-C8-T2	Ethernet	1	2	8	100Mbps	Up
<input type="radio"/>	U2C4B.001.DBJD102-P2-C8-T3	Converged Ethernet	2	1	9	10Gbps	Up
<input type="radio"/>	U2C4B.001.DBJD102-P2-C8-T4	Converged Ethernet	3	1	9	10Gbps	Up
Logical Partition	Location	Device Name					
VIOS2	U2C4B.001.DBJD102-P2-C8-T1-S2	ent3					
VIOS2	U2C4B.001.DBJD102-P2-C8-T1-S3	ent4					
rh65	U2C4B.001.DBJD102-P2-C8-T1-S4	Unknown					
sles11	U2C4B.001.DBJD102-P2-C8-T1-S7	Unknown					
aix71_t13	U2C4B.001.DBJD102-P2-C8-T1-S8	ent3					
IBMi72	U2C4B.001.DBJD102-P2-C8-T1-S13	CMN06					

Close Help

Figure 5-7 SR-IOV adapters status

SR-IOV adapter states

The SR-IOV adapter can have several states, that can be displayed in the HMC GUI. While selecting **Systems Management** → **Servers**. Then, select your managed system, and select **Properties** → **I/O**. Select the SR-IOV Adapter and select the **SR-IOV** tab. From there you can select the available ports and display their configuration.

During a state change, for example, the adapter can show some states like these:

- ▶ Initializing (can take up to 5 minutes)
- ▶ Dumping (a LPADump is generated either automatically or manually)
- ▶ Powering off

Sometimes the adapter might be in one the following states:

- ▶ Failed
- ▶ Powered off

If that applies, follow the normal service procedures to bring the adapter back to a normal operational state.

These states might need user intervention:

- ▶ PCIe ID Mismatch

An incompatible adapter is plugged into the slot. You must replace the current card with an SR-IOV card with the same PCIe ID (or same feature code) as the original one.

- ▶ Missing

The adapter is unplugged from the original slot. You must put the adapter, or a new adapter with same capabilities, back to the original slot.

► Adapter is in initializing state

Sometimes an adapter gets stuck in the initializing state. A possible way to release an adapter out of that status is to force an LPADump with the restart option as described in 5.2.1, “SR-IOV Platform Dump” on page 51. The initializing state can take up to 5 minutes. If the adapters stays in this state longer than 5 minutes, involve the next level of support.

For these cases, if you want to clean up the SR-IOV adapter configuration, you can try to switch the adapter to dedicated mode by clearing the Shared Mode check box and clicking **OK**. If there are configured logical ports, HMC suggests that you release them and try the switching again.

If the switching mode operation does not work, then you might need to force a delete operation, which must be performed on the HMC command line.

The *force delete* procedure requires the HMC command line and the hscpe user with the hmcpe role. This procedure forces unconfiguring logical ports on the adapter and switching the adapter to dedicated mode. This procedure requires the owner partitions to be shut down, otherwise it gives a list of active owner partitions, and does not delete the adapter.

```
hwdbg -m MANAGEDSYSTEM -r sriov -o r -a "slot_id=21010208"
```

After replacing an adapter with a new adapter that has the same capabilities, if the new adapter is plugged into the same slot, the hypervisor automatically associates the old adapter's configuration to the new adapter. A *reallocate adapter* procedure is necessary; if the new adapter is plugged into a different slot, use the following command to re-associate the source adapter configuration to the new adapter. The system must be at standby to ensure partitions are powered off.

```
chhwres -m metsfsp1 -r sriov -rsubtype adapter -o m -a  
"slot_id=21010208,target_slot_id=2101020A"
```

See 5.6, “HMC commands for SR-IOV handling” on page 66 for a more detailed HMC command overview.

5.4 Concurrent maintenance

SR-IOV adapters support concurrent maintenance. The adapters can be added, removed, and replaced without disrupting the system or shutting down the partitions. The HMC provides a GUI for adapter concurrent maintenance operations. Exchanging or replacing one adapter with an adapter of a different type (whether changing from one protocol to another, or changing to a higher performance or higher capacity adapter of the same protocol) is accomplished by using the remove operation, followed by an add operation (two separate procedures). The operation to repair, exchange, and replace (one procedure does all three) is supported only for adapters of the same type. The type of an adapter is generally based on the PCI device ID, vendor ID, subsystem ID, and subsystem vendor ID values.

SR-IOV adapters can also be updated to a new firmware level concurrently. The firmware update does not affect the existing SR-IOV configurations in the server, thus minimizing system down time.

5.4.1 Adapter concurrent and non-concurrent maintenance

Two kinds of SR-IOV capable adapters are available:

- ▶ Integrated adapters: Can only be added or replaced non concurrently.
- ▶ PCIe I/O adapters: Can be added and replaced concurrently and non concurrently.

The HMC is required for those actions.

