Fibre Channel Protocol for Linux and z/VM on IBM System z

FCP concepts

FCP multipathing

N_Port virtualization

Steve Gracin
John Schnitzler Jr
Steve Wilkins
Jon vonWolfersdorf
Frederick Wong

Redbooks

ibm.com/redbooks

International Technical Support Organization

Fibre Channel Protocol for Linux and z/VM on IBM System z

December 2007

SG24-7266-00

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

First Edition (December 2007)

This edition applies to Version 5, Release 2 of z/VM (product number 5741-A05), Linux SLES10 and RHEL5.

Contents

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 about the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

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

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

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

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

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

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

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

COPYRIGHT LICENSE:
This information contains sample application programs in source language, which illustrates 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. 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 IBM application programming interfaces.

Trademarks

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

Redbooks (logo) ®
z/VM®
z/VSE™
zSeries®
z9™
DS6000™
DS8000™
Enterprise Storage Server®
ECKD™
ESCON®
FICON®
IBM®
Redbooks®
S/360™
System i™
System p™
System z™
System z9™
System Storage™
TotalStorage®

The following terms are trademarks of other companies:

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

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

Preface

This IBM® Redbooks® publication builds upon the existing *Linux for zSeriesLinux on zSeries: Fibre Channel Protocol Implementation Guide*, SG24-6344, to produce a System z™ book that includes (among other things):

- A general introduction to Fibre Channel Protocol (FCP) concepts
- Operating system support and usage
- FCP naming and addressing
- FCP devices in the 2.6 Linux® kernel
- N-Port ID Virtualization
- FCP Security topics

The team that wrote this book

This book was produced by a team of specialists from around the world working for the International Technical Support Organization.

Steve Gracin is a Certified Consulting IT Specialist in zVM and Linux with the Advanced Technical Support team in Endicott, NY. He has worked at IBM for 34 years, working the entire time with IBM mainframe computing, beginning with DOS/VS and VM/SP to zVM and Linux. His areas of expertise include z/VM® and Linux, and he enjoys working with connectivity issues. He has contributed to IBM Redbooks and presented at many technical and marketing conferences on the topics z/VSE™, z/VM, and Linux on System z.

John Schnitzler Jr is a Consulting IT Specialist with the z/VM and Linux Advanced Technical Support Team in Endicott, New York. He has 18 years of experience working in both hardware and software technical support. He has worked at IBM for 29 years. His areas of expertise are z/VM, Linux, and large systems hardware and Configuration. He has presented at numerous technical conferences.

Steve Wilkins is a Senior Software Engineer with z/VM Development in Endicott, New York. He has 19 years of experience at IBM and with z/VM Development. He holds a BS degree in Information Science from the University of Pittsburgh. His area of expertise includes operating system exploitation of enterprise channel and device architectures. He has presented at numerous technical conferences on the subject of z/VM device support.

Jon vonWolfersdorf is a Consulting IT Specialist with the z/VM and Linux on System z Advanced Technical Support Team in Endicott, New York. He has seven years of experience working with Linux on System z. He has worked at IBM for 25 years. He holds a degree in Computer Science from the State University of New York at Binghamton. His areas of expertise are z/VM and Linux on System z. He has presented on the topics z/VSE, z/VM, and Linux on System z at numerous technical conferences.

Frederick Wong is a Product Engineer at IBM Poughkeepsie. He has nine years of experience with channel and connectivity in System z. His areas of expertise include 2074, P390, R390, EMIO, FCP, and OSA. Before joining IBM, he worked for 10 years at the University of Buffalo, NYSIS, as a Senior Application Engineer. He holds MS/BS degree in Electrical Engineering from the University of Buffalo. He has written academic papers about superconductivity. He is a member of IEEE.

A special thanks also to the following people for their contributions to this project:

Lydia Parziale, Project Leader
International Technical Support Organization, Poughkeepsie Center

Gerhard Banzhaf, Stefan Amann
IBM Boeblingen, Germany

Eric Farman, Leslie Geer III, Jim McGinniss, Lisa Reese
IBM Endicott, New York

Richard Lewis
IBM Gaithersburg, Maryland

George P. Kuch, Glenn D. Brooks, Daniel F. Clarke, Leonard J. Dibella
IBM Poughkeepsie, New York

Mark Spencer, Sr. System z Solutions Architect
Red Hat - The Americas

Brad Hinson, Sr. Support Engineer Lead, System z
Red Hat, Inc.

Become a published author

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

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

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

ibm.com/redbooks/residencies.html

Comments welcome

Your comments are important to us!

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

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

  ibm.com/redbooks

- Send your comments in an email to:

  redbook@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

1

Fibre Channel Protocol concepts

This chapter discusses Fibre Channel Protocol (FCP) concepts and explains some major hardware and software requirements. Topics include:

- FCP support
- FCP topologies
- FCP terminology
- FCP access control
- FCP considerations and limitations

1.1 FCP support

The Fibre Channel (FC) standard was developed by the National Committee of Information Technology Standards (NCITS). The System z FCP I/O architecture conforms to the FC standards specified by the NCITS. More detailed information about the FC standards can be obtained from the following Web sites:

http://www.t10.org
http://www.t11.org

System z FCP support enables z/VM and Linux running on System z to access industry-standard SCSI devices. For disk applications, these FCP storage devices utilize Fixed Block (512-byte) sectors rather than Extended Count Key Data (ECKD™) format.

A new channel-path identifier (CHPID) type has been defined for the FICON/FICON® Express feature cards, called *FCP*. The FCP CHPID type is supported on the FICON and FICON Express features of all System z processors.

1.2 FCP topologies

This section provides a high-level overview of the Fibre Channel topologies. The FC architecture defines three separate topologies to support connectivity between endpoints:

- Point-to-point
- Arbitrated loops
- Switched fabric

1.2.1 Point-to-point

This is the simplest topology to configure. A point-to-point configuration is a direct connection between two endpoints. Typically, it consists of a host, a device (such as a disk controller), and a dedicated fiber link (Figure 1-1).

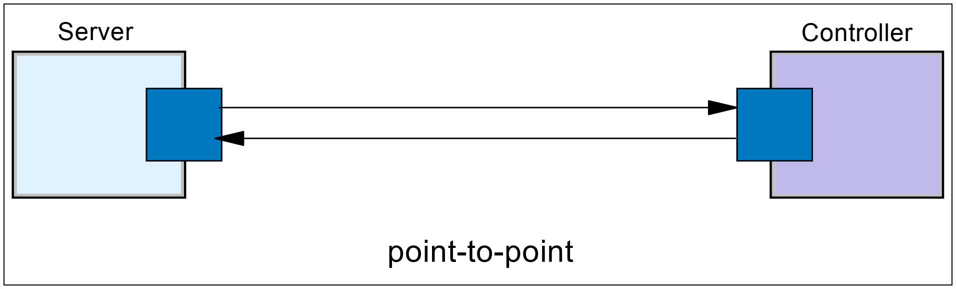


Figure 1-1 Point-to-point configuration

1.2.2 Arbitrated loop

An arbitrated loop is a ring topology that shares the Fibre Channel bandwidth among multiple endpoints. The loop is implemented within a hub that interconnects the endpoints (Figure 1-2). An arbitrated scheme is used to determine which endpoint gets control of the loop. The maximum number of ports is 127.

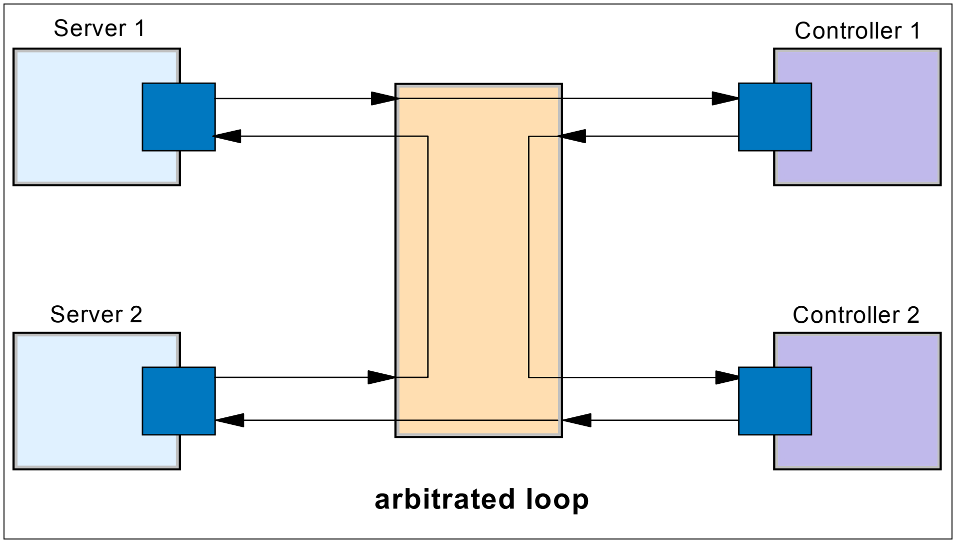


Figure 1-2 Arbitrated loop configuration

1.2.3 Switched fabric

This topology provides the most flexibility and makes the best use of the aggregated bandwidth by the use of switched connections between endpoints. One or more switches are interconnected to create a fabric to which the endpoints are connected (Figure 1-3).

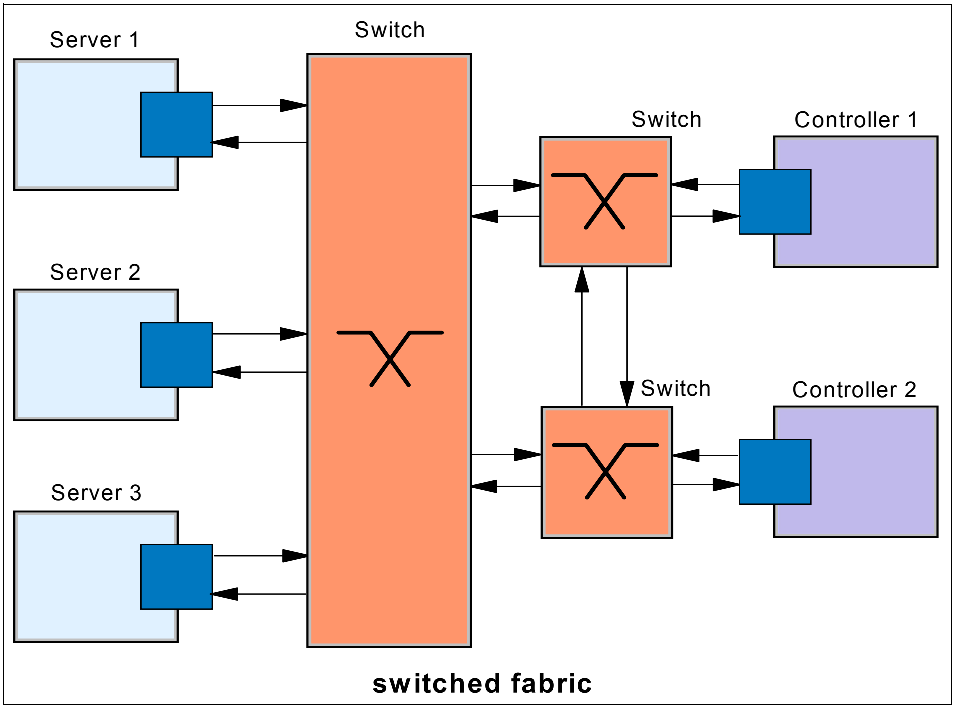


Figure 1-3 Switched fabric configuration

1.2.4 Supported topologies

The supported topologies for the System z FCP function include:

- A point-to-point dedicated fibre link to an FCP device
- A Fibre Channel through a single switch or multiple switches to an FCP device
- A Fibre Channel through a single switch or multiple switches to a Fibre Channel-to-SCSI bridge

Restriction: Point-to-point topology is only supported on the z9™, z990, and z890 processors. Point-to-point topology is supported by the current releases of Redhat and SuSE Linux for System z. z/VM will include point-to-point topology support in Version 5 Release 3. Arbitrated loop topology is not supported as part of the System z FCP enablement.

1.3 FCP terminology

This section discusses some general terms that are used in the FC environment when operating in a point-to-point, arbitrated loop, switched, or bridged configuration.

1.3.1 Node

A node is an endpoint that contains information. It can be a computer (host), a device controller, or a peripheral device (such as disk or tape drives). A node has a unique 64-bit identifier known as the Node_Name. The Node_Name is used for system management purposes.

1.3.2 Port

Each node must have at least one port (hardware interface) to connect the node to the FC topology. This node port is referred to as an N_Port.

Each N_Port has a Port_Name, which is a unique 64-bit identifier that is assigned at the time it is manufactured. The N_Port is used to associate an access point to a node's resources.

Other port types include:

E_Port An *expansion port* is used to interconnect switches and build a switched fabric.

F_Port A *fabric port* is used to connect an N_Port to a switch that is not loop-capable.

FL_Port A *fabric loop port* is used to connect NL_Ports to a switch in a loop configuration.

G_Port A *generic port* is a port that has not assumed a role in the fabric.

L_Port A *loop port* is a port in a Fibre Channel Arbitrated Loop (FC-AL) topology.

NL_Port A *node loop port* is an N_Port with loop capabilities.

The port type is determined by the node’s role in the topology, as shown in Figure 1-4.

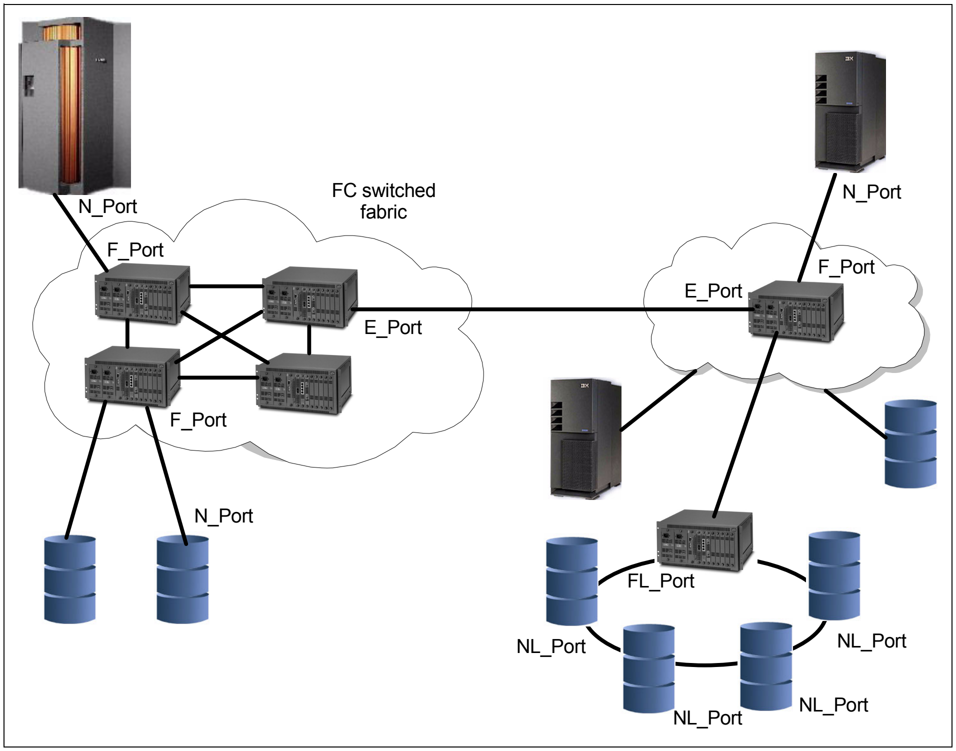


Figure 1-4 FC port types

1.3.3 Link

The port connects to the topology through a link. The link is a fiber optic cable that has two strands, one used to transmit a signal and the other to receive a signal (see Figure 1-5). A link is used to interconnect nodes, switches, or both.

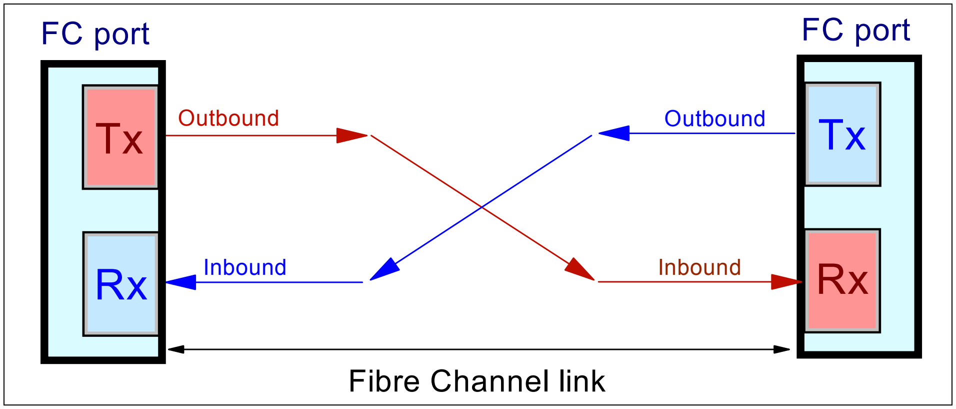


Figure 1-5 Fibre Channel link

For example, a Fibre Channel link (port-to-port connection) can be:

- Node-to-node link (N_Port-to-N_Port)
- Node-to-switch link (N_Port-to-F_Port)
- Loop node-to-switch link (NL_Port-to-FL_Port)
- Switch-to-switch link (E_Port-to-E_Port)

1.3.4 Worldwide names

As mentioned previously, nodes and ports have unique 64-bit addresses that are used to identify them in an FC topology. These addresses are assigned by the manufacturer, with a vendor-specific portion defined by the IEEE standards committee. These addresses (in the FC standard) are called *node names* and port names, and when they are worldwide unique, they are referred to as:

- Worldwide node name (WWNN)
- Worldwide port name (WWPN)

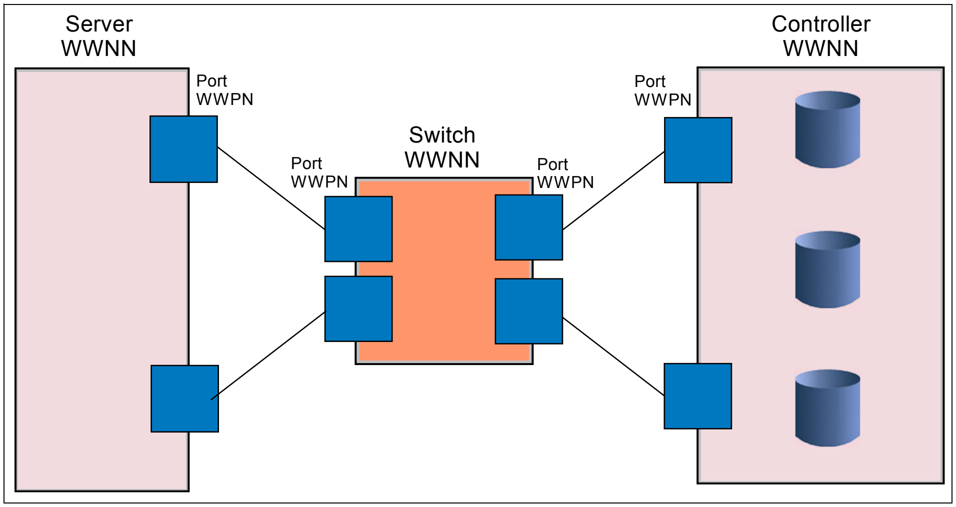


Figure 1-6 Example of worldwide names

1.4 FCP access control

The ability to control access to nodes and devices is provided as a function in switches and controllers, and is called *LUN masking* and *zoning*. Logical unit number (LUN) masking and zoning can be used to prevent servers from accessing storage that they are not permitted to access.

1.4.1 LUN masking

A LUN represents a portion of a controller, such as a disk device. With the use of LUNs, a controller can be logically divided into independent partitions. Access to these LUNs can be restricted to distinctive WWPNs as part of the controller configuration. This method is known as LUN masking. Details are discussed in 9.1, “LUN masking” on page 162.

1.4.2 Zoning

Segmentation of a switched fabric is achieved through zoning. It can be used to partition off certain portions of the switched fabric, allowing only the members of a zone to communicate within that zone. All others attempting to access from outside that zone are rejected. Details on zoning can be found in 9.2, “Zoning” on page 163.

1.5 FCP considerations and limitations

FCP usage from System z is new, and will certainly evolve over time. Likewise, the switched fabric capabilities will continue to evolve.

1.5.1 Channel sharing

When multiple Linux images share an FCP channel, each Linux image is identified to the fabric by the permanent WWPN of the FCP port and has access to all devices that are connected to the FCP fabric. Because the Linux images all share the same FCP channel, each Linux image uses the same WWPN to enter the fabric (and thus become indistinguishable from one another). This design means that the use of zoning and LUN masking cannot effectively create appropriate access controls among the Linux images.

N_Port ID Virtualization (NPIV), a new feature available with IBM System z9™ 109 servers, allows a single FCP port to register multiple WWPNs with a fabric name server. Each registered WWPN is assigned a unique N_Port ID. With NPIV, a single FCP port can appear as multiple WWPNs in the FCP fabric.

Use of NPIV along with zoning and LUN masking can ensure data integrity among Linux images sharing an FCP channel. For design considerations and the prerequisites of NPIV, see 10.1, “Overview of the NPIV feature” on page 170.

1.5.2 Device sharing

You can share a device among logical partitions by:

- Using a separate channel for each logical partition

  In order for two or more unique operating system instances to share concurrent access to a single Fibre Channel or SCSI device LUN, each of these operating systems has to access this device through a different System z FCP channel.

- Using a shared channel

  Although multiple operating systems can concurrently access the same remote Fibre Channel port via a single FCP channel, Fibre Channel devices (identified by their LUNs) can only be serially reused.

1.5.3 Supported devices

A large number of SCSI devices exist in the marketplace, especially when older devices are included. IBM cannot fully test all of these for use in System z systems. For a complete list of supported devices, consult:

`http://www.ibm.com/servers/eserver/zseries/connectivity/`

2

Installing z/VM to SCSI disks

In this chapter we describe how to install and configure z/VM v5.2 to FCP-attached SCSI disks. We explain the hardware and software configurations that were used for the installations, and why. z/VM v5.2 was first installed to a ESS2105 Disk Storage Subsystem, then a DS6000™.

2.1 Installation planning

To begin the planning of our installation of z/VM v5.2 on SCSI we reviewed the *Guide For Automated Installation and Service v5r2*, GC-24-6099-02, the latest edition available. The Plan Your Installation sections of the manual were reviewed and the associated worksheets were completed. There is detailed information you will be prompted for during installation. You will want the answers readily available. It is important that the worksheets be completed prior to installation.

The major items for installation of z/VM are memory, processors, TCP/IP connectivity, storage, and console device. These are items typically discussed and agreed upon before an LPAR is defined during a solutions assurance review.

2.2 Hardware configuration

The installation of z/VM v5.2 to SCSI was accomplished using an LPAR of a z900 processor that had eight standard central processors (CPs) and 512 Meg memory. To support TCP/IP, the LPAR had a Gigabit OSA/Express2 card connected to the IBM corporate network. These parts of the install configuration remained constant. The storage subsystem used was the only difference between the two installations of z/VM v5.2.

2.3 z/VM v5.2 distribution

The z/VM v5.2 installation media ordered was a DVD to be installed from the Z900 Hardware Management Console (HMC). Installing to SCSI disks is not an option with the zVM5.2 tape distribution.

2.4 z/VM operating with SCSI disks overview

As described in its installation/service guide, z/VM v5.2 requires four SCSI disks as the system volumes: 520RES, 520SPL, 520PAG, and 520W01, each with a minimum of 3.5 Gig. They look to z/VM as FBA DASD 8000, 8001, 8002, and 8003 (see Figure 2-1 on page 13).

In order to make use of a SCSI disk there is code in z/VM CP called the SCSI Container. This code manages z/VM SCSI device drivers that make attaching SCSI disks by FCP channel to z/VM possible. There is a z/VM CP FBA

Emulation Layer that operates above the SCSI Container that makes the SCSI disks look to the rest of the z/VM operating system as FBA DASD. The disks appear as physical 9336 FBA device types.

To install z/VM v5.2 to SCSI disks, z/VM first needs to be aware of the disks. This will be the FCP/WWPN/LUN triplet. Then for z/VM to use the disks for CMS minidisks, for example, the EDEVICE command provides an emulated FBA address. See Figure 2-1.

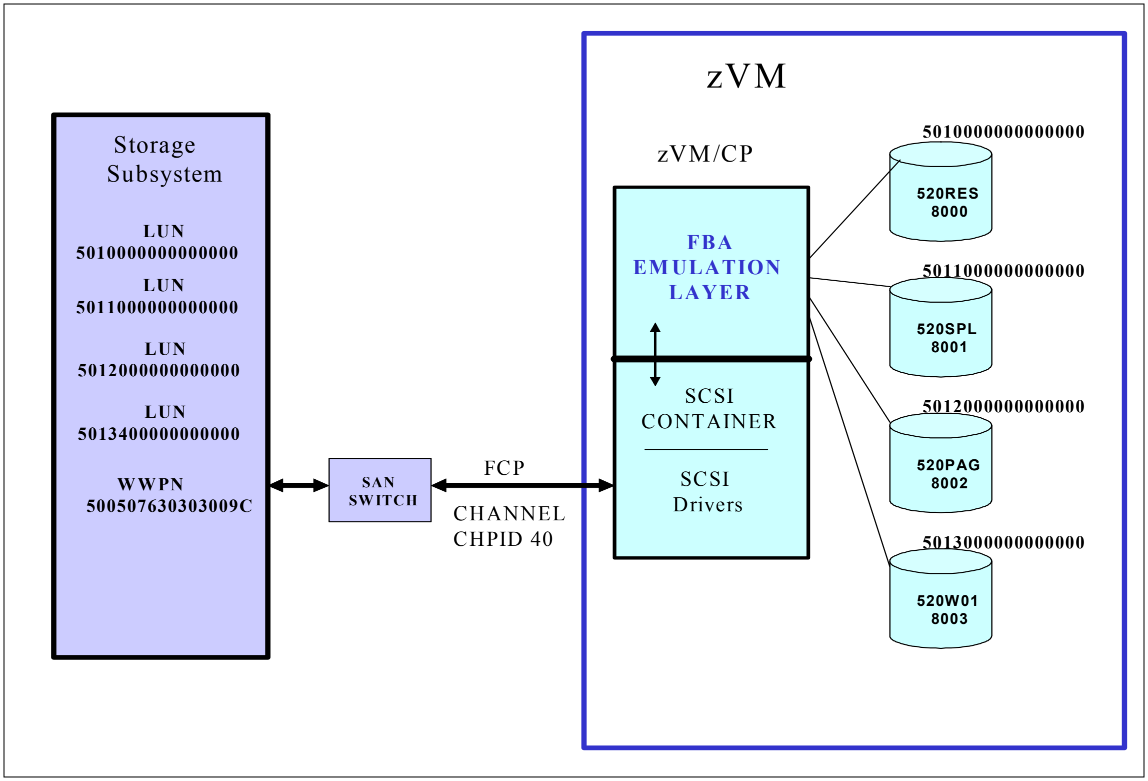


Figure 2-1 z/VM and SCSI disks

2.5 Installation with ESS2105-F20

The storage-related information required for the installation of z/VM v5.2 on the DS8000™ is listed in Table 2-1. These are all the items needed to define the SCSI disks to z/VM as emulated FBA devices. Four disks are required.

Table 2-1 Storage information

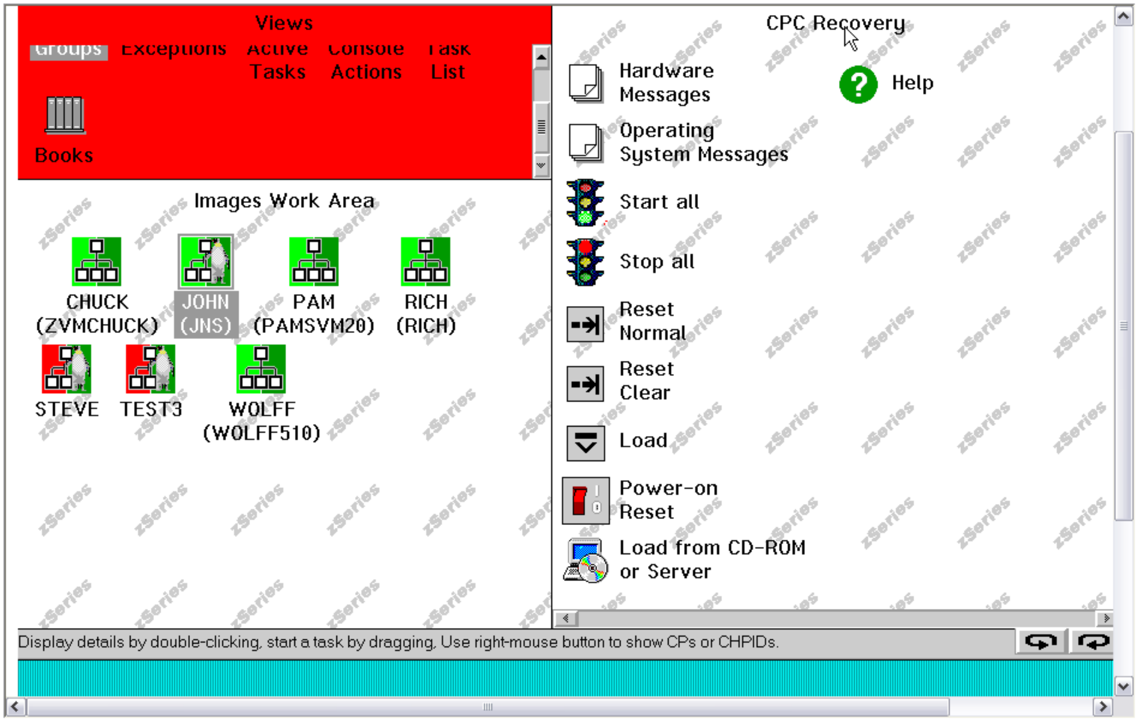
VOLID	EDEVICE	FCP ADDRESS	WWPN	LUN
520RES	8000	6002	500507640100556E	5010000000000000
520SPL	8001	6003	500507640100556E	5011000000000000
520PAG	8002	6006	500507640100556E	5012000000000000
520W01	8003	6007	500507640100556E	5013000000000000

The information items required for the installation on SCSI disk are the target WWPN and the LUNs. z/VM does not make use of a WWNN. The EDEVICE addresses are virtual addresses that z/VM uses to address the SCSI disks as emulated FBA devices. The EDEVICE addresses do not have to be a particular address range. These are available virtual addresses the you decide to use. The FCP addresses are the physical addresses that are defined in the IOCP and assisgned in our case to CHPID 40. The function in z/VM CP will make the SCSI LUNs look to z/VM itself as 9336 FBA DASD at the EDEVICE addresses.

All WWPN, LUN, and FCP address information was provided to us by the storage management team. Our decision was what EDEVICE addresses to use.

2.6 Installation from DVD

The z/VM v5.2 installation from DVD requires that the z/VM DVD be mounted in the Hardware Management Console (HMC) DVD drive and that we log in to the HMC as user SYSPROG (see Figure 2-2). Our image, STEVE, was placed into Single Image Operations mode, which means that we connect to and use the desktop of the System Element (SE) from the HMC. The SE is a mobile computer within the z9, which has the Load from CD-ROM DVD or server icon. We noticed that it was not possible to start the integrated 3270 Console once in Single Image Operations mode, so be sure to start the Integrated 3270 Console before entering Single Image Operations mode.



Display details by double-clicking, start a task by dragging, Use right-mouse button to show CPs or CHPIDs.

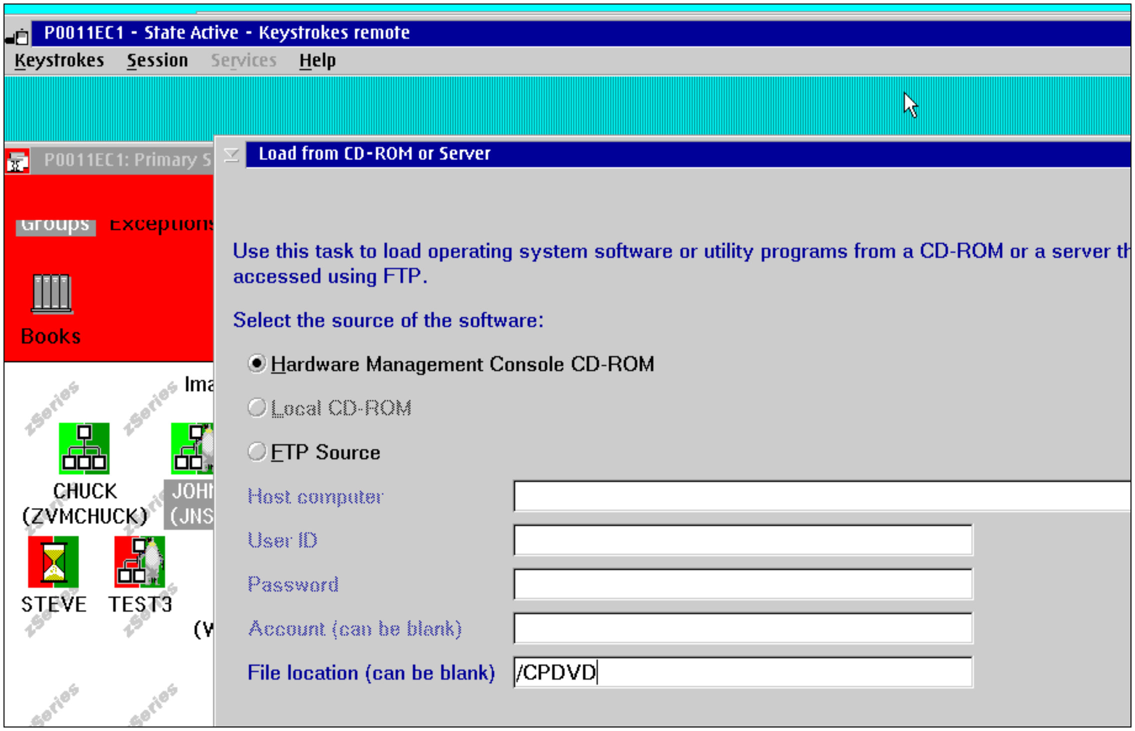
Figure 2-2 The HMC

2.7 Loading From DVD

In Single Image Operations mode we then activated the Load from CD-ROM DVD or Server by selecting the icon on the desktop. The menu that was displayed next allows us to enter the information required to boot z/VM v5.2 from DVD. For our installation, which was z/VM v5.2 first level, we used the First-Level

Installation Method path of the z/VM Guide for Automated Installation and Service manual. The method of installation is one of the questions asked and answered when completing the installation worksheets in the Installation and Service Guide.

In the Load from CD-ROM or Server window (see Figure 2-3) we picked the **Hardware Management Console CD-ROM/DVD** selection. The only field that required an entry was the file location, which is not optional in this case. We entered /CPDVD.



Load from CD-ROM or Server

Use this task to load operating system software or utility programs from a CD-ROM or a server th accessed using FTP.