Concurrent add

To add an SR-IOV capable adapter to the system, use the HMC GUI. To determine the SR-IOV capable I/O slots, check the properties. Select the server, then select **Properties**; the I/O tab then shows the PCIe slots (Figure 5-8).

Slot	Description	Bus	I/O Pool Id	Owner	Type	SR-IOV Capable(Logical Port Limit)
U2C48.001.DBJD102-P2-C9-T1	PCI-E SAS Controller	520	Unassigned	VIOS2		No
U2C48.001.DBJD102-P2-C9-T2	PCI-E SAS Controller	521	Unassigned	VIOS2		No
U2C48.001.DBJD102-P2-C8-T5	Universal Serial Bus UHC Spec	522	Unassigned	Unassigned		No
U2C48.001.DBJD102-P2-C8-T1	Integrated MultiFunction Card w/ 10GbE RJ45 & Copper Twinax	523	Unassigned	Hypervisor		Yes(40)
U2C48.001.DBJD102-P2-C6	Empty slot	524	Unassigned	Unassigned		Yes(96)
U2C48.001.DBJD102-P2-C5	Quad 8 Gigabit Fibre Channel Adapter	525	Unassigned	Unassigned		No
U2C48.001.DBJD102-P2-T3	RAID Controller	512	Unassigned	Unassigned		No
U2C48.001.DBJD102-P2-C8-T7	Generic XT-Compatible Serial Controller	513	Unassigned	Unassigned		No
U2C48.001.DBJD102-P2-C4	Empty slot	514	Unassigned	Unassigned		Yes(96)
U2C48.001.DBJD102-P2-C3	Empty slot	515	Unassigned	Unassigned		Yes(96)
U2C48.001.DBJD102-P2-C2	Empty slot	516	Unassigned	Unassigned		Yes(96)
U2C48.001.DBJD102-P2-C1	Dual 1 Gigabit Ethernet-TX PCI-E Adapter	517	Unassigned	VIOS2		No

Total: 12 Filtered: 12

Figure 5-8 I/O slot properties

Use **Serviceability** → **Hardware** → **MES Tasks** → **Add FRU** to start the concurrent add. See Figure 5-9.

Add/Install/Remove Hardware - Add FRU, Select FRU Type - Server1

Select an installed enclosure type from the drop down list and choose a FRU type, then Click the Next button to locate and add the selected FRU type.

Selected System: 9179-MHD*102EC8P
 Enclosure type: System Unit, Model MHD

FRU types:

Select	Description
<input type="radio"/>	GX Adapter Card
<input type="radio"/>	Memory DIMM
<input type="radio"/>	RAID Enablement Card
<input checked="" type="radio"/>	PCI Adapter Card
<input type="radio"/>	DVD Drive
<input type="radio"/>	Disk Drive (DASD)

< Back Next > Finish Cancel

Figure 5-9 Concurrent adapter add

All necessary steps to install the adapter are covered in the installation instructions, available at the following website:

<http://pic.dhe.ibm.com/infocenter/powersys/v3r1m5/index.jsp>

Concurrent exchange and replace

The SR-IOV device mappings panel can be used to verify the partitions with logical ports configured on a specific physical port.

Select **Hardware information** → **Adapters** → **SR-IOV End to End Mapping** in the HMC GUI and then select the appropriate port. Figure 5-10 shows an example.

Select	Physical Port	Type	Port ID	Configured LPs	Available LPs	Speed	Link Status
<input checked="" type="radio"/>	U2C4B.001.DBJD102-P2-C8-T1	Ethernet	0	3	7	10Gbps	Up
<input type="radio"/>	U2C4B.001.DBJD102-P2-C8-T2	Ethernet	1	1	9	10Gbps	Up
<input type="radio"/>	U2C4B.001.DBJD102-P2-C8-T3	Converged Ethernet	2	1	9	10Gbps	Up
<input type="radio"/>	U2C4B.001.DBJD102-P2-C8-T4	Converged Ethernet	3	1	9	10Gbps	Up

Logical Partition	Location	Device Name
VIOS2	U2C4B.001.DBJD102-P2-C8-T1-S2	ent3
VIOS2	U2C4B.001.DBJD102-P2-C8-T1-S3	ent4
rh65	U2C4B.001.DBJD102-P2-C8-T1-S4	Unknown

Close Help

Figure 5-10 SR-IOV end-to-end mapping

To replace the adapter, all the logical ports must be deconfigured. The Exchange FRU task determines the logical resources that must be deconfigured. Use the appropriate OS commands to deconfigure the logical port before replacing the adapter.

Note: All logical ports (VFs) must be deconfigured to successfully replace the adapter.

The adapter must be replaced with a card of the same type. The configuration is preserved during this replacement. When the replacement procedure is done, the logical ports (VFs) must be configured again with OS-specific commands (AIX uses `cfgmgr`, IBM i uses `VRYCFG`).