Select the source of the software:

● Hardware Management Console CD-ROM

○ Local CD-ROM

○ FTP Source

Host computer

User ID

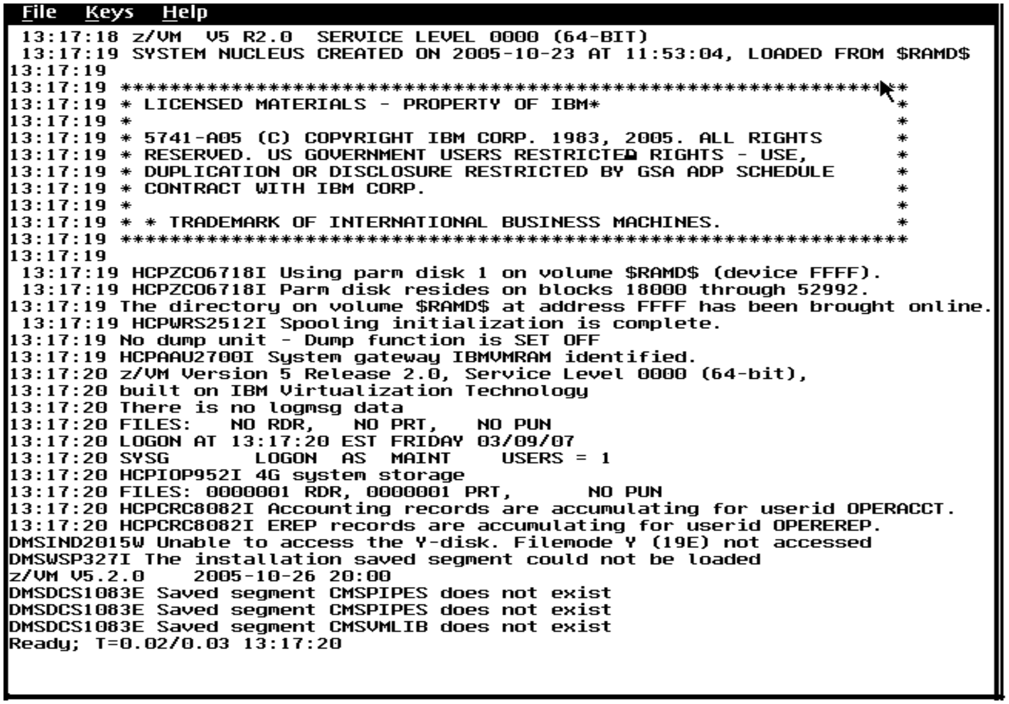
Password

Account (can be blank)

File location (can be blank) /CPDVD

Figure 2-3 Indicating the source of the software

We replied **OK** to begin the load of z/VM v5.2, and after some time the result was IPL messages from the z/VM v5.2 system on the Integrated 3270 Console (see Figure 2-4). When the ready prompt appeared, the RAMDISK System is established.



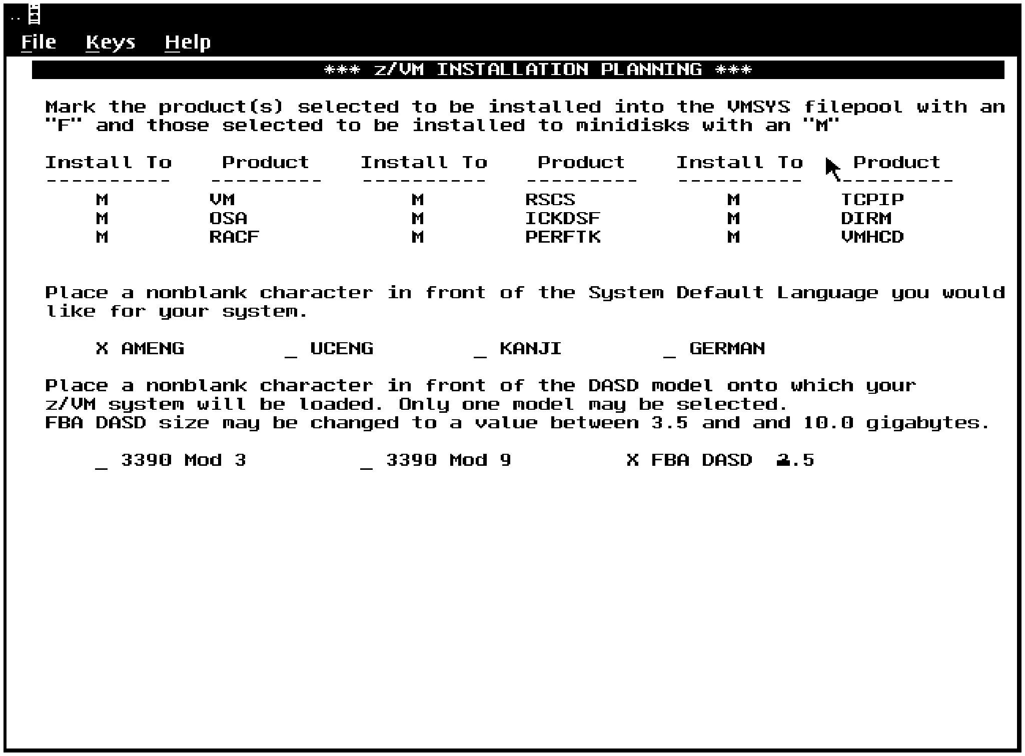
```
 13:17:18 z/VM  V5 R2.0  SERVICE LEVEL 0000 (64-BIT)
 13:17:19 SYSTEM NUCLEUS CREATED ON 2005-10-23 AT 11:53:04, LOADED FROM $RAMD$
13:17:19
13:17:19 ****************************************************************
13:17:19 * LICENSED MATERIALS - PROPERTY OF IBM*                        *
13:17:19 *                                                              *
13:17:19 * 5741-A05 (C) COPYRIGHT IBM CORP. 1983, 2005. ALL RIGHTS      *
13:17:19 * RESERVED. US GOVERNMENT USERS RESTRICTED RIGHTS - USE,       *
13:17:19 * DUPLICATION OR DISCLOSURE RESTRICTED BY GSA ADP SCHEDULE     *
13:17:19 * CONTRACT WITH IBM CORP.                                      *
13:17:19 *                                                              *
13:17:19 * * TRADEMARK OF INTERNATIONAL BUSINESS MACHINES.              *
13:17:19 ****************************************************************
13:17:19
 13:17:19 HCPZCO6718I Using parm disk 1 on volume $RAMD$ (device FFFF).
 13:17:19 HCPZCO6718I Parm disk resides on blocks 18000 through 52992.
13:17:19 The directory on volume $RAMD$ at address FFFF has been brought online.
 13:17:19 HCPWRS2512I Spooling initialization is complete.
13:17:19 No dump unit - Dump function is SET OFF
13:17:19 HCPAAU2700I System gateway IBMVMRAM identified.
13:17:20 z/VM Version 5 Release 2.0, Service Level 0000 (64-bit),
13:17:20 built on IBM Virtualization Technology
13:17:20 There is no logmsg data
13:17:20 FILES:   NO RDR,   NO PRT,   NO PUN
13:17:20 LOGON AT 13:17:20 EST FRIDAY 03/09/07
13:17:20 SYSG       LOGON  AS  MAINT    USERS = 1
13:17:20 HCPIOP952I 4G system storage
13:17:20 FILES: 0000001 RDR, 0000001 PRT,      NO PUN
13:17:20 HCPCRC8082I Accounting records are accumulating for userid OPERACCT.
13:17:20 HCPCRC8082I EREP records are accumulating for userid OPEREREP.
DMSIND2015W Unable to access the Y-disk. Filemode Y (19E) not accessed
DMSWSP327I The installation saved segment could not be loaded
z/VM V5.2.0    2005-10-26 20:00
DMSDCS1083E Saved segment CMSPIPES does not exist
DMSDCS1083E Saved segment CMSPIPES does not exist
DMSDCS1083E Saved segment CMSVMLIB does not exist
Ready; T=0.02/0.03 13:17:20
```

Figure 2-4 Loading z/VM v5.2

2.8 z/VM RAMDISK

A difference between installing z/VM v5.2 from tape versus booting from the DVD distribution is that z/VM v5.2 from DVD boots to the zVM RAMDISK system. You have a z/VM v5.2 running from a virtual disk in processor memory.

After z/VM v5.2 is established and logged on as MAINT, the INSTPLAN EXEC is run (see Figure 2-5), which is the same EXEC used for installation from any media. There is an option presented by this EXEC that is unique to SCSI Disk. To let the zVM install process know that SCSI Disk will be used, the option FBA DASD 3.5 is selected.

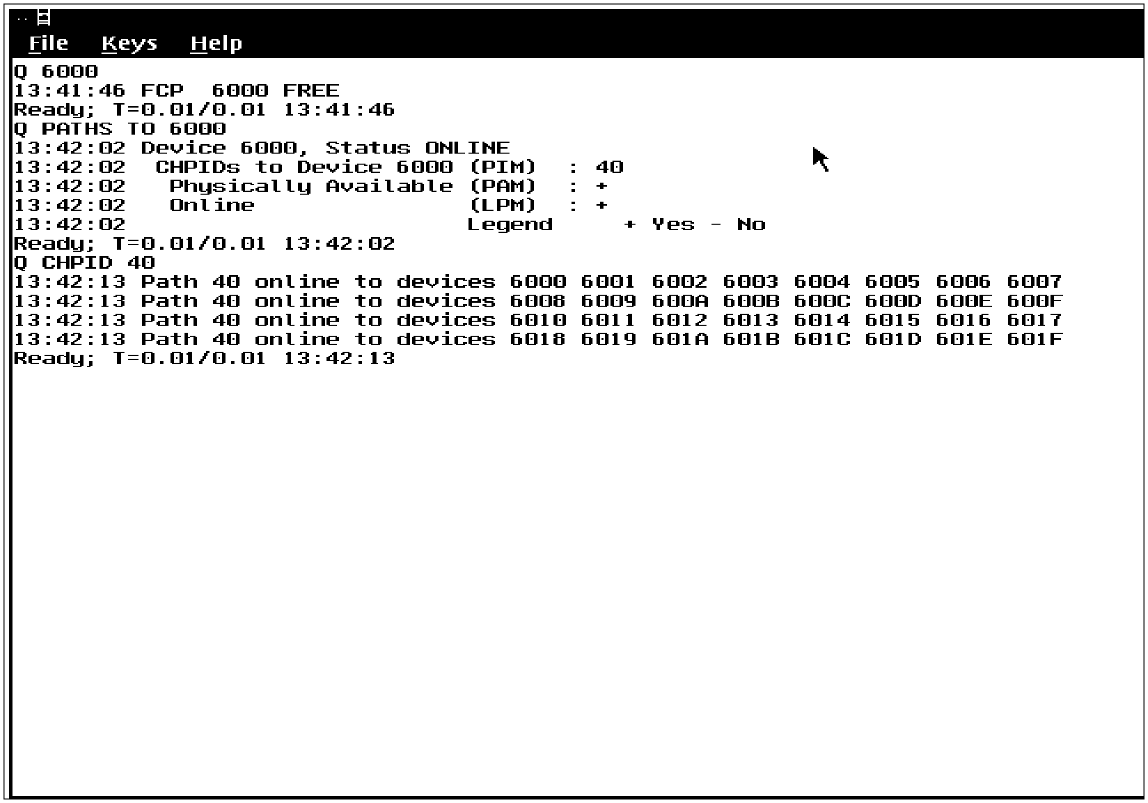


```
                   *** z/VM INSTALLATION PLANNING ***

 Mark the product(s) selected to be installed into the VMSYS filepool with an
 "F" and those selected to be installed to minidisks with an "M"

 Install To    Product    Install To    Product    Install To    Product
 ----------   ---------   ----------   ---------   ----------   ---------
     M        VM              M        RSCS            M        TCPIP
     M        OSA             M        ICKDSF          M        DIRM
     M        RACF            M        PERFTK          M        VMHCD


 Place a nonblank character in front of the System Default Language you would
 like for your system.

     X AMENG        _ UCENG        _ KANJI        _ GERMAN

 Place a nonblank character in front of the DASD model onto which your
 z/VM system will be loaded. Only one model may be selected.
 FBA DASD size may be changed to a value between 3.5 and and 10.0 gigabytes.

     _ 3390 Mod 3         _ 3390 Mod 9         X FBA DASD  3.5
```

Figure 2-5 Installation planning - the INSTPLAN EXEC

When FBA DASD 3.5 is selected, it implies a 3.5 Gig LUN, a SCSI device that will appear to z/VM v5.2 as FBA DASD, which will be used for the zVM volumes 520RES, 520SPL, 520PAG, and 520W01. z/VM supports from 3.5 Gig to a maximum of 10 Gig in size for the install volumes. We found the default of 3.5 Gig per volume to be a good choice.

After the selections were made, we used PF5 to start the processing of the EXEC. Upon successful completion, the next step was to verify that the FCP addresses were available. For the first level install of z/VM v5.2, we queried channel path ID 40 to verify the existence of the FCP addresses (see Figure 2-6).



```
Q 6000
13:41:46 FCP  6000 FREE
Ready; T=0.01/0.01 13:41:46
Q PATHS TO 6000
13:42:02 Device 6000, Status ONLINE
13:42:02  CHPIDs to Device 6000 (PIM)  : 40
13:42:02   Physically Available (PAM)  : +
13:42:02   Online               (LPM)  : +
13:42:02                        Legend     + Yes - No
Ready; T=0.01/0.01 13:42:02
Q CHPID 40
13:42:13 Path 40 online to devices 6000 6001 6002 6003 6004 6005 6006 6007
13:42:13 Path 40 online to devices 6008 6009 600A 600B 600C 600D 600E 600F
13:42:13 Path 40 online to devices 6010 6011 6012 6013 6014 6015 6016 6017
13:42:13 Path 40 online to devices 6018 6019 601A 601B 601C 601D 601E 601F
Ready; T=0.01/0.01 13:42:13
```

Figure 2-6 Query chpid command

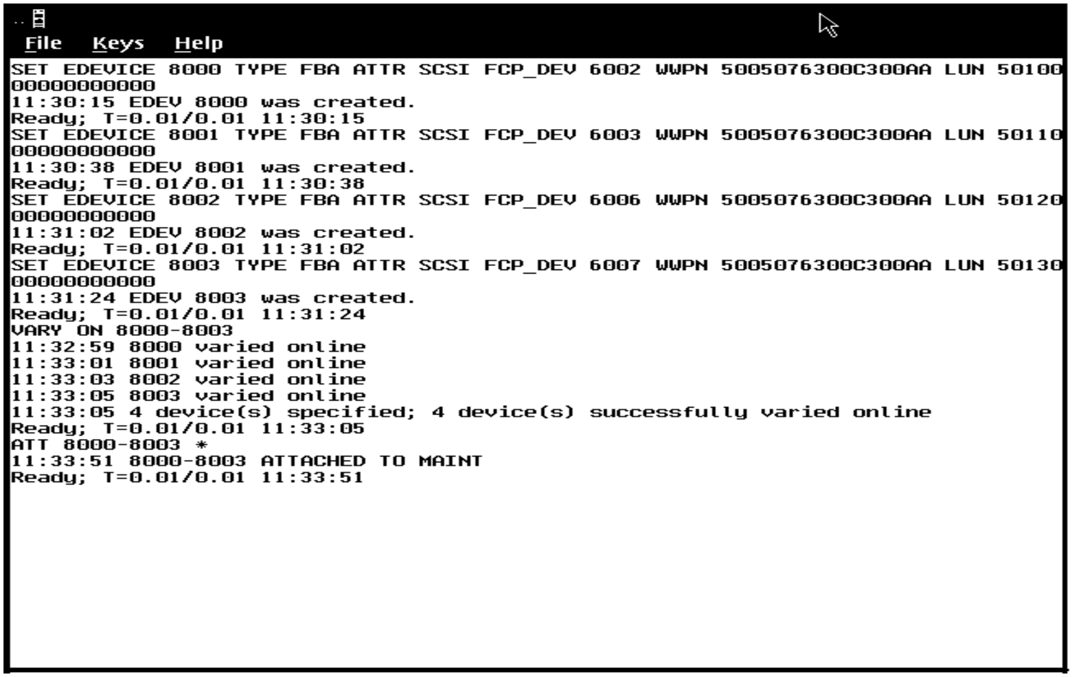
The result of the **`Query Chpid`** command shows that 6000 through 6000F and 6010 through 601F are available. At this point it is possible to define the devices to z/VM. The LUNs will be made to look to z/VM as Fixed Block Architecture (FBA) DASD. z/VM uses this Emulated FBA for its system volumes.

2.9 The Set EDEVICE command

The association in z/VM between what is referred to as a path triplet of FCP_DEV, WWPN, and LUN, to a virtual address that will represent FBA DASD is set by the EDEVICE command. An Emulated Device command is issued for each z/VM volume to be used for the installation procedure. In our case the

EDEVICEs are addresses 8000 through 8003. These addresses are virtual, and there is no special significance to 8000 through 8003 other then that this is the virtual address range we decided to use. They could have be any four available virtual addresses.

After ensuring that the FCP channel paths to the SCSI devices are available, a **`set edvice`** command was issued for each emulated device address in our range of 8000 to 8003 (see Figure 2-7 for output from this command).

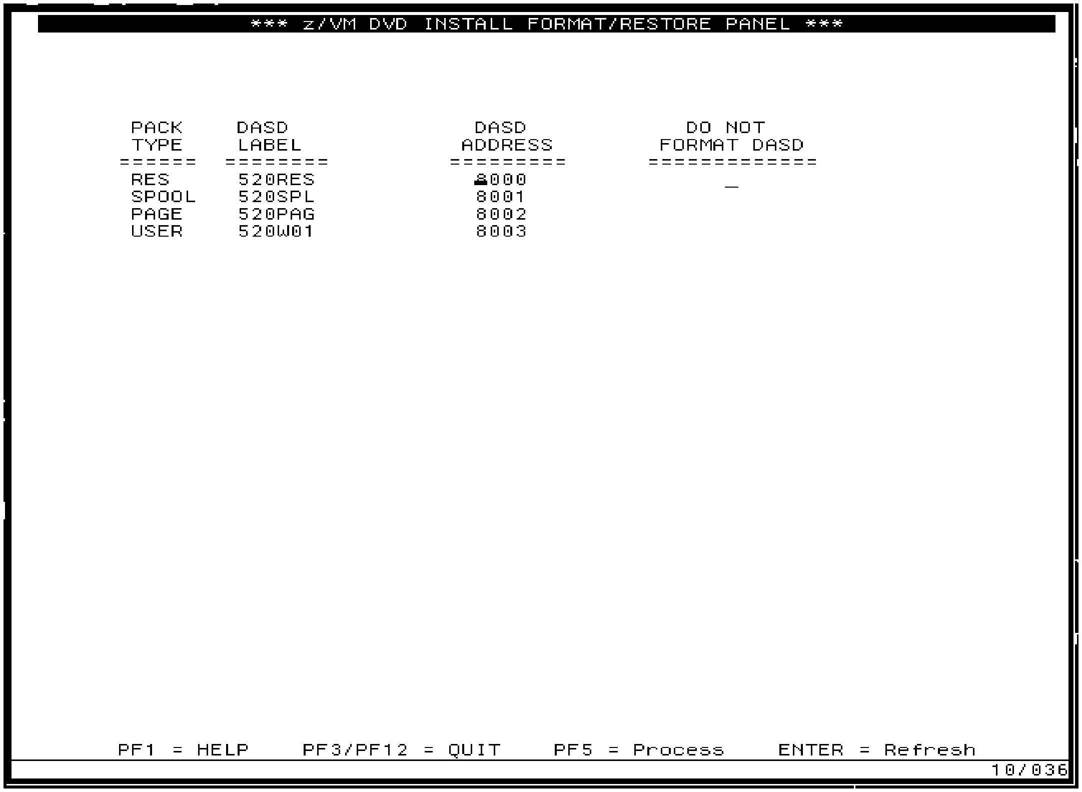


```
SET EDEVICE 8000 TYPE FBA ATTR SCSI FCP_DEV 6002 WWPN 5005076300C300AA LUN 50100
00000000000
11:30:15 EDEV 8000 was created.
Ready; T=0.01/0.01 11:30:15
SET EDEVICE 8001 TYPE FBA ATTR SCSI FCP_DEV 6003 WWPN 5005076300C300AA LUN 50110
00000000000
11:30:38 EDEV 8001 was created.
Ready; T=0.01/0.01 11:30:38
SET EDEVICE 8002 TYPE FBA ATTR SCSI FCP_DEV 6006 WWPN 5005076300C300AA LUN 50120
00000000000
11:31:02 EDEV 8002 was created.
Ready; T=0.01/0.01 11:31:02
SET EDEVICE 8003 TYPE FBA ATTR SCSI FCP_DEV 6007 WWPN 5005076300C300AA LUN 50130
00000000000
11:31:24 EDEV 8003 was created.
Ready; T=0.01/0.01 11:31:24
VARY ON 8000-8003
11:32:59 8000 varied online
11:33:01 8001 varied online
11:33:03 8002 varied online
11:33:05 8003 varied online
11:33:05 4 device(s) specified; 4 device(s) successfully varied online
Ready; T=0.01/0.01 11:33:05
ATT 8000-8003 *
11:33:51 8000-8003 ATTACHED TO MAINT
Ready; T=0.01/0.01 11:33:51
```

Figure 2-7 Output from the SET EDEVICE command

2.10 INSTDVD EXEC

The INSTDVD EXEC begins with a panel in which you are required to enter a DASD ADDRESS for each system pack. The DASD ADDRESS that is assigned to each DASD LABEL is an emulated FBA device that was previously defined with an EDEVICE command. See Figure 2-8 for the output from this command in our system. The DASD ADDRESS is how an individual pack is known to z/VM. The DASD ADDRESS to FCP_DEV/WWPN/LUN translation is accomplished by the SCSI driver.



```
                  *** z/VM DVD INSTALL FORMAT/RESTORE PANEL ***



     PACK      DASD                DASD              DO NOT
     TYPE      LABEL              ADDRESS          FORMAT DASD
    ======    ========           =========        =============
     RES       520RES              8000                 _
     SPOOL     520SPL              8001
     PAGE      520PAG              8002
     USER      520W01              8003




    PF1 = HELP    PF3/PF12 = QUIT    PF5 = Process    ENTER = Refresh
                                                                  10/036
```

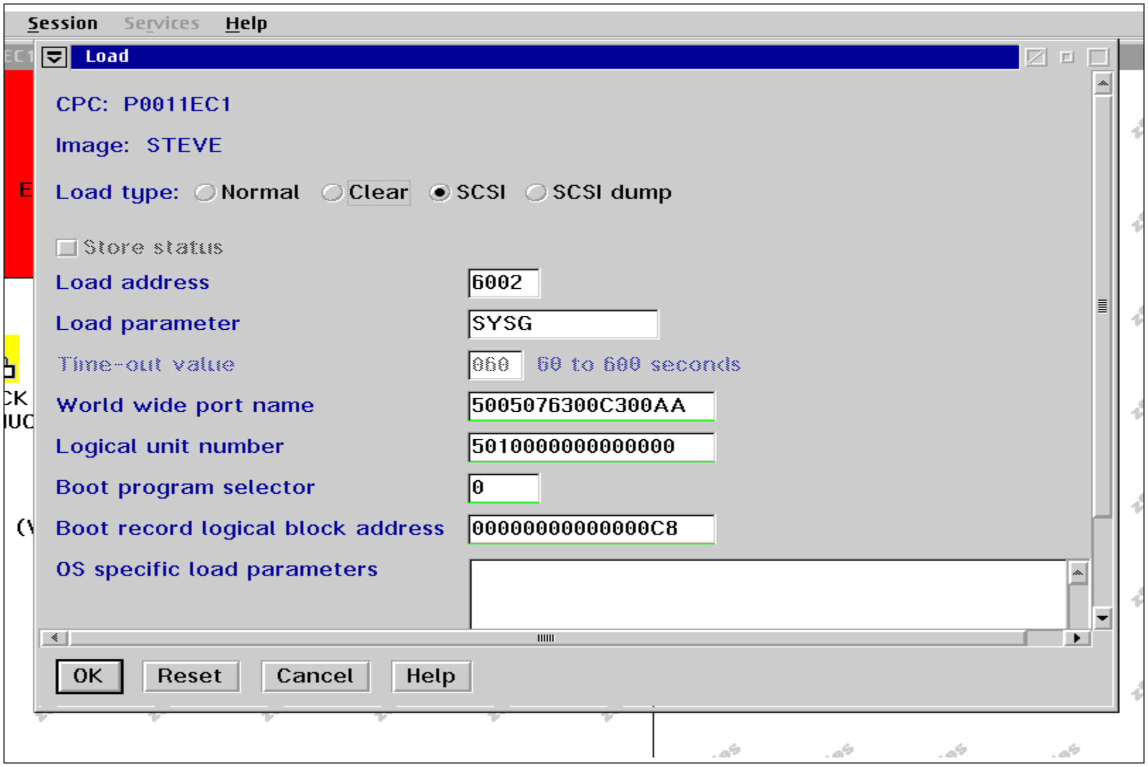
Figure 2-8 Output from the INSTDVD EXEC

With all the entries complete, press PF5 to begin processing and loading z/VM from the DVD.

2.10.1 IPL

When INSTDVD completes, the next step is to IPL the system.

At the HMC CPC Recovery panel we selected the Load icon (Figure 2-9). Selecting the **SCSI** button of the Load panel allows entering the WWPN and LUN information of the 520RES pack.



CPC: P0011EC1

Image: STEVE

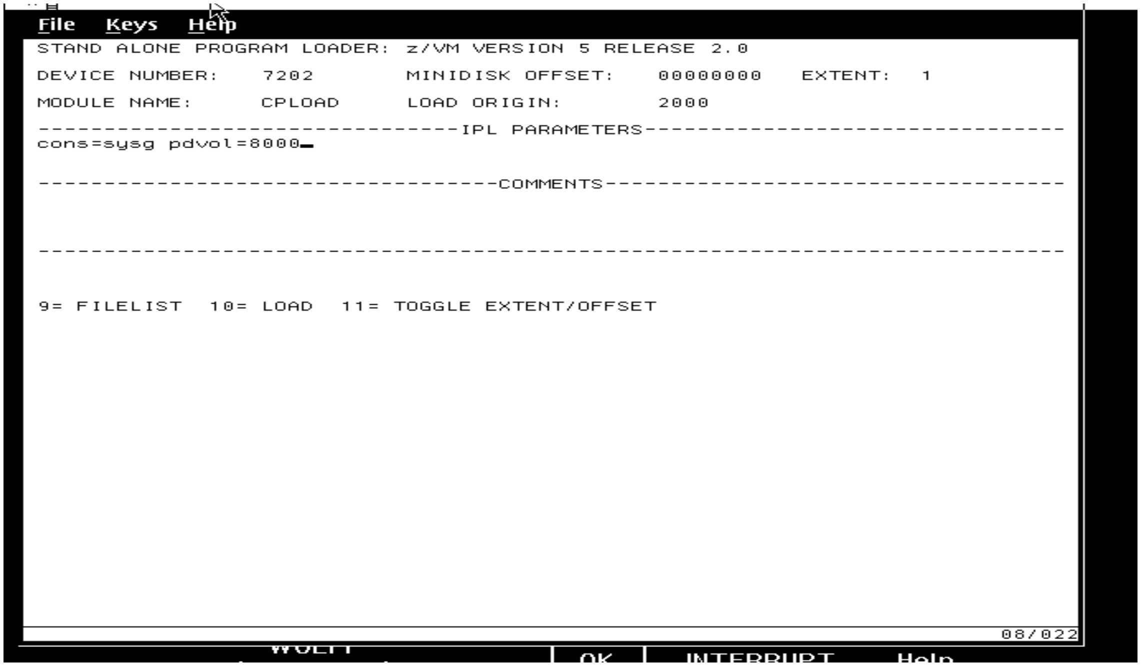
Load type: ○ Normal ○ Clear ● SCSI ○ SCSI dump

□ Store status

Load address	6002
Load parameter	SYSG
Time-out value	060 60 to 600 seconds
World wide port name	5005076300C300AA
Logical unit number	5010000000000000
Boot program selector	0
Boot record logical block address	00000000000000C8
OS specific load parameters	

Figure 2-9 Selecting the SCSI button of the Load panel

The IPL of z/VM v5.2 from that SCSI disk requires that the z/VM Stand Alone Program Loader (SAPL) provide console and parameter disk information (Figure 2-10).



```
STAND ALONE PROGRAM LOADER: z/VM VERSION 5 RELEASE 2.0

DEVICE NUMBER:    7202        MINIDISK OFFSET:    00000000   EXTENT:   1

MODULE NAME:      CPLOAD      LOAD ORIGIN:        2000

--------------------------------IPL PARAMETERS---------------------------------
cons=sysg pdvol=8000

-----------------------------------COMMENTS-----------------------------------------


-------------------------------------------------------------------------------


9= FILELIST  10= LOAD  11= TOGGLE EXTENT/OFFSET

                                                                        08/022
```

Figure 2-10 The Stand Alone Program Loader

The cons=sysg z/VM IPL parameter instructs zVM to use the HMC Integrated 3270 Console for the IPL, the Parm Disk Volume parameter. pdvol=8000 tells z/VM v5.2 that 8000 is the address of the Emulated FBA device that contains the z/VM/CP IPL parameters (520RES). z/VM CP uses the EDEVICE information to associate the Emulated FBA device 8000 with its FCP/WWPN/LUN.

2.11 Installation with DS6000

There is no difference in steps needed for the installation of z/VM v5.2 from the DVD distribution to a DS6000 Storage Subsystem versus a ESS2105-F20 or a DS8000. You still must fill out the worksheets and obtain the same information, FCP_DEV/WWPN/LUN triplets, and edevice addresses.

There are two items that must be taken into account because of the DS6000. One is that this machine has a preferred path. The preferred path is the path triplet of FCP_DEV/WWPN/LUN that is the best one (fastest) to use when accessing a LUN. The DS6000, due to its physical, design has this one path that will result in best performance, so it is the preferred path to use.

Secondly, a different attribute, 1750 instead of SCSI, is used by zVM5.2 for the DS6000. Both the attr of 1750 and the preferred path (pref) differences are accounted for in the **set edevice** command. For a DS6000, our **set edevice** command would look like:

```
set edevice 8000 type fba attr 1750 fcp_dev 1f60 wwpn 500507630303009c
lun 4011401a00000000 pref
```

The **set edevice** command is the only difference between installing on a DS6000 versus a ESS2105-F20, no matter what the storage subsystem z/VM sees.

3

Multipath FCP with z/VM

This chapter discusses multipath FCP support within zVM5.2. Multipath support allows access to a LUN through multiple physical paths. Each path is defined as an FCP_DEV/WWPN/LUN triplet. The multiple paths provide backup in the event of a hardware failure.

3.1 Adding a path

When the installation of z/VM v5.2 is complete, there is one path through chpid 40 to each EDEVICE, 8000 through 8003. In order to have multipath and take advantage of hardware redundancy, a second path through a different chpid and switch needs to be added for each EDEVICE.

3.1.1 Initial single path configuration

The z/VM v5.2 installation process added the SCSI Definition Statements section to the file SYSTEM CONFIG that resides on the CF1 disk of user MAINT. These statements define the triple path to each EDEVICE for the system at IPL time. You can see an example of these statements in Example 3-1.

```
/**********************************************************************/
/*   SCSI Definition Statements                                       */
/**********************************************************************/

  edevice 8000 type fba attr SCSI fcp_dev 6002,
          wwpn 5005076300C300AA lun 5010000000000000
  edevice 8001 type fba attr SCSI fcp_dev 6003,
          wwpn 5005076300C300AA lun 5011000000000000
  edevice 8002 type fba attr SCSI fcp_dev 6006,
          wwpn 5005076300C300AA lun 5012000000000000
  edevice 8003 type fba attr SCSI fcp_dev 6007,
          wwpn 5005076300C300AA lun 5013000000000000
```

Example 3-1 SCSI definition statements

The EDVICE statement for 8000, which is LUN 501000000000, specifies the one available path, which is through fcp_dev 6002, wwpn 5005076300C300AC. We want to add a second path to each edevice.

A new path must consist of a different FCP_DEV and WWPN than are currently in use. The path is also through a different switch for hardware redundancy.

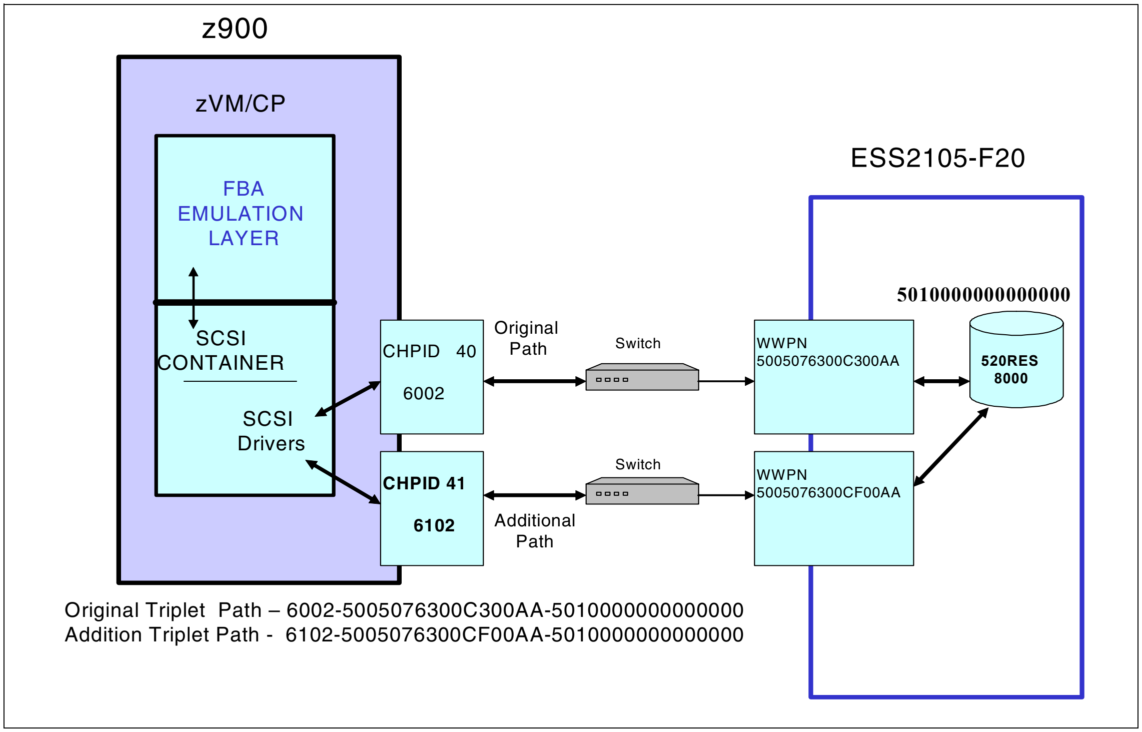
Having fcp_dev in another CHPID means that different FCP channel hardware will be used, and a different WWPN means that a different Host Bus Adapter (HBA) in the ESS2105 will be used. A different switch means that a completely different physical path from the Z900 processor to the ESS2105 will be used, which provides multipath support with hardware redundancy to the LUN.

Our existing definitions for edevices 8000 through 8003 use fcp_dev devices on CHPID 40. Another FCP CHPID that was an available path to the ESS2105 and through another switch was CHPID 41.

3.1.2 Adding additional paths

z/VM controls the distribution of I/O to an LUN when multiple paths are available, not the FCP channel hardware. When multiple paths are available to an LUN the z/VM SCSI Driver uses a round-robin scheme to distribute I/O to one path then the other, with the I/O completed on the path that it is started on.

An additional triplet/path was added for each of the EDEVICEs to enable a multipath configuration with hardware redundancy. Example 3-2 is what our configuration looked like prior to and after adding the additional path.



Example 3-2 Configuration architecture before adding additional path

The additional path to LUN 5010000000000000 through CHPID 41 consists of a FCP_DEV/WWPN/LUN triplet whose FCP_DEV and WWPN are different from the path through CHPID 40. This provides the multipath. A different switch was used to provide hardware redundancy, giving a completely different physical path for failover. Definitions were manually added to the SCSI Definition Statements section of the SYSTEM CONFIG file, as shown in Example 3-3.

```
/**********************************************************************/
/*   SCSI Definition Statements                                       */
/**********************************************************************/

  edevice 8000 type fba attr SCSI fcp_dev 6002,
          wwpn 5005076300C300AA lun 5010000000000000,
  fcp_dev 6102 wwpn 5005076300CF00AA lun 5010000000000000
  edevice 8001 type fba attr SCSI fcp_dev 6003,
          wwpn 5005076300C300AA lun 5011000000000000,
  fcp_dev 6103 wwpn 5005076300CF00AA lun 5011000000000000
  edevice 8002 type fba attr SCSI fcp_dev 6006,
          wwpn 5005076300C300AA lun 5012000000000000,
  fcp_dev 6106 wwpn 5005076300CF00AA lun 5012000000000000
  edevice 8003 type fba attr SCSI fcp_dev 6007,
          wwpn 5005076300C300AA lun 5013000000000000,
  fcp_dev 6107 wwpn 5005076300CF00AA lun 5013000000000000
```

Example 3-3 SCSI definition statements

With the new fcp_dev/wwpn statements added after an IPL we were able to see the result with a QUERY EDEVICE command, which shows the two available paths to EDEVICEs.