Concurrent removal

If a PCIe SR-IOV adapter must be completely removed, the logical ports (VFs) must be removed completely for all active partitions. This can be done by taking the SR-IOV adapter out of the shared mode. Using the HMC GUI, select **Serviceability** → **Hardware** → **MES Tasks** → **Remove FRU** and remove the PCIe adapter. After the adapter is removed, its configuration is discarded.

Non-concurrent removal

When the adapter has to be removed completely from the System in the power off state, the adapter must be taken out of the SR-IOV mode before the removal. If this is not done, the adapter appears as missing status in the HMC.

This situation can be resolved by taking the removed, but still configured, adapter out of the SR-IOV mode in the HMC GUI. The orphaned configuration data is then deleted and there is no more missing SR-IOV adapter.

Note: To prevent having to manually remove the configuration, take the adapter out of SR-IOV mode before removing it from the system.

Move and relocate an SR-IOV adapter

It is possible to move an SR-IOV capable adapter from one SR-IOV capable slot to another SR-IOV capable slot inside the same system. It can be done non-concurrently and the configuration is preserved.

The scenario is as follows:

- ▶ A provisioned SR-IOV adapter resides in slot Uxxxx.001.xxxxxxx-Px-Cx.
- ▶ Power off the system.
- ▶ Move the SR-IOV adapter from the Uxxxx.001.xxxxxxx-Px-Cx slot to the Uxxxx.001.xxxxxxx-Px-Cy slot.
- ▶ Power on the system.
- ▶ The SR-IOV configuration and provisioning are preserved.

Concurrent maintenance: Error reasons

Multiple reasons can cause a concurrent maintenance to fail. These are some reasons:

- ▶ Replacing the SR-IOV adapter with a non-like SR-IOV capable adapter has these results:
 - The Exchange FRU procedure fails.
 - A B2006002 SRC code is logged in the event log.

You can recover from this error by retrying the Exchange FRU procedure with an SR-IOV adapter of the same type.

- ▶ An SR-IOV capable adapter is being added to a non SR-IOV capable slot has this result:
 - It is not possible to put the adapter in SR-IOV shared mode.

You can recover from this error by running a Remove FRU operation for the adapter remove FRU for the adapter from the incorrect slot and then use Add FRU to add the adapter to the correct slot

Live Partition Mobility usage

If moving a partition for maintenance reasons is necessary, Live Partition Mobility (LPM) can help to move the partition. Consider these important notes to make LPM operations successful in conjunction with an SR-IOV adapter:

- ▶ If the mobile partition is configured with Link Aggregation of virtual Network Interface Controller (vNIC) adapters, the mobile partition can be migrated to the destination server only when the destination server supports vNIC adapters as well as there are ports available to use as backing devices for the vNICs and the switch ports are configured to connect to the target server for the Link Aggregation.
- ▶ LPM is not allowed with logical ports assigned directly to partitions because logical ports are treated like physical I/O adapters.
- ▶ LPM is allowed if the logical ports are assigned to VIOS and the lpars have only virtual devices.

Consider this information for AIX, Linux, and IBM i:

- ▶ AIX and Linux
 - Logical ports can be unconfigured and then dynamically removed to allow mobility.
 - If the logical ports are configured for Network Interface Backup on AIX or channel bonding, with a virtual Ethernet device as the backup device, failover to the backup device can be performed. Then, the logical port can be unconfigured and dynamically removed without taking down the network interface.

- ▶ IBM i
 - LPM is not allowed on partitions that have logical ports directly assigned.
 - Logical ports cannot be dynamically added to partitions that are enabled for LPM.

Diagnostics

Performing diagnostics on an SR-IOV adapter can be performed in either dedicated (non SR-IOV) mode or in SR-IOV mode. When the diagnostics are used in the dedicated mode, the diagnostic tools are used just like for any other dedicated adapter. When the adapter is set to SR-IOV mode, a logical port must be configured for diagnostics. To display the current logical port settings, select a managed server in the Systems Management pane and select **Dynamic partitioning** → **SR-IOV Logical ports** (Figure 5-11).