Querying the EDEVICE using the **q edev** command shows that there are now multiple paths from z/VM to each LUN/EDEVICE, as shown in Example 3-4.

```
q edev 8000-8003 details
EDEV 8000 TYPE FBA ATTRIBUTES SCSI
  PATHS:
    FCP_DEV: 6002 WWPN: 5005076300C300AA LUN: 5010000000000000
    FCP_DEV: 6102 WWPN: 5005076300CF00AA LUN: 5010000000000000
EDEV 8001 TYPE FBA ATTRIBUTES SCSI
  PATHS:
    FCP_DEV: 6003 WWPN: 5005076300C300AA LUN: 5011000000000000
    FCP_DEV: 6103 WWPN: 5005076300CF00AA LUN: 5011000000000000
EDEV 8002 TYPE FBA ATTRIBUTES SCSI
  PATHS:
    FCP_DEV: 6006 WWPN: 5005076300C300AA LUN: 5012000000000000
    FCP_DEV: 6106 WWPN: 5005076300CF00AA LUN: 5012000000000000
EDEV 8003 TYPE FBA ATTRIBUTES SCSI
  PATHS:
    FCP_DEV: 6007 WWPN: 5005076300C300AA LUN: 5013000000000000
    FCP_DEV: 6107 WWPN: 5005076300CF00AA LUN: 5013000000000000
Ready; T=0.01/0.01 14:37:52
```

Example 3-4 Querying the EDEVICE

4

FCP devices in the 2.6 Linux kernel

This chapter discusses accessing Fibre Channel Protocol devices from Linux guests. We cover:

- Addressing concepts for FCP-attached SCSI
- FCP addresses
- Linux device names
- The sysfs file system
- Linux FCP mapping
- SCSI IPL feature
- Multipathing concepts

4.1 Addressing concepts for FCP-attached SCSI

FCP on Linux for System z uses both SCSI and System z addressing styles and concepts. To make it work, there must be ways to map one type of address to another. As FCP address mapping may be new to experienced System z users, a short review of some of the concepts that are involved might be helpful.

4.1.1 SCSI addresses

Traditional SCSI devices have simple addresses. Older devices use a number in the range of 0–7, with address 7 usually reserved for the SCSI controller. These are known as *target* addresses. Newer SCSI architecture extends this range to 0–15 (or 0–F in hexadecimal), with address 7 usually reserved for the controller.

The target address of the SCSI device is usually set by a hardware function, such as a thumb switch or jumper pins. The user must ensure that two devices are not set to the same target address. More advanced implementations (often used with hot-pluggable drives) set the target address automatically using backplane hardware logic.

Each of the devices (targets) can have subsidiary units known as LUNs provided in the SCSI architecture. Older systems seldom used anything other than LUN0, and they use this by default. The system administrator often did not *see* the LUN. The older SCSI architecture had a maximum of seven LUNs (3 bits for addressing), while newer definitions provide for much larger numbers.

More advanced SCSI adapters provide several *bus* interfaces. Each bus can have a full compliment of target devices (and each target can have its own LUNs). Several SCSI adapters (each with several bus interfaces) can be used, although this is usually seen only in larger systems.

RAID adapters often map their logical drives to separate SCSI target addresses. Some RAID adapters have an option to map their logical drives to different LUN numbers at a single target address, but this setup is seldom encountered in typical small SCSI systems.

4.1.2 System z device numbers

In S/360™, an S/360 address had a specific meaning. For instance, the unit at address 327 was on channel 3, control unit 2, unit 7. This was later expanded to allow two hex digits for the channel address, so address B237 meant channel (or CHPID) number B2, control unit 3, device 7.

Later system evolution changed the address to a *device number* (still informally known as an address or software address). The device number is set by the

IOCDS and is an arbitrary 16-bit number expressed as four hexadecimal digits. It is no longer related to actual channel, control unit, or device addresses unless the person who is constructing the IOCDS happens to assign device numbers to match some elements of the hardware addresses involved.

4.1.3 System z hardware addresses

At the System z hardware level, a number of addresses are involved in using an I/O device. These include:

- One or more channel path identifiers (CHPIDs)
- An optional channel switch for which the appropriate port addresses are needed (the LINK addresses)
- One or more logical control unit addresses (CUADDs)
- A unit address (UNITADD)

The key addressing elements (CHPID, CUADD, UNITADD) are involved for all System z devices, although their exact meaning can vary depending on the device type. For example, the meaning of UNITADD is quite different for a 3390 disk and an OSA Express (LAN) adapter using QDIO interfaces. For QDIO devices, the IOCDS actually defines a queue.

4.2 FCP addresses

A new addressing scheme developed for FCP is built around worldwide names (WWN) that are eight bytes long:

- Part of the name represents an address type.
- Part is a number that identifies the manufacturer.
- Part is a unique number assigned by the manufacturer for each *port* or *node*.

A node is typically a box containing information and having one or more *ports*. A given box may have several addresses — one for the node itself and one for each FC port contained in the node. Common abbreviations are:

- WWNN is a worldwide node name.
- WWPN is a worldwide port name.
- WWN is any worldwide name (WWNN or WWPN).

A WWN is usually written in sets of two hex digits separated by colons (for example, 08:45:12:56:43:00:D5:A0).

In addition to the WWNN and WWPN names, addressable units (such as a disk drive, or logical drive on a RAID controller) are assigned a unit name that is also an 8-byte number. These assigned names are created by the FCP node that owns the device. The devices are known as logical units (LUNs) and the names are LUN names. For these network devices, it is important to remember that a *physical drive* can be a LUN. Think of a LUN as an addressable object within a node.

4.3 Linux device names

Linux works with device names instead of device addresses. For example, /dev/hda might be a disk drive. The exact scheme and name format varies between Linux platforms and is a function of the device drivers being used.

Basic Linux for System z uses names such as this:

/dev/dasda	The name of a whole drive (such as a 3390-3)
/dev/dasda1	The name of a partition on the dasda drive
/dev/sda	The name of a whole scsi fcp drive
/dev/sda1	This would be the name for the first partition

4.4 The sysfs file system

The Linux 2.6 kernel introduces the sysfs file system. The sysfs file system is simply described as a union of the proc, devfs, and devpty file systems. Sysfs enumerates the devices and busses that are attached to the system into a file system hierarchy that can be accessed from userspace. It is designed to handle the device and driver-specific options that have previously resided in /proc, and encompass the dynamic device addition previously offered by devfs. The /proc/scsi/zfcp/map interface used in the Linux 2.4 kernel is no longer available or required.

With the implementation of sysfs, you may find that many drivers and utilities still refer to the old proc entries. However, it is understood that sysfs is the agreed-upon direction that will be used in the future.

Figure 4-1 is a partial diagram of the 2.6 kernel sysfs directory structure.

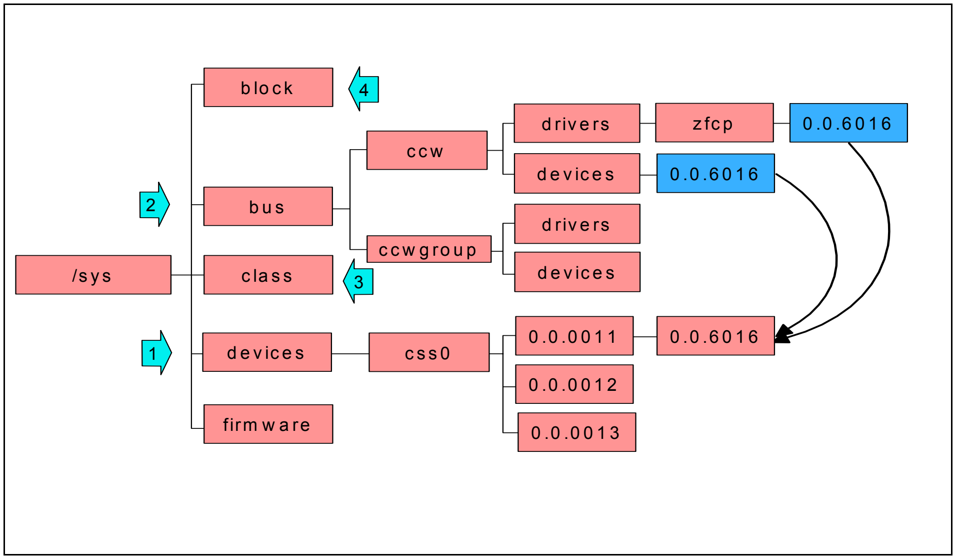


Figure 4-1 The sysfs directory structure

The sysfs file system is mounted at /sys and contains directories that organize the devices attached to the system in several different ways. The /sysfs subdirectories include:

1. The devices directory

   This directory contains the css0 directory. Its subdirectories represent all subchannels detected by the Linux kernel. Subchannel directories are named in the form 0.0.*nnnn*, where *nnnn* is the subchannel number in hex between 0 and ffff.

   Subchannel directories in turn contain status files and another subdirectory that represents the actual device. The device directory is named 0.0.*xxxx,* where *xxxx* is the unit address for the device. The device directory also contains status information as well as configuration options for the device.

2. The bus directory

   This contains a ccw subdirectory and a ccwgroup subdirectory. CCW devices are accessed using channel command words. Devices in the ccw directory only use one subchannel on the mainframe's channel subsystem. CCW group devices are also accessed with channel command words, but they use more than one subchannel per device. For example, a 3390-3 DASD device uses one subchannel. A QDIO network connection for an OSA adapter uses three subchannels.

The ccw directory and the ccwgroup directory both contain directories called devices and drivers:

- The devices directory contains symbolic links to the device directories in the /sys/devices/css0 directory.
- The drivers directory contains directories for each device driver that is currently loaded on the system. The zfcp driver has a directory here. The driver directory contains settings for the device driver, as well as symbolic links to the devices it is using (in /sys/devices/css0 directory).

3. The class directory

   This contains directories that group together similar devices such as ttys, SCSI tape drives, network devices, and other miscellaneous devices.

4. The block directory

   This contains directories for each of the block devices on the system. These are mostly the disk type devices such as real DASD, loopback devices, and software RAID block devices.

The difference that will be noticed between older Linux systems and ones that use sysfs will be the need to refer to devices by their sysfs name. On an SLES8 system, the zfcp driver was passed 0xb002 as its device addresses. On the 2.6 kernel systems the driver will get passed 0.0.b002. FCP configuration using sysfs.

> **Note:** For an introduction to FCP on Linux for System z and for details on FCP device mapping, see *Linux on zSeries: Fibre Channel Protocol Implementation Guide*, SG24-6344.

Dynamically adding and removing SCSI disks is simplified in the Linux 2.6 kernel, in contrast to the Linux 2.4 kernel, using the sysfs file system. The following examples illustrate the procedure.

> **Note:** The zfcp device driver must be loaded *before* FCP device entries appear in sysfs. To load the driver, issue the **modprobe** command:
>
> ```
> modprobe zfcp
> ```
>
> Entries in the /sys directory allow us to configure a SCSI device. In these examples, we assume that the current directory is /sys/bus/ccw/drivers/zfcp.

1. Set the virtual FCP adapter online. To bring FCP device 6116 online:

```
# echo 1 > 0.0.6116/online
```

To bring the device offline:

```
# echo 0 > 0.0.6116/online
```

As an alternative, use the **chccwdev** command:

```
# chccwdev -e 0.0.6116
Setting device 0.0.6116 online
Done
# chccwdev -d 0.0.6116
Setting device 0.0.6116 offline
Done
```

2. Add a WWPN. To add WWPN 0x5005076300cb00aa to the FCP device:

   ```
   # echo 0x5005076300cb00aa > 0.0.6116/port_add
   ```

   The WWPN appears as a subdirectory for the device (the 0.0.6116/0x5005076300cb00aa subdirectory).

3. Add a Logical Unit (LUN) to the WWPN. To add LUN 0x5024000000000000 to the WWPN:

   ```
   # echo 0x5024000000000000 > 0.0.6116/0x5005076300cb00aa/unit_add
   ```

   The LUN is now available to Linux. As this is the first SCSI disk, it maps to the sda device. This can be checked using:

   ```
   # cat /sys/block/sda/device/fcp_lun
   0x5024000000000000
   ```

   When added, the LUN is mapped to the SCSI subsystem in sysfs. A subdirectory of the form *host*:*bus*:*target*:*lun* is added to the /sys/bus/scsi/devices directory:

   ```
   # cat /sys/bus/scsi/devices/1:0:0:0/fcp_lun
   0x5024000000000000
   ```

4.4.1 Querying FCP device information

Information for an FCP device is available in sysfs. For example, attributes for the device 6016 can be read from files found in the /sys/bus/ccw/drivers/zfcp/0.0.6016 directory.

> **Note:** A complete list of all attributes for an FCP device can be found in *Linux on System z, Device Drivers, Features, and Commands February, 2007*, SC33-8289.

Both RHEL5 and SLES10 still maintain the /proc/scsi/scsi file interface (Figure 4-2).

```
[root@jns3 ~]# cat /proc/scsi/scsi
Attached devices:
Host: scsi0 Channel: 00 Id: 00 Lun: 00
  Vendor: IBM      Model: 2105F20          Rev: .144
  Type:   Direct-Access                    ANSI SCSI revision: 03
Host: scsi1 Channel: 00 Id: 00 Lun: 00
  Vendor: IBM      Model: 2105F20          Rev: .144
  Type:   Direct-Access                    ANSI SCSI revision: 03
Host: scsi2 Channel: 00 Id: 00 Lun: 00
  Vendor: IBM      Model: 2105F20          Rev: .144
  Type:   Direct-Access                    ANSI SCSI revision: 03
```

Figure 4-2 File interface in RHEL5 and SLES10

SLES10 and RHEL5 both provide additional commands to query SCSI devices, such as the **lsscsi** command, which uses sysfs to report information about SCSI devices (Figure 4-3).

```
[root@jns3 /]# lsscsi
[0:0:0:0]    disk    IBM      2105F20          .144  /dev/sda
[1:0:0:0]    disk    IBM      2105F20          .144  /dev/sdb
[2:0:0:0]    disk    IBM      2105F20          .144  /dev/sdc
```

Figure 4-3 Using the lsscsi command

Figure 4-3 shows the path for a SCSI device, information about the type of device (that is, disk and IBM 2105F20), and the Linux block special name currently assigned to the device. The path column, which is the information within the brackets [0:0:0:0] in column one of Figure 4-3, consists of four fields ([a:b:c:d]) with the following meaning:

a. SCSI_host
b. Channel
c. Target_id
d. LUN

- Fields a and b

  In the Linux for System z context, the channel is always 0, and taken together with the SCSI_host field represents the FCP subchannel that the device is associated with. Seeing the values 0:0 for fields a and b indicate in our case a LUN on subchannel 0.0.6016. In our example above the value of 1:0:0:0 would indicate the first LUN on the first WWPN of subchannel 0.0.6017.

- Field c

  The Target_id represents a particular target WWPN used to access the LUN. Thus, a value of zero represents the first target WWPN defined (in our case, 0x5005076300c300aa), and a value of 1 represents the second target WWPN defined (again, in our case, 0x5005076300c700aa).

- Field d

  The LUN number is just a sequential number starting at 0 that represents the first, second, third, and so on, LUN that has been defined on a particular FCP subchannel, with a particular target WWPN. So [0:0:0:0] represents the first LUN defined on FCP subchannel 0.0.6016 with target WWPN 0x5005076300c300aa. Similarly, [0:0:0:1] represents the second LUN defined on FCP subchannel 0.0.6016 with target WWPN 0x5005076300c300aa, and [0:0:1:0] represents the first LUN defined on FCP subchannel 0.0.6016 with target WWPN of 0x5005076300c700aa.

It is possible to verify which target WWPN and device LUN number are associated with any given row in the lsscsi output from the /sys file system. In the case of the first row [0:0:0:0] you would enter the following commands:

Enter : `cd /sys/class/scsi_device/0:0:0:0/device`

Enter : `cat fcp_lun` (to see the lun id, for example, 0x5020000000000000)

Enter : `cat wwpn` (to see the target WWPN, for example, 0x5005076300c300aa)

Enter : `cat hba_id` (to see the FCP subchannel, for example, 0.0.6016)

The **systool** command uses sysfs to report devices by bus, class, and topology. The -v option reports all device attributes. The -b option reports device information for a specific bus (Figure 4-4).

```
[root@jns3 ~]# systool -vb scsi
Bus = "scsi"

  Device = "0:0:0:0"
  Device path =
"/sys/devices/css0/0.0.0011/0.0.6016/host0/rport-0:0-0/target0:0:0/0:0:0:
0"
    delete              = <store method only>
    device_blocked      = "0"
    fcp_lun             = "0x5023000000000000"
    hba_id              = "0.0.6016"
    iocounterbits       = "32"
    iodone_cnt          = "0xff9"
    ioerr_cnt           = "0x0"
    iorequest_cnt       = "0xff9"
    model               = "2105F20         "
    queue_depth         = "32"
    queue_type          = "simple"
    rescan              = <store method only>
    rev                 = ".144"
    scsi_level          = "4"
    state               = "running"
    timeout             = "60"
    type                = "0"
    uevent              = <store method only>
    vendor              = "IBM     "
    wwpn                = "0x5005076300c300aa"
```

Figure 4-4   Using the systool command

The **udevinfo** command reports device information from udev (Figure 4-5).

```
[root@jns3 ~]# udevinfo -q all -n /dev/sda1
P: /block/sda/sda1
N: sda1
S: disk/by-id/scsi-1IBM_2105_02312320-part1
S:
disk/by-path/ccw-0.0.6016-zfcp-0x5005076300c300aa:0x5023000000000000-part1
S: disk/by-uuid/0799f3b8-682e-4cf1-be99-0910bf773dfb
S: disk/by-label/boot
E: ID_VENDOR=IBM
E: ID_MODEL=2105F20
E: ID_REVISION=.144
E: ID_SERIAL=1IBM_2105_02312320
E: ID_TYPE=disk
E: ID_BUS=scsi
E: ID_PATH=ccw-0.0.6016-zfcp-0x5005076300c300aa:0x5023000000000000
E: ID_FS_USAGE=filesystem
E: ID_FS_TYPE=ext3
E: ID_FS_VERSION=16777216.0
E: ID_FS_UUID=0799f3b8-682e-4cf1-be99-0910bf773dfb
E: ID_FS_LABEL=/boot
E: ID_FS_LABEL_SAFE=boot
```

Figure 4-5 Using the udevinfo command

The -q all option specifies that all sysfs device attributes are to be reported. The -n /dev/sda1 option displays disk and partition information for the /dev/sda1 device.

4.5 Linux FCP mapping

With FCP mapping, some of the mapping elements are assigned on the System z side, and others are created within the FC fabric. LUN addresses appear on both the System z side and the FCP side. Each accessed LUN requires a map entry. (The format is shown in Figure 4-6.)