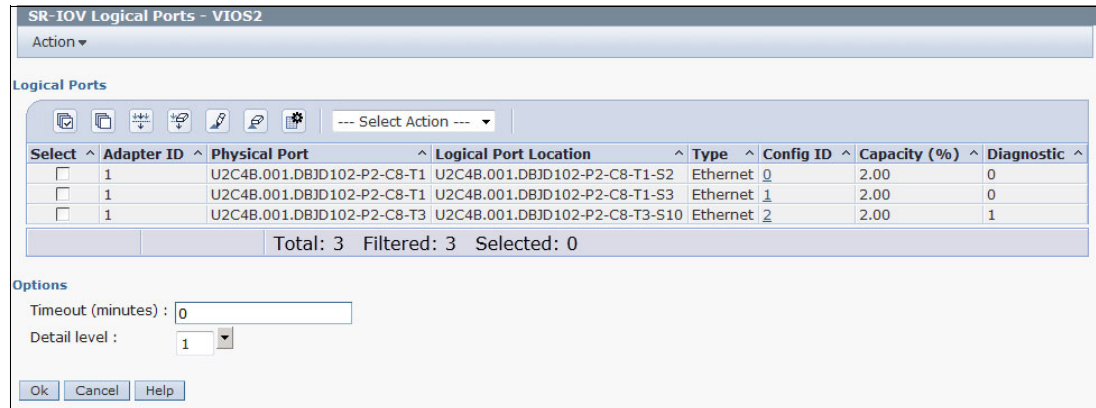


Figure 5-11 SR-IOV logical ports

This overview shows the Diagnostic column. Select the appropriate port, and then click **Action** → **Edit Logical Port** and either select or clear the Diagnostic flag.

Note: Only one logical port per physical port can be set to diagnostic permissions.

Figure 5-12 shows the Logical Port Properties, with the Diagnostic attribute selected.

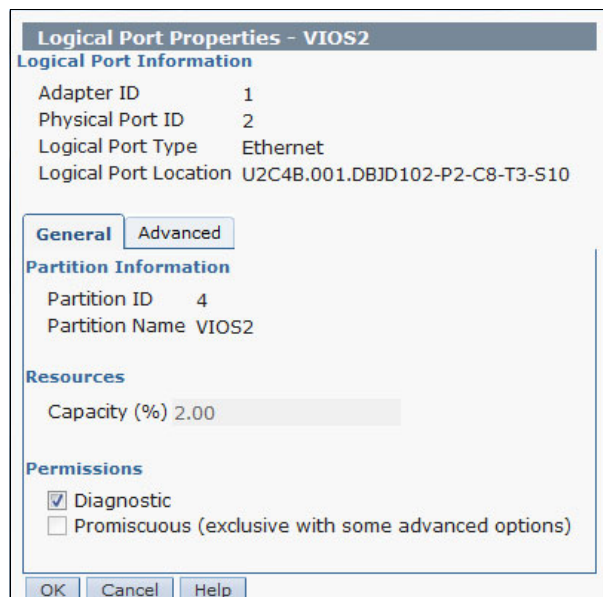


Figure 5-12 Logical Port Properties

The AIX diagnostics can be used in either normal mode or advanced mode. If the partition is running AIX then it might be necessary to put the SR-IOV Adapter back to dedicated mode and then use the Standalone AIX Diagnostics to perform extended tests on the adapter. See Table 5-1 for an overview which diagnostic test is possible under which diagnostic mode.

Table 5-1 Diagnostic modes

Test	AIX diag Normal Mode	AIX diag Advance Mode
Internal Loopback	Dedicated: yes SR-IOV: no	Dedicated: yes SR-IOV: no
External Loopback ^a	Dedicated: no SR-IOV: no	Dedicated: yes SR-IOV: yes

a. Wrap Plug is required for the physical port

External Loopback testing provides an end to end testing of the adapters physical port. This is possible with the use of an wrap plug that needs to be connected to the physical port at the time of testing. Part numbers for the wire plufs are provided during the diagnostic test.

IBM i

There are no facilities within IBM i for performing detailed diagnostics by a user. All advanced diagnostics, and analysis must be coordinated with IBM technical support.

5.5 IBM i performance metrics

Regular Ethernet LAN protocols statistics are recorded by the Collection Services, the license program 5761-PT1. A partition usually collects metrics for its own adapter, and is reported in the QAPMETH database file.

With IBM i V7.1 TR8, performance statistics are introduced regarding SR-IOV adapters. When an IBM i partition becomes eligible to collect internal performance information, it can collect and report the adapter's physical port statistics for all traffic flowing.

To enable the performance collections, select **Allow performance information collection** for the LPAR properties, as shown in Figure 5-13 on page 65.