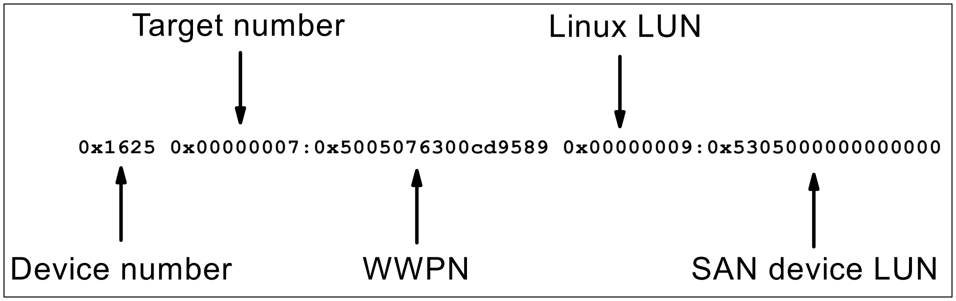

Target number
Linux LUN
0x1625 0x00000007:0x5005076300cd9589 0x00000009:0x5305000000000000
Device number
WWPN
SAN device LUN

Figure 4-6 The format of an FCP map entry

The five elements in a map entry are:

- Device number (explained in the next section)
- Target number
- WWPN
- Linux LUN
- SAN device LUN

4.5.1 Device number

The first number is a System z device number. This is defined in the IOCP/IOCDS, and it is assigned to the FCP channel. The same device number may used for all FCP connections, or multiple device numbers (assigned to the FCP channel) may be used.

The following rules apply to device numbers:

- In the simple case, a *device number* corresponds to a *unit* on the channel. In more complex cases, we consider both *device numbers* and *units*.

  Note: The *unit* on an FCP channel is an abstraction that does not correspond to any network entity (such as a LUN). Rather, it is a communication path (QDIO queue pair) between the FCP channel and the operating system.

- The maximum number of FCP channels depends on the specific System z configuration. There are two or four channels per adapter card depending on the feature code. Each may be used as a FICON channel or an FCP channel. The maximum number of devices for a single FCP channel is:
  - 240 for a z800 or z900
  - 480 for a z890, z990, z9
- The same device number can be used for all FCP maps in a given Linux system. In the unlikely event that you exceed the maximum number of target addresses and LUN addresses possible on a single unit address, you might need additional device numbers.

  If an error occurs with a SCSI device, operation of *all devices connected to the device number* may be stopped while recovery takes place. Consider using different device numbers for unreliable devices.

  Note: We recommend separate device numbers for each tape device.

- Each z/VM Linux guest must use a different device number to access FCP devices. This imposes a limit to the number of Linux guests that can access one FCP channel:
  - 240 on a z800 or z900
  - 480 on a z890, z990, z9
- The IOCDS can define the same device numbers for multiple LPARs. However, the System z channel firmware internally creates separate units for each LPAR. The total number of devices cannot exceed the maximum for your configuration. For example, an IOCDS that defines 64 device numbers and is shared by two LPARs uses (2 x 64 =) 128 units on the FCP channel. This applies to the maximum of number Linux images that can use a single FCP channel (whether they run in an LPAR, under z/VM LPARs, or any combination).
- Device numbers are four-digit hexadecimal numbers (following normal System z convention). High-order zeros may be omitted.
- Unit addresses in the IOCDS definition for device numbers must *not* include FC, FD, FE, and FF. These are reserved for future functions.

4.5.2 Target number

The second number of the map specifies the SCSI target number Linux is to use. This is assigned by you.

The basic rules are:

- Target address 0 cannot be used. (This addresses the FCP adapter itself.) Usable target numbers range from 1 to any positive 32-bit number.
- Normal usage starts with target address 1 and increments by 1 for each new WWPN used. However, any positive 32-bit number can be used for a target number, and the target number need not be sequential.
- A different target address must be used for each different WWPN in the FCP map. However, target addresses may be reused with different device numbers.
- All LUNs on a single WWPN address (using the same device number) must use the same target number.

4.5.3 WWPN

The third map number is the worldwide port name (WWPN) of the device containing the LUN. This is the WWPN as seen by the FC switches. In practice, this number is always written in hexadecimal and must be 16 hexadecimal digits.

The WWPN is built into the FCP device. To determine the proper number, query the FCP node or switch.

4.5.4 Linux LUN

The fourth number is the LUN to be used by Linux. You assign this number.

4.5.5 SAN device LUN

The fifth number is the LUN used by the remote device. This is the LUN address that is assigned by the node controller. In practice, the number is always written in hexadecimal and must be exactly 16 hexadecimal digits.

> **Note:** All of the devices we worked with use only the high-order four digits of the 16 hexadecimal digit LUN address. Some devices check only the first four hexadecimal digits. Therefore, different LUN addresses having the same high-order four hexadecimal digits access the same device. This can cause errors or erratic operation. Thus, we recommend that the low-order 12 hexadecimal digits be zero.

It is possible for two LUNs on two different storage controllers to use the same LUN (because they are connected via different WWPNs).

4.5.6 Map entry ordering

Map entries can be in any order, but we recommend ordering within a particular WWPN by:

1. Device number
2. Target number
3. Linux LUN number

This makes it simpler to manage FCP mapping. In particular, this enables you to spot gaps that can affect how LUNs are automatically attached to the SCSI system.

4.5.7 Syntax rules

Some basic syntax rules for FCP maps include:

- Numbers in the FCP maps are C language decimal by default. Use a leading *0x* prefix for hexadecimal.

> **Note:** The *C language* condition means that a leading zero (with no 0x prefix) indicates an octal number. For example, 010 is interpreted as an octal number equaling decimal 8. To avoid confusion, we recommend using hexadecimal notation.

- Comments may appear in maps that exist as separate lines in a file. They are indicated by the number character (#) at the beginning of the comment. The comment can be a whole line or the right-hand portion of a line.

> **Note:** Be careful using the # comment character in a script. The map will be generated successfully but entries following the # character are ignored.

- When supplied as a module parameter, entries are separated with semicolons. The backslash (\) escape character should be used as a continuation character. Some command formats require the whole map to be enclosed in double quotation marks:

```
# insmod zfcp map=”\
0x0600 0x1:0x1234567887654321 0x0:0x0000000000000000;\
0x0600 0x1:0x1234567887654321 0X1:0x0001000000000000”
```

4.6 SCSI IPL feature

The traditional initial program load (IPL) process relies on accessing a device using System z channel attachment. For IPL from a FCP-attached device, this is not possible. In this case, the IPL record resides on an SCSI disk and the device cannot be accessed from the information defined in the IOCDS (channel path, control unit, device number). A new tool (referred to as the *machine loader*) is required to perform an IPL from FCP-attached devices.

To help understand SCSI features for System z, we discuss:

- “IPL Linux from SCSI disk in an LPAR” on page 46

  In this section, we introduce operation of the machine loader.
- “IPL Linux from SCSI disk in a VM guest” on page 46

  We examine how the machine loader operates for Linux IPL in a VM guest.
- “The CP SET LOADDEV command” on page 47

  A new CP command is available to support Linux IPL from SCSI.

- "The LOADDEV user directory statement" on page 47

  This user directory statement can automate Linux IPL from SCSI.

- "For details on the LOADDEV user directory statement, consult CP Planning and Administration, SC24-6043." on page 48

  If running Linux in an LPAR, you can use the dump tools with SCSI disks.

4.6.1 IPL Linux from SCSI disk in an LPAR

When a Linux IPL from a SCSI disk is initiated from the Service Element (SE) or Hardware Management Console (HMC), the machine loader is copied from the service element into LPAR memory. Parameters that are specified on the load panel (the WWPN and LUN number of the disk to access) are also loaded into memory.

Note: The machine loader uses these parameters to build the SCSI commands that are required to access the SCSI disk. The load panel provides parameters to the machine loader.

At this point:

1. The operating system loader (OS loader) is then read from the SCSI disk into LPAR memory. The machine loader must move itself into LPAR memory to avoid being overwritten.
2. When the OS loader is in LPAR memory, the machine loader is no longer required and can be overwritten. At this point, the OS loader loads the operating system and IPL proceeds.

For details on the operation of the System z IPL from SCSI feature, consult "SCSI initial program loading for zSeries" by G. Banzhaf, et al., *IBM Journal of Research and Development,* Vol 48, No. 3/4, 2004.

4.6.2 IPL Linux from SCSI disk in a VM guest

When a Linux IPL from a SCSI disk is initiated from a VM guest, the machine loader is read from the hardware system area (HSA) into the guest's virtual memory.

Note: If the SCSI IPL feature is installed and enabled, the machine loader is loaded into the HSA during initial machine load (IML).

The parameters that are required to access the SCSI disk are also read into the guest's virtual memory. IPL then proceeds as described in 4.6.1, "IPL Linux from SCSI disk in an LPAR" on page 46.

4.6.3 The CP SET LOADDEV command

The new CP SET LOADDEV command is used to provide the parameters that are needed to access a SCSI disk to the machine loader. Parameters to SET LOADDEV include the WWPN and LUN number of the SCSI disk on which Linux resides. For example, to IPL Linux on the SCSI disk that is located at WWPN 5005076300C300AA and LUN 5023000000000000, use the command:

```
SET LOADDEV PORTNAME 50050763 00C300AA LUN 50230000 00000000
```

The CP QUERY LOADDEV command reports parameters set for the machine loader:

```
CP Q LOADDEV
PORTNAME 50050763 00C300AA    LUN  50230000 00000000    BOOTPROG 0
BR_LBA   00000000 00000000
```

For details on the SET LOADDEV and QUERY LOADDEV commands, consult *CP Command and Utility Reference*, SC24-6008.

4.6.4 The LOADDEV user directory statement

It is possible to set the machine loader parameters in the user directory entry using the LOADDEV statement. In Figure 4-7, we set the LOADDEV WWPN and LUN parameters.

```
USER JNS3 JNS3   1024M 1024M  BG
   MACHINE ESA
   CPU 00
   CPU 01
   IPL CMS
   LOADDEV PORT 5005076300c300aa
   LOADDEV LUN 5023000000000000
   CONSOLE 001F 3215 T
   SPOOL 000C 2540 READER *
   SPOOL 000D 2540 PUNCH A
   SPOOL 000E 1403 A
   .......
   .......
```

Figure 4-7 Setting LOADDEV parameters

For details on the LOADDEV user directory statement, consult *CP Planning and Administration*, SC24-6043.

4.6.5 Dumping to a SCSI disk

If Linux is running in an LPAR, the machine loader is used to load the SCSI dump tool from the SCSI disk.

Note: Dump to SCSI is not supported for Linux running under z/VM. In this case, dumps must be performed to ECKD DASD or to tape.

When performing a dump to a SCSI disk:

- A dump is initiated from the load panel on the SE or HMC. Portions of LPAR memory are copied to the HSA in order to free space for the machine loader.
- The machine loader (including the SCSI access parameters) is loaded into LPAR memory. The machine loader builds the SCSI commands to access the SCSI disk.
- The SCSI dump tool is loaded from SCSI disk to LPAR memory. It copies data previously saved in the HSA to the dump device, then dumps the remaining LPAR memory to SCSI disk.

4.7 Multipathing concepts

Multipathing is supported for FCP-attached SCSI disks. In this section, we compare multipathing with FICON to multipathing with FCP.

In general, there are two reasons for establishing a multiple paths to a device:

- *High availability* provides several physical paths to a device and offers high availability in a failure scenario. If one path fails, other paths to the device are still available.
- *Performance* is achieved by using multiple paths simultaneously to write to or read from a device.

In this section, we compare multipathing with FICON to multipathing with FCP.

4.7.1 Multipathing with FICON

With the standard channel subsystem on System z machines, multipath access to disk subsystems is a basic hardware feature. Both ESCON® and FICON connections support multiple hardware paths to any physical disk device. The z/Series microcode handles multipathing invisibly to the operating system. In the ESCON and FICON models, a single device is presented to the operating system to do I/O operations on. Multipathing happens automatically and is handled by the I/O subsystem. Figure 4-8 illustrates FICON multipathing.

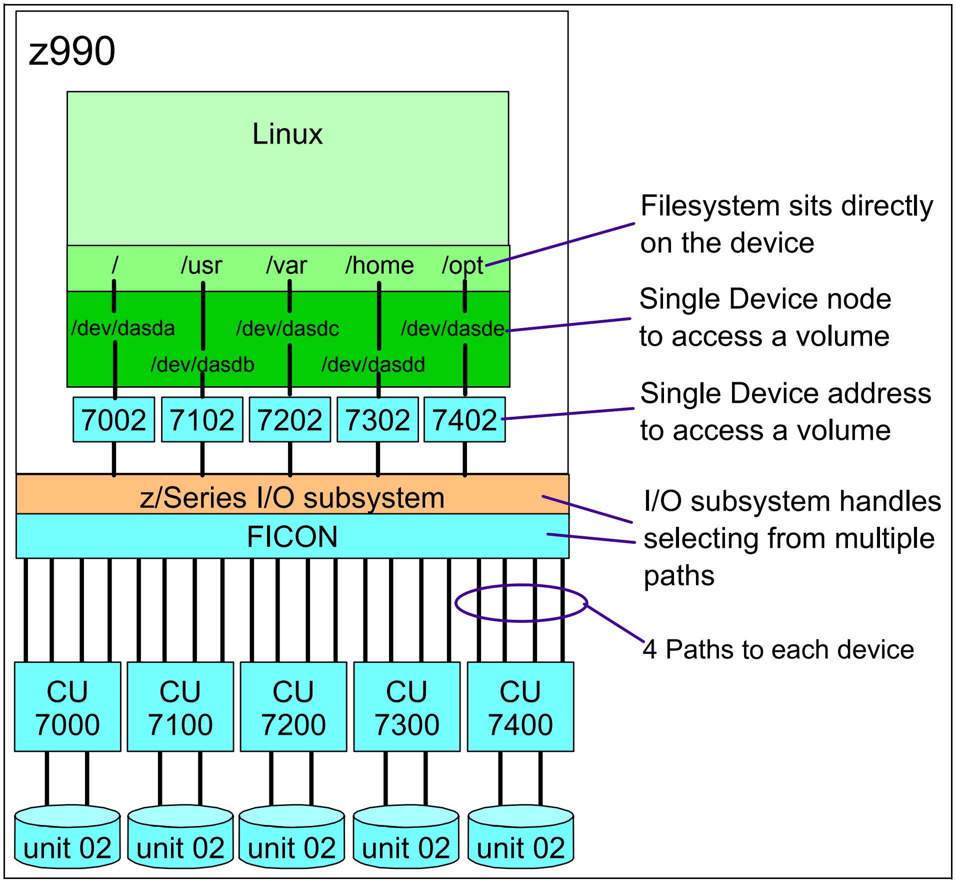


Figure 4-8 FICON multipathing

The complexity of choosing from the multiple FICON paths is handled by the z/Series I/O subsystem and is hidden from the Linux OS image. Multipathing is managed across the entire z/Series system.

4.7.2 Multipathing with FCP

Multipathing over FCP is another matter. With FCP multipathing on Linux on System z, each path to each LUN appears to the operating system as a separate device. For example, if there are four paths to five LUNs, the Linux system sees 20 SCSI devices. This means that there must be another layer of code between the Linux file system layer and the SCSI subsystem. This extra layer handles all of the coordination between the raw paths and the higher level file system:

- On Red Hat, this layer is handled by mdadm.
- On a SUSE this layer is handled by EVMS, LVM2, and MPTOOLS.

Figure 4-9 illustrates multipathing with FCP.

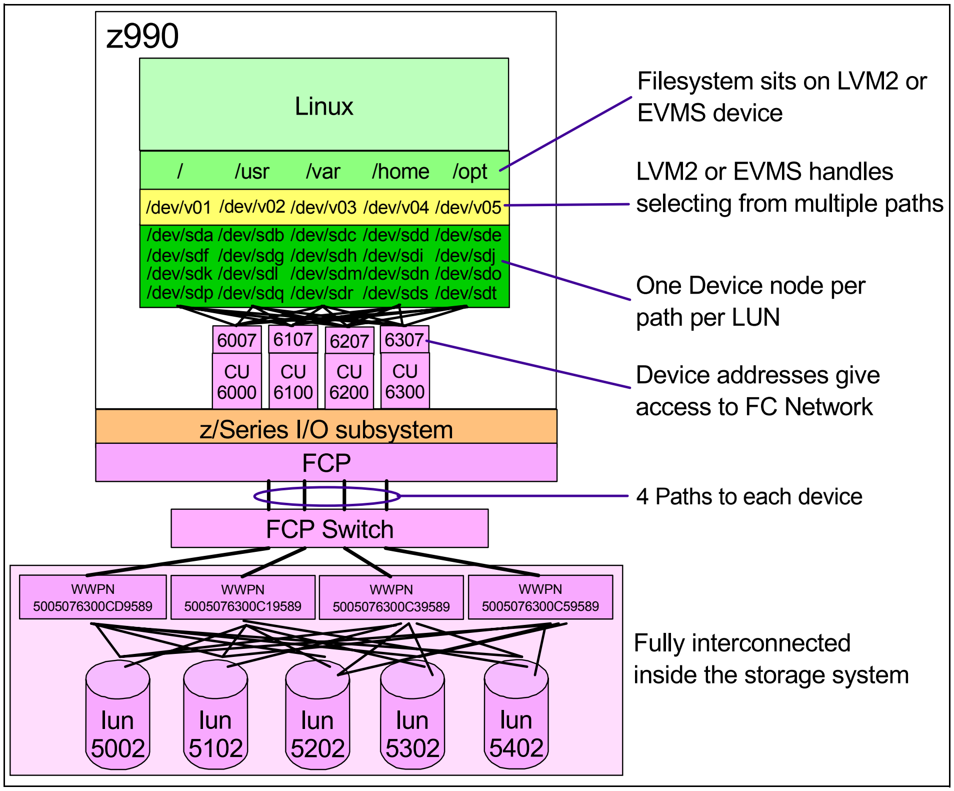

z990
Linux
/ /usr /var /home /opt
/dev/v01 /dev/v02 /dev/v03 /dev/v04 /dev/v05
/dev/sda /dev/sdb /dev/sdc /dev/sdd /dev/sde
/dev/sdf /dev/sdg /dev/sdh /dev/sdi /dev/sdj
/dev/sdk /dev/sdl /dev/sdm/dev/sdn /dev/sdo
/dev/sdp /dev/sdq /dev/sdr /dev/sds /dev/sdt
6007 6107 6207 6307
CU 6000
CU 6100
CU 6200
CU 6300
z/Series I/O subsystem
FCP
FCP Switch
WWPN 5005076300CD9589
WWPN 5005076300C19589
WWPN 5005076300C39589
WWPN 5005076300C59589
lun 5002
lun 5102
lun 5202
lun 5302
lun 5402
Filesystem sits on LVM2 or EVMS device
LVM2 or EVMS handles selecting from multiple paths
One Device node per path per LUN
Device addresses give access to FC Network
4 Paths to each device
Fully interconnected inside the storage system

Figure 4-9 FCP multipathing

In addition to getting FCP itself running on Linux, multipathing also requires getting this extra layer of virtualization configured and running. In order to get the most benefit of multipathing, multiple FCP CHPIDs and one storage system host adapter per CHPID should be used. FCP multipathing is managed at the Linux system level. There is no global multipathing scheduler that works across the entire z/Series system for FCP.