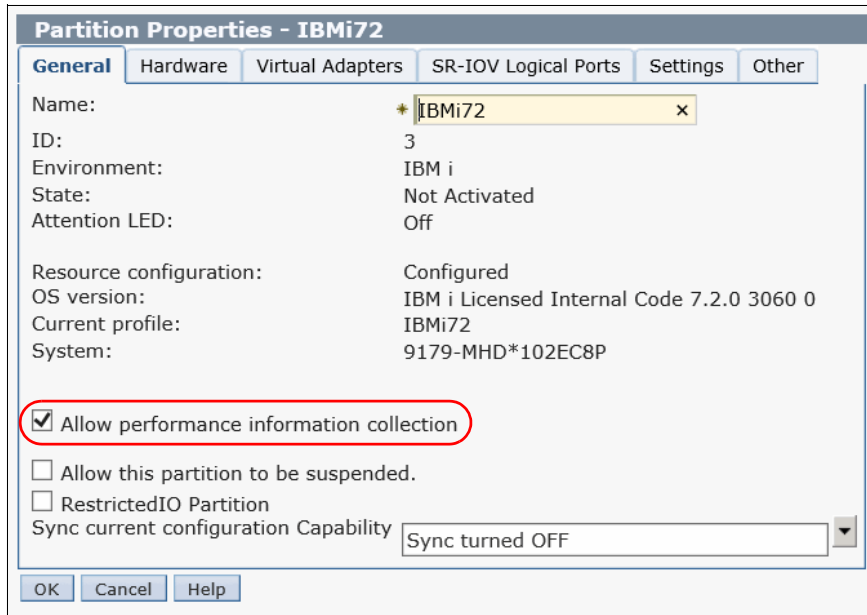


Figure 5-13 The check box that activates performance collection by the LPAR

The physical port metrics for Ethernet ports are stored in the QAPMETHP file. The following metrics are stored in the file:

- ▶ Port resource name
- ▶ Frames transmitted without error
- ▶ Frames received without error
- ▶ CRC error
- ▶ More than 16 retries
- ▶ Out of window collisions
- ▶ Alignment error
- ▶ Carrier loss
- ▶ Discarded inbound frames
- ▶ Receive overruns
- ▶ Memory error
- ▶ Signal quality
- ▶ More than 1 retry to transmit
- ▶ Exactly one retry to transmit
- ▶ Deferred conditions
- ▶ Total MAC bytes received ok
- ▶ Total MAC bytes transmitted ok
- ▶ Transmit frames discarded
- ▶ Unsupported protocol frames

Example 5-6 shows performance data stored in QAPMETHP. The field MAC Bytes Transmitted shows all data transferred by the physical port.

Example 5-6 Physical port performance data from the QAPMETHP file

Display Report

Report width : 396

Position to line Shift to column

Line +...33...+...34...+...35...+...36...+...37...+...38...+...39...+.

	MAC Bytes Transmitted	Transmit Frames Discarded	Unsupported protocol frames
000001	25,368,484	0	0
000002	598	0	0
000003	876	0	0
000004	2,081	0	0
000005	4,942	0	0
000006	660,732,396	0	0
000007	2,087,810,081	0	0
000008	4,278	0	0
***** ***** End of report *****			

Bottom

F3=Exit F12=Cancel F19=Left F20=Right F21=Split

5.6 HMC commands for SR-IOV handling

Most functions that are used to configure the SR-IOV adapters on the HMC in the GUI are also available in the command line. Some special functions, like force delete as shown in “SR-IOV adapter states” on page 58, are available only in the command line.

List SR-IOV adapters, physical ports, logical ports on a managed system

Use the `lshwres` command to complete the following tasks:

- ▶ List the SR-IOV adapter (Example 5-7 shows the command output):

```
lshwres -m MANAGEDSYSTEM -r sriov --subtype adapter
```

Example 5-7 lshwres SR-IOV adapter

```
hscpe@slcb27a:~>lshwres -m Server1 -r sriov --subtype adapter
adapter_id=1,slot_id=2101020b,adapter_max_logical_ports=40,config_state=sriov,func
tional_state=1,logical_ports=40,phys_loc=U2C4B.001.DBJD102-P2-C8-T1,phys_por
ts=4,sriov_status=running,alternate_config=0
```

- ▶ List the converged Ethernet ports (Example 5-8 on page 67 shows the command output):

```
lshwres -m MANAGEDSYSTEM -r sriov --subtype physport --level ethc
```

Example 5-8 List the converged (ethc) Ethernet ports

```
hscpe@slcb27a:~>lshwres -m Server1 -r sriov --rsubtype physport --level ethc
adapter_id=1,phys_port_id=2,phys_port_label=,phys_port_sub_label=,phys_port_loc
=U2C4B.001.DBJD102-P2-C8-T3,phys_port_type=ethc,state=1,config_logical_ports=1,
phys_port_max_logical_ports=10,supported_max_eth_logical_ports=10,max_eth_logic
al_ports=10,curr_eth_logical_ports=1,max_diag_ports=1,max_promisc_ports=1,"capa
bilities=pvid_priority_capable,port_vlan_id_capable,mac_vlan_consistency_capabl
e,clear_phys_port_stat_capable,clear_logical_port_stat_capable",conn_speed=1000
0,config_conn_speed=10000,max_rcv_packet_size=1500,config_max_rcv_packet_size
=1500,rcv_flow_control=0,config_rcv_flow_control=0,trans_flow_control=0,confi
g_trans_flow_control=0,veb_mode=1,vepa_mode=0,priority_flow_control_active=1
adapter_id=1,phys_port_id=3,phys_port_label=,phys_port_sub_label=,phys_port_loc
=U2C4B.001.DBJD102-P2-C8-T4,phys_port_type=ethc,state=1,config_logical_ports=1,
phys_port_max_logical_ports=10,supported_max_eth_logical_ports=10,max_eth_logic
al_ports=10,curr_eth_logical_ports=1,max_diag_ports=1,max_promisc_ports=1,"capa
bilities=pvid_priority_capable,port_vlan_id_capable,mac_vlan_consistency_capabl
e,clear_phys_port_stat_capable,clear_logical_port_stat_capable",conn_speed=1000
0,config_conn_speed=10000,max_rcv_packet_size=1500,config_max_rcv_packet_size
=1500,rcv_flow_control=0,config_rcv_flow_control=0,trans_flow_control=0,confi
g_trans_flow_control=0,veb_mode=1,vepa_mode=0,priority_flow_control_active=1
```