5

Installing SUSE Linux Enterprise Server 10 on FCP-attached SCSI Devices

This chapter provides the details for installing SLES10 onto FCP devices, as well as how to add FCP devices to an existing SLES10 system. SLES10 was chosen because it is the most current release of SuSE Linux for System z. Although there may be slight differences in the GUI installation panels for SLES9, both releases are based on the Linux 2.6 kernel and the process is the same. Topics include:

- Installing SLES10 to an FCP device
- Adding FCP devices to an existing SLES10 system

5.1 Installing SLES 10 to an FCP device

The SLES10 installation system supports installing Linux, including the boot partition, onto FCP devices. The following window shots depict the GUI versions of the YaST installation panels. The ncurses text mode version of the YaST installation panels will display or request the same information.

This document assumes that the reader is familiar with the process of installing Linux on System z as a guest of z/VM, and skips some of the basic initial installation steps. The focus is on the specific tasks required to install onto FCP devices.

1. After the Linux for System z installation RAM system is up and running, start a VNC client on a workstation and establish a session with the SLES 10 Linux guest. YaST displays the window shown in Figure 5-1.

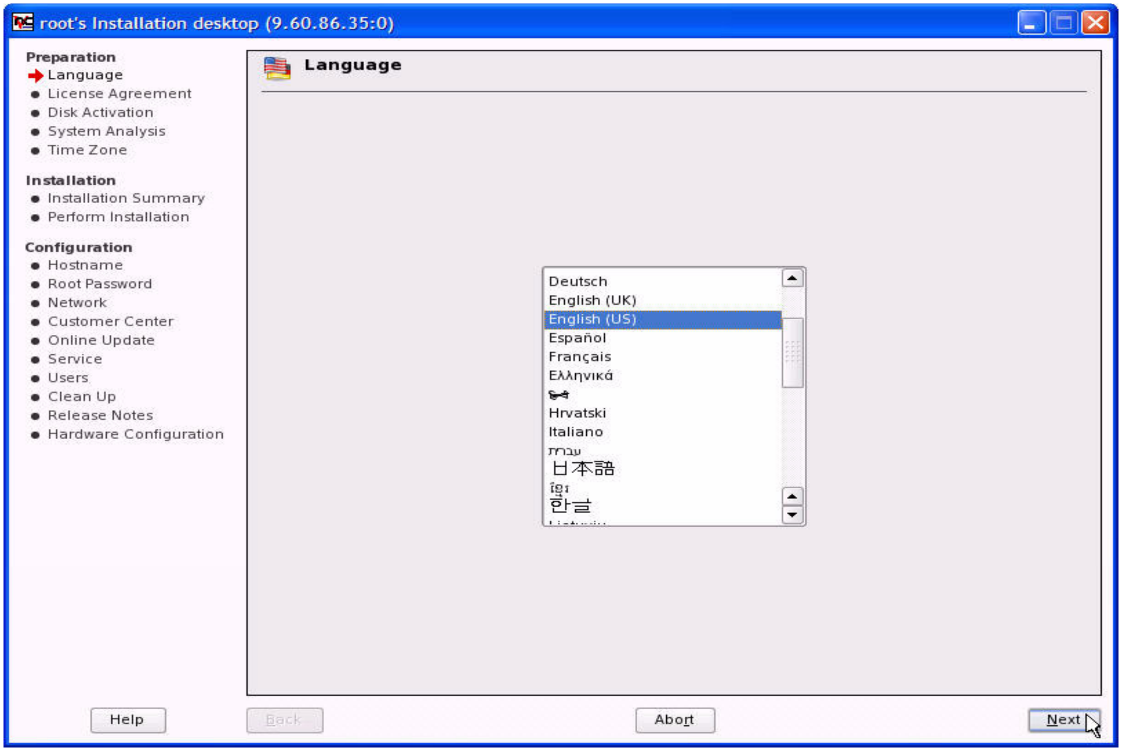
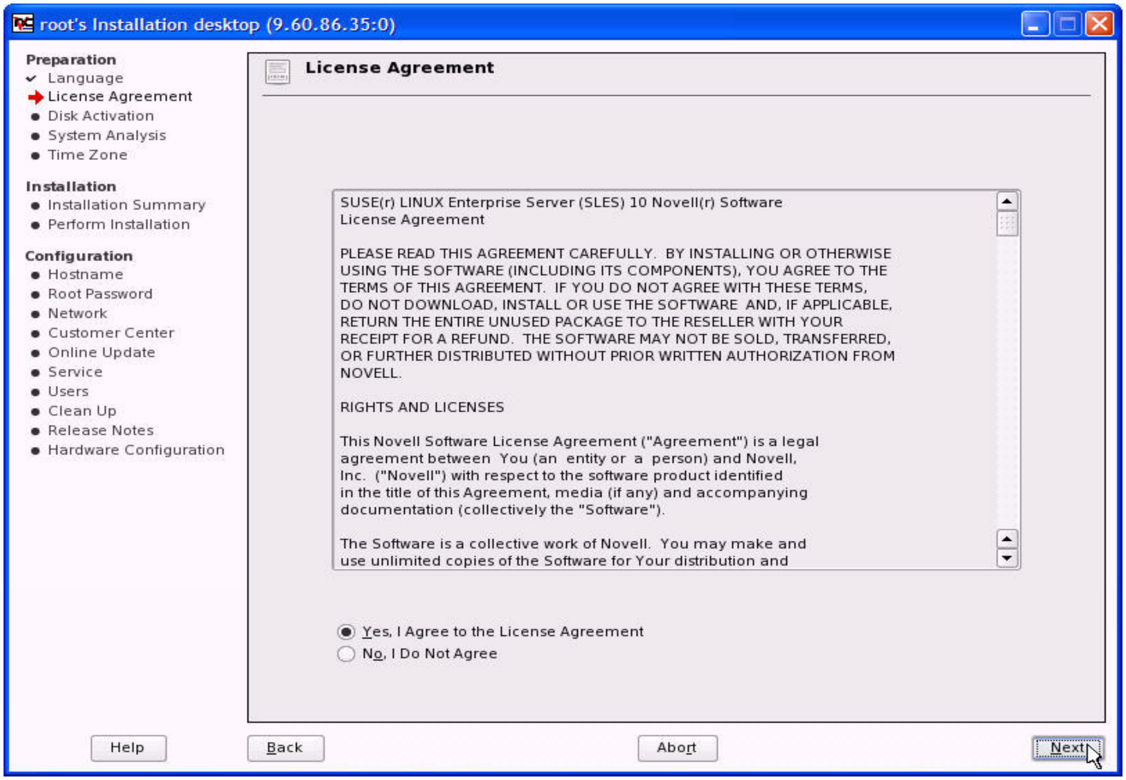


Figure 5-1 Language selection panel

2. Select the language that will be used for the system, then click **Next** to continue to Figure 5-2.



SUSE(r) LINUX Enterprise Server (SLES) 10 Novell(r) Software License Agreement

PLEASE READ THIS AGREEMENT CAREFULLY. BY INSTALLING OR OTHERWISE USING THE SOFTWARE (INCLUDING ITS COMPONENTS), YOU AGREE TO THE TERMS OF THIS AGREEMENT. IF YOU DO NOT AGREE WITH THESE TERMS, DO NOT DOWNLOAD, INSTALL OR USE THE SOFTWARE AND, IF APPLICABLE, RETURN THE ENTIRE UNUSED PACKAGE TO THE RESELLER WITH YOUR RECEIPT FOR A REFUND. THE SOFTWARE MAY NOT BE SOLD, TRANSFERRED, OR FURTHER DISTRIBUTED WITHOUT PRIOR WRITTEN AUTHORIZATION FROM NOVELL.

RIGHTS AND LICENSES

This Novell Software License Agreement ("Agreement") is a legal agreement between You (an entity or a person) and Novell, Inc. ("Novell") with respect to the software product identified in the title of this Agreement, media (if any) and accompanying documentation (collectively the "Software").

The Software is a collective work of Novell. You may make and use unlimited copies of the Software for Your distribution and

● Yes, I Agree to the License Agreement
○ No, I Do Not Agree

Figure 5-2 Licence Agreement panel

3. Select the **Yes, I Agree to the License Agreement** radio button, then click **Next** to continue to Figure 5-3.

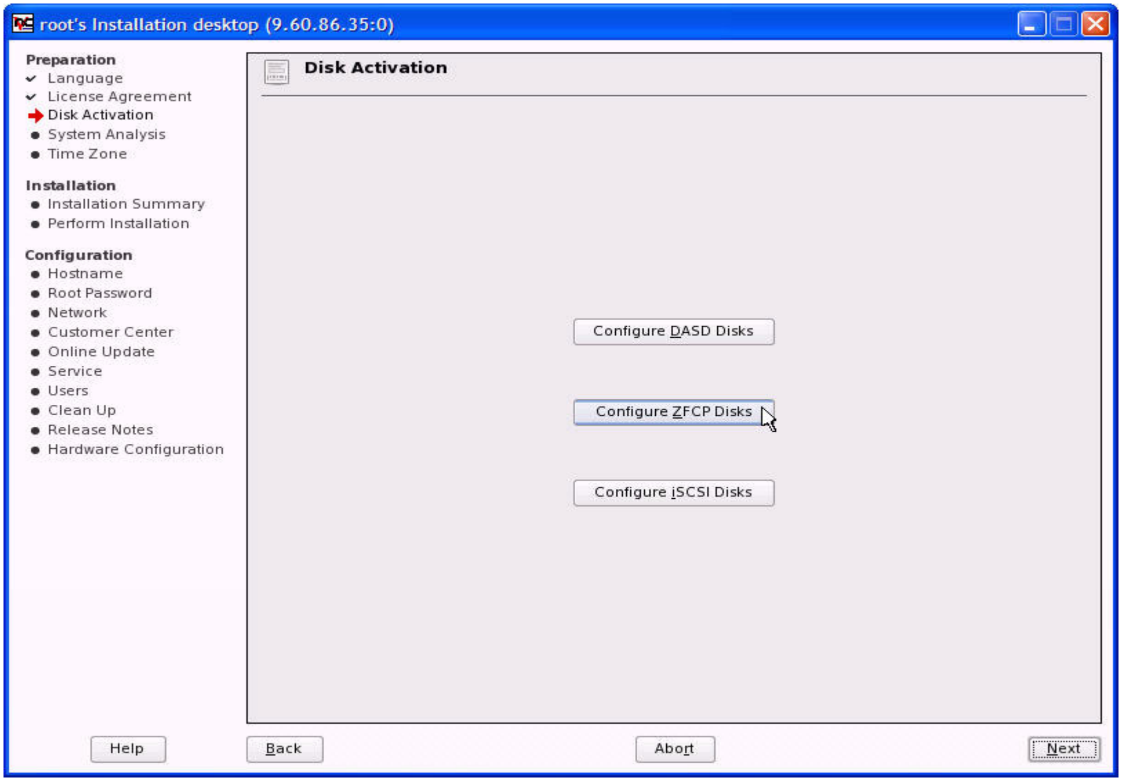
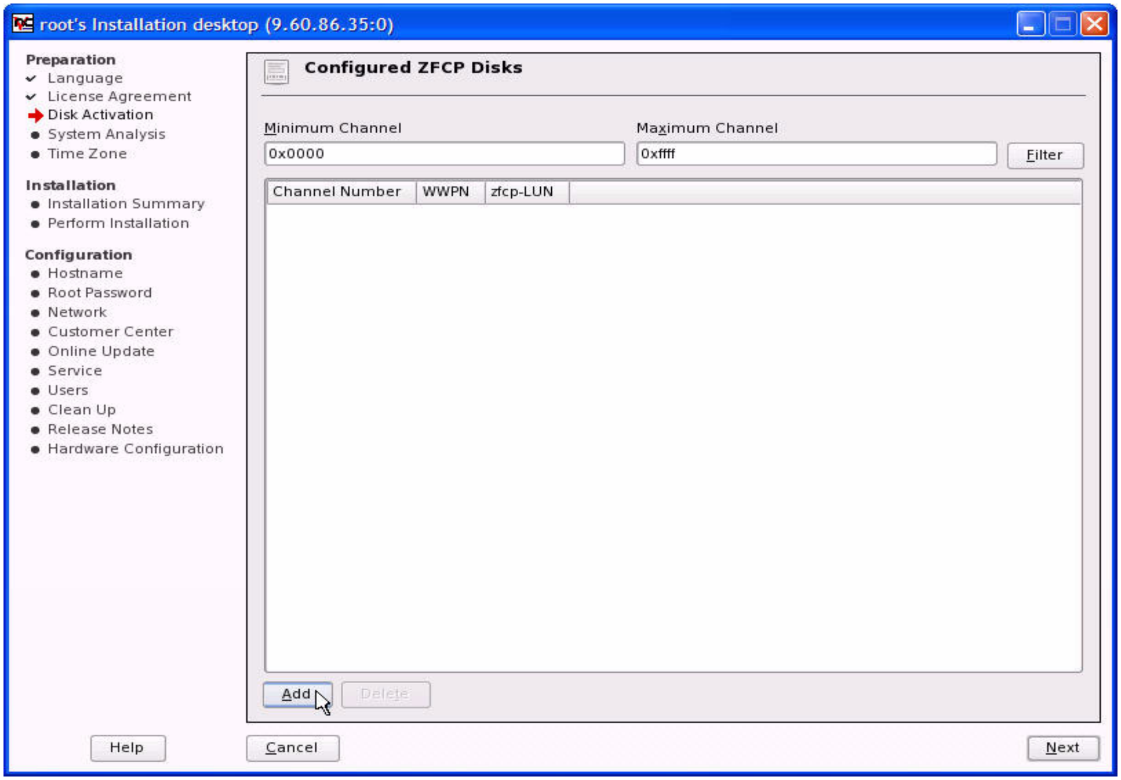


Figure 5-3 Disk Activation panel

4. Click **Configure ZFCP Disks** to continue to the window shown in Figure 5-4.



Configured ZFCP Disks

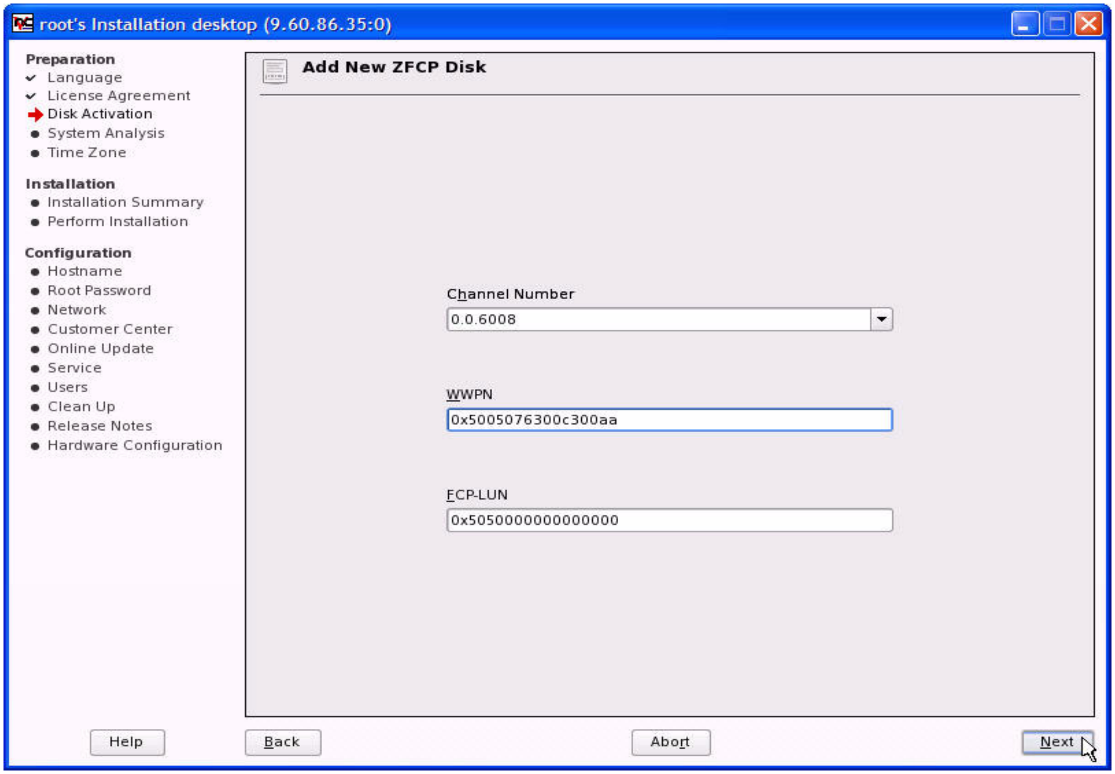
Minimum Channel: 0x0000

Maximum Channel: 0xffff

Channel Number	WWPN	zfcp-LUN

Figure 5-4 Configured ZFCP Disks panel

5. This window shows that there are currently no zfcp disks configured. Click **Add** to display the Add New ZFCP Disk panel (Figure 5-5).



root's Installation desktop (9.60.86.35:0)