- List the physical Ethernet ports configured on the SR-IOV adapter (Example 5-9 shows the Ethernet port command output):

```
lshwres -m MANAGEDSYSTEM -r sriov -rsubtype physport -level eth
```

Example 5-9 List the Ethernet ports

```
hscpe@slcb27a:~>lshwres -m Server1 -r sriov --rsubtype physport --level eth
adapter_id=1,phys_port_id=0,phys_port_label=,phys_port_sub_label=,phys_port_loc
=U2C4B.001.DBJD102-P2-C8-T1,phys_port_type=eth,state=1,config_logical_ports=6,p
hys_port_max_logical_ports=10,supported_max_eth_logical_ports=10,max_eth_logica
l_ports=10,curr_eth_logical_ports=6,max_diag_ports=1,max_promisc_ports=1,"capab
ilities=pvid_priority_capable,port_vlan_id_capable,mac_vlan_consistency_capable
,clear_phys_port_stat_capable,clear_logical_port_stat_capable",conn_speed=100,c
onfig_conn_speed=auto,max_rcv_packet_size=1500,config_max_rcv_packet_size=150
0,rcv_flow_control=0,config_rcv_flow_control=0,trans_flow_control=0,config_tr
ans_flow_control=0,veb_mode=1,vepa_mode=0
adapter_id=1,phys_port_id=1,phys_port_label=,phys_port_sub_label=,phys_port_loc
=U2C4B.001.DBJD102-P2-C8-T2,phys_port_type=eth,state=1,config_logical_ports=2,p
hys_port_max_logical_ports=10,supported_max_eth_logical_ports=10,max_eth_logica
l_ports=10,curr_eth_logical_ports=2,max_diag_ports=1,max_promisc_ports=1,"capab
ilities=pvid_priority_capable,port_vlan_id_capable,mac_vlan_consistency_capable
,clear_phys_port_stat_capable,clear_logical_port_stat_capable",conn_speed=100,c
onfig_conn_speed=auto,max_rcv_packet_size=1500,config_max_rcv_packet_size=150
0,rcv_flow_control=0,config_rcv_flow_control=0,trans_flow_control=0,config_tr
ans_flow_control=0,veb_mode=1,vepa_mode=0
```

- List the logical Ethernet ports configured on the SR-IOV adapter (Example 5-10 on page 68 shows the command output):

```
lshwres -m MANAGEDSYSTEM -r sriov --rsubtype logport --level eth
```

Example 5-10 Logical Ethernet ports

```
hscpe@slcb27a:~>lshwres -m Server1 -r sriov --rsubtype logport --level eth
config_id=0,lpar_name=VIOS2,lpar_id=4,lpar_state=Running,is_required=1,adapter_id=1,logical_port_id=27004002,logical_port_type=eth,drc_name=PHB
4098,location_code=U2C4B.001.DBJD102-P2-C8-T1-S2,functional_state=1,phys_port_id=0,debug_mode=0,diag_mode=0,huge_dma_window_mode=0,capacity=2.0,promisc_mode=0,mac_addr=2e840a550200,curr_mac_addr=5cf3fccf0a20,allowed_os_mac_addrs=all,allowed_vlan_ids=all,port_vlan_id=0
config_id=1,lpar_name=VIOS2,lpar_id=4,lpar_state=Running,is_required=1,adapter_id=1,logical_port_id=27004003,logical_port_type=eth,drc_name=PHB
4099,location_code=U2C4B.001.DBJD102-P2-C8-T1-S3,functional_state=1,phys_port_id=0,debug_mode=0,diag_mode=0,huge_dma_window_mode=0,capacity=2.0,promisc_mode=0,mac_addr=2e840d60e701,curr_mac_addr=5cf3fccf0a20,allowed_os_mac_addrs=all,allowed_vlan_ids=all,port_vlan_id=0
...
```

- ▶ List all unconfigured logical ports on the SR-IOV adapter (Example 5-11 shows the command output):

```
lshwres -m MANAGEDSYSTEM -r sriov --rsubtype logport
```

Example 5-11 Unconfigured logical ports

```
hscpe@slcb27a:~>lshwres -m Server1 -r sriov --rsubtype logport
adapter_id=1,logical_port_id=2700400e,logical_port_type=unconfigured,drc_name=PHB
4110,location_code=U2C4B.001.DBJD102-P2-C8-T1-S14
adapter_id=1,logical_port_id=2700400f,logical_port_type=unconfigured,drc_name=PHB
4111,location_code=U2C4B.001.DBJD102-P2-C8-T1-S15
adapter_id=1,logical_port_id=27004010,logical_port_type=unconfigured,drc_name=PHB
4112,location_code=U2C4B.001.DBJD102-P2-C8-T1-S16
adapter_id=1,logical_port_id=27004011,logical_port_type=unconfigured,drc_name=PHB
4113,location_code=U2C4B.001.DBJD102-P2-C8-T1-S17
...
```

Changing the adapter modes and attributes

To manually change the adapter attributes and modes, use the **chhwres** command.

- ▶ Switch adapter to shared mode:

```
chhwres -m sys1 -r sriov --rsubtype adapter -o a -a
"slot_id=21010208,adapter_id=1"
```

- ▶ Switch adapter to dedicated mode:

```
chhwres -m sys1 -r sriov --rsubtype adapter -o r -a "slot_id=21010208"
```

- ▶ Set physical port attributes:

```
chhwres -m sys1 -r sriov --rsubtype physport -o s -a
"adapter_id=1,phys_port_id=1,phys_port_label=test,phys_port_sub_label=internet,
conn_speed=10000,max_rcv_packet_size=1500,rcv_flow_control=1,trans_flow_control=1,veb_mode=0,vepa_mode=1,max_eth_logical_ports=10"
```

Note: In this command, conn_speed=10000 refers to 10000 Mbps.

- ▶ Switch an SR-IOV adapter to shared mode:

```
chhwres -r sriov -m managed-system --rsubtype adapter -o a -a "attributes"
```

- ▶ Switch an SR-IOV adapter to dedicated mode:

```
chhwres -r sriov -m managed-system --rsubtype adapter -o r -a "attributes"
```


- ▶ Move the configuration of a failed SR-IOV adapter to a new adapter:
`chhwres -r sriov -m managed-system --rsubtype adapter -o m -a "attributes"`
- ▶ Set SR-IOV physical port attributes:
`chhwres -r sriov -m managed-system --rsubtype physport -o s -a "attributes"`
- ▶ Add or remove an SR-IOV logical port, or to set SR-IOV logical port attributes:
`chhwres -r sriov -m managed-system --rsubtype logport -o {a | r | s} {-p
partition-name | --id partition-ID} -a "attributes" [-w wait-time] [-d
detail-level] [--force]`
- ▶ Reset statistics for an SR-IOV logical or physical port:
`chhwres -r sriov -m managed-system -o rs --rsubtype {logport | physport} -a
"attributes"`

See Table 5-2 for valid **chhwres** attribute names for changing an SR-IOV physical port.

Note: When an attribute for an SR-IOV physical port is changed, a short network interruption might occur for all partitions that share the physical port.

Table 5-2 Valid *chhwres* attribute names

Attribute	Value
adapter_id	Required
phys_port_id	Required
conn_speed	Possible valid values: <ul style="list-style-type: none"> ▶ auto ▶ 10: 10 Mbps ▶ 100: 100 Mbps ▶ 1000: 1 Gbps ▶ 10000: 10 Gbps ▶ 100000: 100 Gbps
max_eth_logical_ports	An integer value less than or equal to the maximum number of Ethernet logical ports allowed on any physical port on the adapter
max_recv_packet_size	<ul style="list-style-type: none"> ▶ 1500 - 1500 bytes ▶ 9000 - 9000 bytes (jumbo frames)
phys_port_label	1 - 16 characters Specify none to clear the physical port label
phys_port_sub_label	1 - 8 characters Specify none to clear the physical port sublabel
recv_flow_control	<ul style="list-style-type: none"> ▶ 0: disable ▶ 1: enable
trans_flow_control	<ul style="list-style-type: none"> ▶ 0: disable ▶ 1: enable
veb_mode	<ul style="list-style-type: none"> ▶ 0: disable Virtual Ethernet Bridge mode ▶ 1: enable Virtual Ethernet Bridge mode
vepa_mode	<ul style="list-style-type: none"> ▶ 0: disable Virtual Ethernet Port Aggregator mode ▶ 1: enable Virtual Ethernet Port Aggregator mode

Command examples for changing values

Examples of the **chhwres** command are shown in these tasks:

- ▶ Switch an SR-IOV adapter to shared mode:

```
chhwres -r sriov -m sys1 --rsubtype adapter -o a -a "slot_id=21010202"
```
- ▶ Switch an SR-IOV adapter to dedicated mode:

```
chhwres -r sriov -m sys1 --rsubtype adapter -o r -a "slot_id=21010202"
```
- ▶ Set the connection speed for SR-IOV physical port 0 to 1000 Gbps:

```
chhwres -r sriov -m sys1 --rsubtype physport -o s -a  
"adapter_id=1,phys_port_id=0,conn_speed=1000000"
```
- ▶ Add an SR-IOV Ethernet logical port (using defaults) to partition lpar1:

```
chhwres -r sriov -m sys1 --rsubtype logport -o a -p lpar1 -a  
"adapter_id=1,phys_port_id=1,logical_port_type=eth"
```
- ▶ Remove an SR-IOV Ethernet logical port from partition lpar1:

```
chhwres -r sriov -m sys1 --rsubtype logport -o r -p lpar1 -a  
"adapter_id=1,logical_port_id=27004001"
```
- ▶ Change the port VLAN ID for an SR-IOV Ethernet logical port in partition lpar1:

```
chhwres -r sriov -m sys1 --rsubtype logport -o s -p lpar1 -a  
"adapter_id=1,logical_port_id=27004001,port_vlan_id=2"
```
- ▶ Set the connection speed for SR-IOV physical port 0 to 100 Gbps:

```
chhwres -r sriov -m sys1 --rsubtype physport -o s -a  
"adapter_id=1,phys_port_id=0,conn_speed=100000"
```
- ▶ Add an SR-IOV Ethernet logical port (using defaults) to partition lpar1:

```
chhwres -r sriov -m sys1 --rsubtype logport -o a -p lpar1 -a  
"adapter_id=1,phys_port_id=1,logical_port_type=eth"
```
- ▶ Remove an SR-IOV Ethernet logical port from partition lpar1:

```
chhwres -r sriov -m sys1 --rsubtype logport -o r -p lpar1 -a  
"adapter_id=1,logical_port_id=27004001"
```
- ▶ Change the port VLAN ID for an SR-IOV Ethernet logical port in partition lpar1:

```
chhwres -r sriov -m sys1 --rsubtype logport -o s -p lpar1 -a  
"adapter_id=1,logical_port_id=27004001,port_vlan_id=2"
```
- ▶ Reset the statistics for an SR-IOV physical port:

```
chhwres -r sriov -m sys1 --rsubtype physport -o rs -a  
"adapter_id=1,phys_port_id=0"
```

For more details about the command syntax, values, and attributes, see the man pages.

DLPAR

With the **chhwres** command, you can also perform some DLPAR operations such as add, edit, or remove. The remove function might be necessary if a logical port is assigned to a partition that should be moved to another system with the Live Partition Mobility function.

- ▶ Add an logical port (VF):

```
chhwres -m astrosfsp1 -r sriov --rsubtype logport --id 2 -o a -a
"adapter_id=1,phys_port_id=3,logical_port_type=eth,capacity=4,promisc_mode=0,po
rt_vlan_id=2,pvid_priority=5,allowed_vlan_ids=\"100,101\",allowed_os_mac_addrs=
\"02123456789a,02123456789b\""
```

- ▶ Edit some settings:

```
chhwres -m sys1 -r sriov --rsubtype logport -p mylpar -o s -a
"adapter_id=1,logical_port_id=27004001,allowed_vlan_ids+=102"
```

- ▶ Remove a logical port (VF):

```
chhwres -m sys1 -r sriov --rsubtype logport -p mylpar -o r -a
"adapter_id=1,logical_port_id=27004001" (remove)
```




IBM Power Systems SR-IOV Technical Overview and Introduction



**See how SR-IOV
minimizes contention
with CPU and memory
resources**

**Explore powerful
adapter-based
virtualization for
logical partitions**

**Learn about industry
standard PCI
specification**

This IBM Redpaper publication describes the adapter-based virtualization capabilities that are being deployed in high-end IBM POWER7+ processor-based servers.

Peripheral Component Interconnect Express (PCIe) single root I/O virtualization (SR-IOV) is a virtualization technology on IBM Power Systems servers. SR-IOV allows multiple logical partitions (LPARs) to share a PCIe adapter with little or no run time involvement of a hypervisor or other virtualization intermediary.

SR-IOV does not replace the existing virtualization capabilities that are offered as part of the IBM PowerVM offerings. Rather, SR-IOV complements them with additional capabilities.

This paper describes many aspects of the SR-IOV technology:

- ▶ A comparison of SR-IOV with standard virtualization technology
- ▶ Architectural overview of SR-IOV
- ▶ Overall benefits of SR-IOV
- ▶ Planning requirements
- ▶ SR-IOV deployment models that use standard I/O virtualization
- ▶ Configuring the adapter for dedicated or shared modes
- ▶ Tips for maintaining and troubleshooting your system
- ▶ Scenarios for configuring your system

This paper is directed to clients, IBM Business Partners, and system administrators who are involved with planning, deploying, configuring, and maintaining key virtualization technologies.

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