Add New ZFCP Disk

Channel Number: 0.0.6008

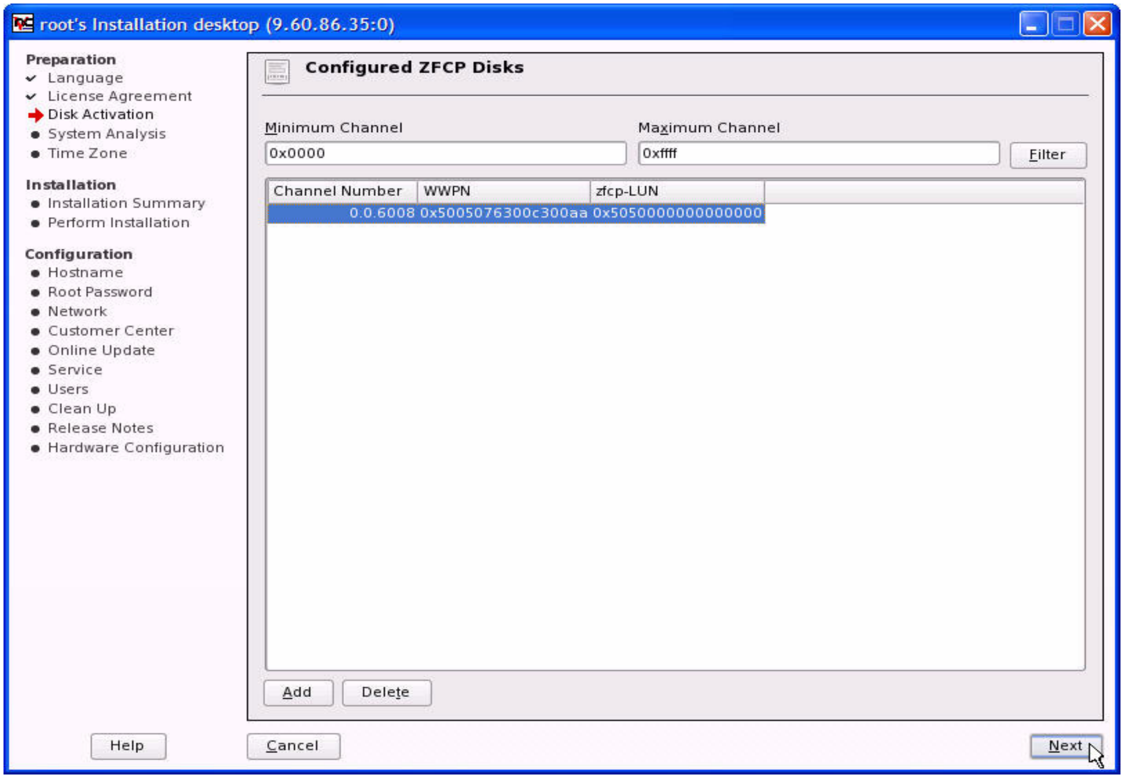
WWPN: 0x5005076300c300aa

FCP-LUN: 0x5050000000000000

Figure 5-5 Add New ZFCP Disk panel

6. Select the FCP device from the Channel Number drop-down list. Enter the WWPN of the storage device that holds the LUN into the WWPN field. Enter the LUN in the FCP-LUN field. Click **Next** to continue to the panel shown in Figure 5-6.

> **Caution:** The WWPN and FCP-LUN fields must be entered in lowercase letters as a 16-digit hex value with all trailing zeros and begin with the character string prefix 0x. Leaving this prefix off the WWPN field resulted in the error message Invalid WWPN pop-up window when **Next** was clicked to continue.



Minimum Channel: 0x0000

Maximum Channel: 0xffff

Channel Number	WWPN	zfcp-LUN
0.0.6008	0x5005076300c300aa	0x5050000000000000

Figure 5-6 Configured ZFCP Disks panel

7. When added, the disk shows in the list of configured zfcp disks. Click **Next** to continue to the panel shown in Figure 5-7.

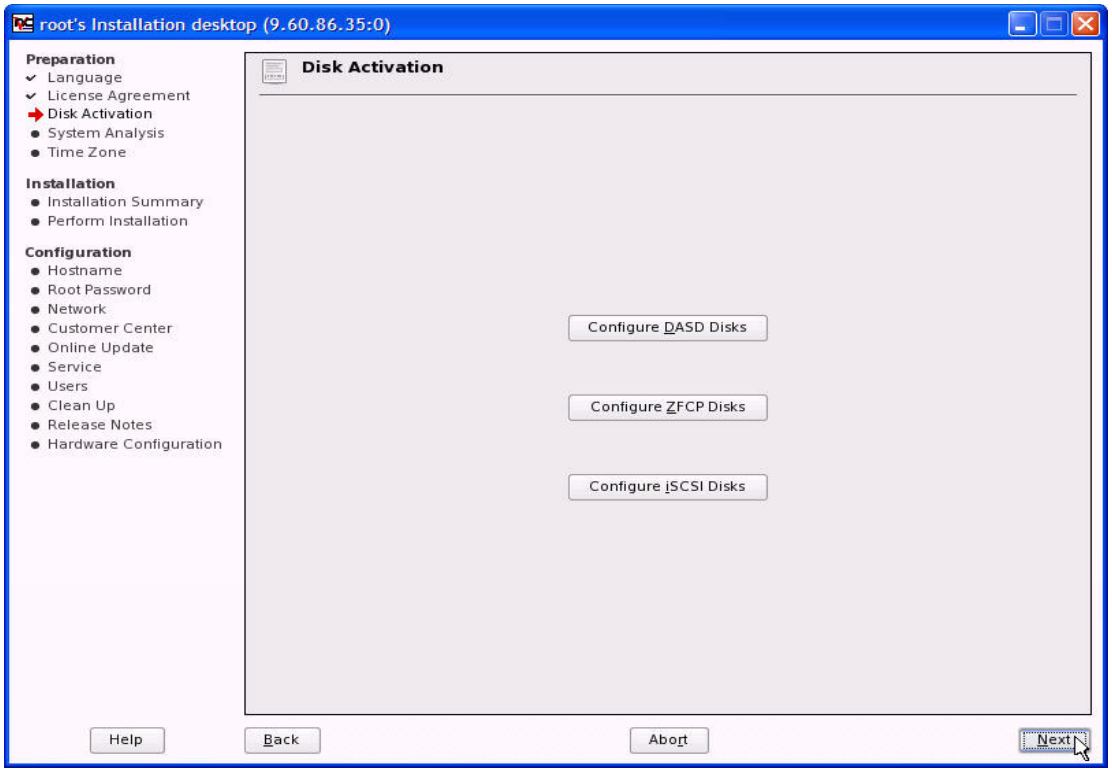
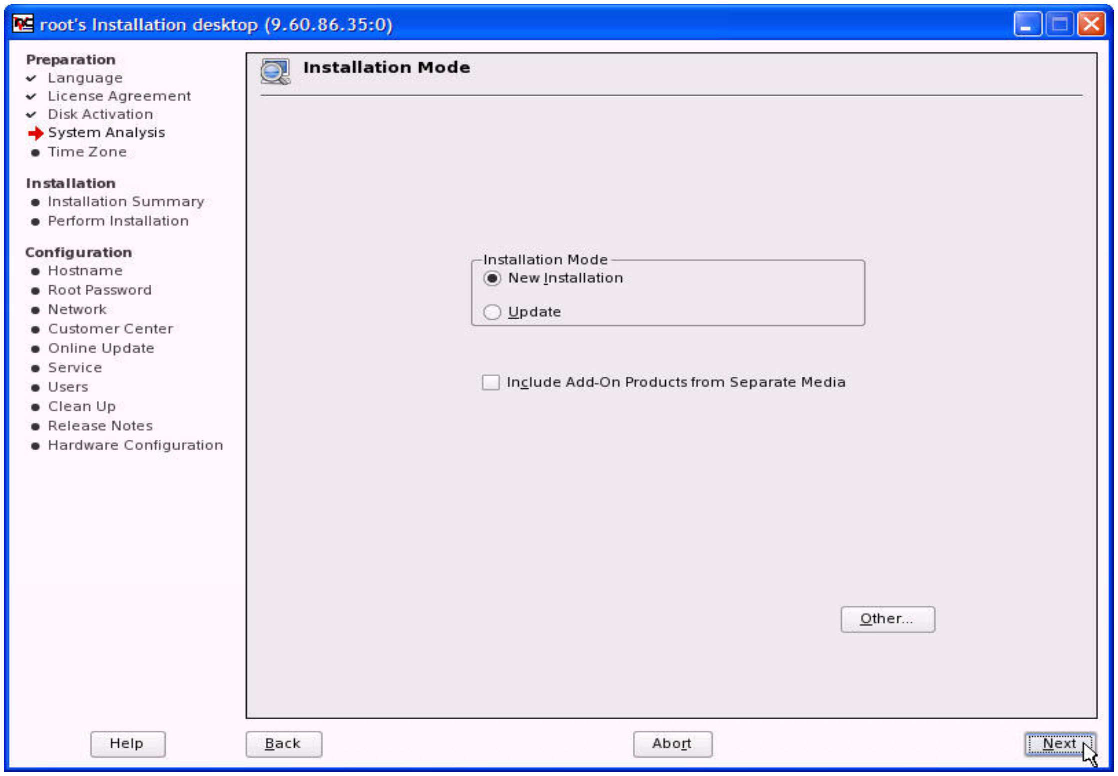


Figure 5-7 Disk Activation panel

8. Click **Next** to continue to the panel shown in Figure 5-8. More disks can be added at this time by clicking **Configure DASD Disks** to add ECKD/FBA devices, clicking **Configure ZFCP Disks** to add FCP devices, or clicking **Configure ISCSI Disks** to add ISCSI devices.



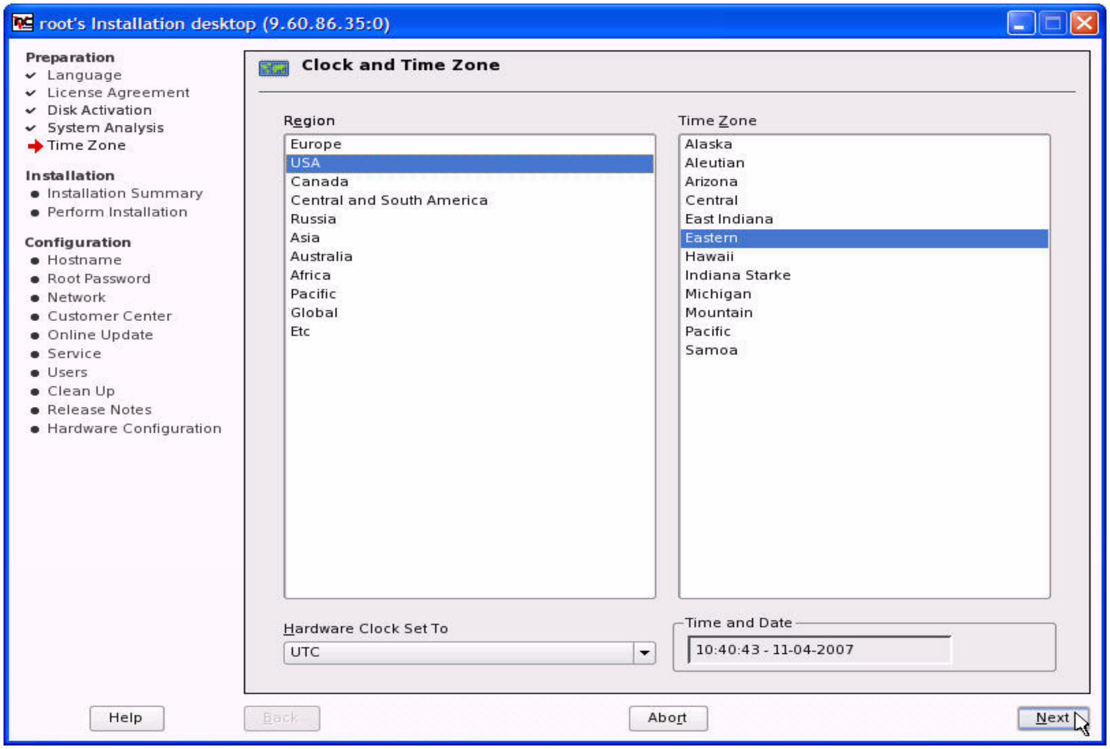
Installation Mode

Installation Mode
● New Installation
○ Update

□ Include Add-On Products from Separate Media

Figure 5-8 Installation Mode panel

9. Select the **New installation** radio button and then click **Next** to continue to Figure 5-9.



Region: USA

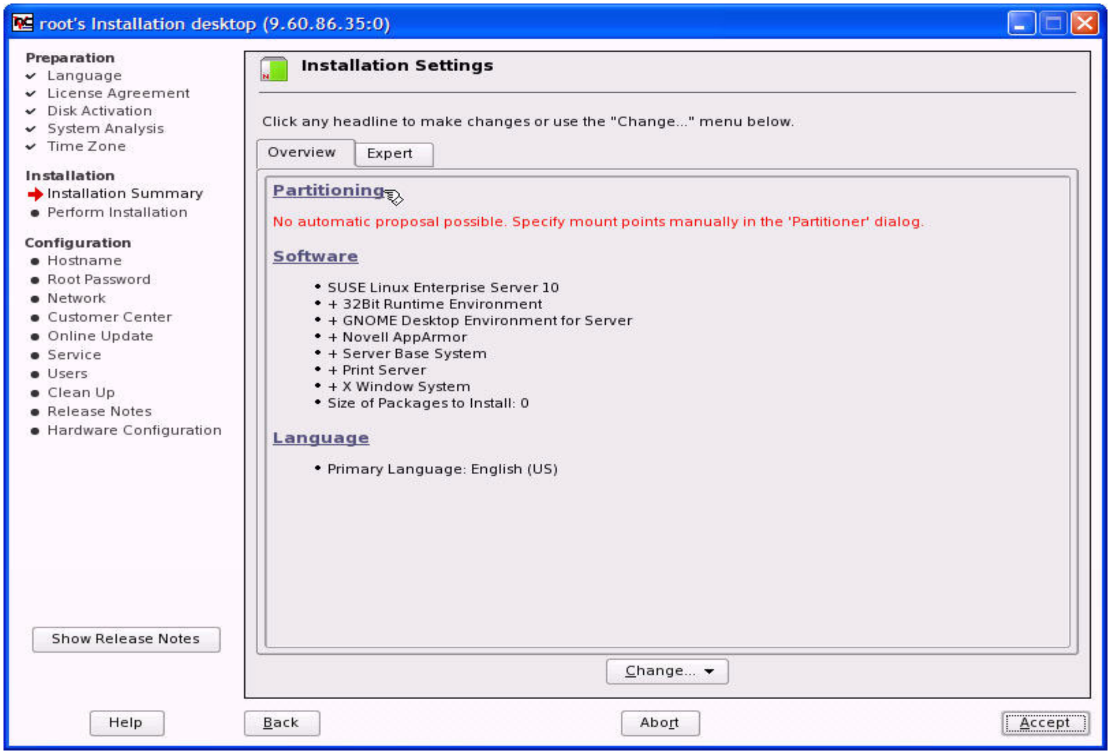
Time Zone: Eastern

Hardware Clock Set To: UTC

Time and Date: 10:40:43 - 11-04-2007

Figure 5-9 Clock and Time Zone panel

10. Select the desired region and time zone to set the system clock, then click **Next** to continue to the panel shown in Figure 5-10.



Installation Settings

Click any headline to make changes or use the "Change..." menu below.

Partitioning

No automatic proposal possible. Specify mount points manually in the 'Partitioner' dialog.

Software

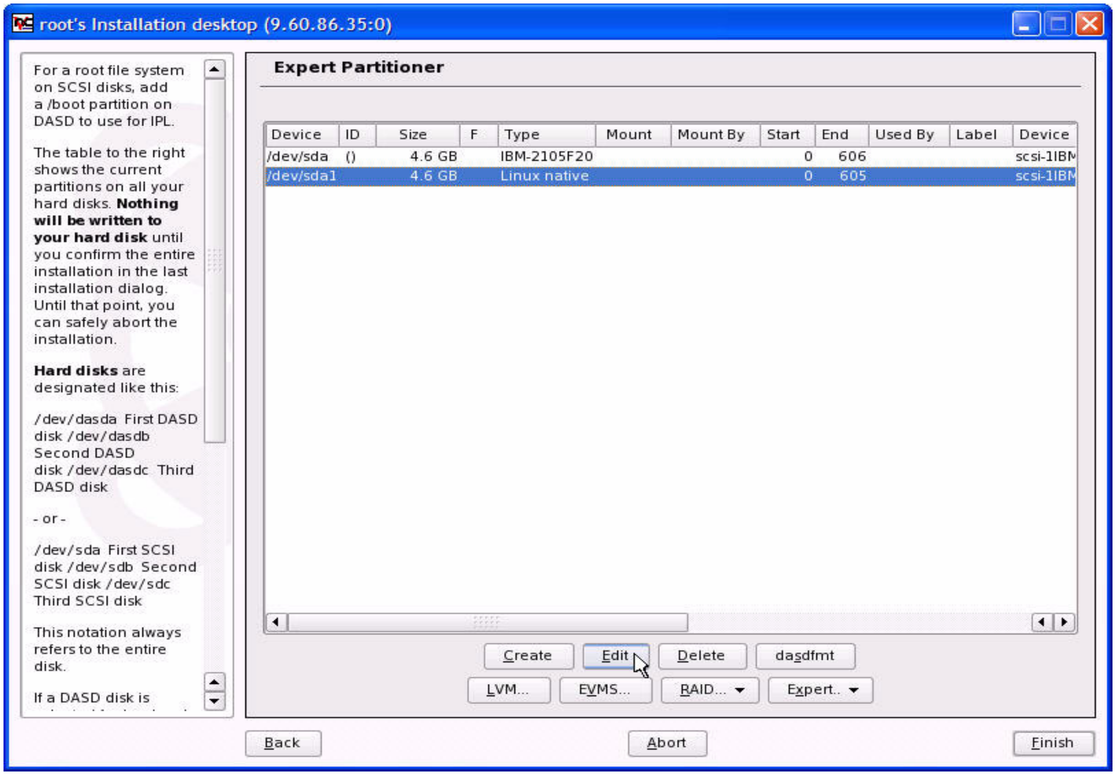
- SUSE Linux Enterprise Server 10
- + 32Bit Runtime Environment
- + GNOME Desktop Environment for Server
- + Novell AppArmor
- + Server Base System
- + Print Server
- + X Window System
- Size of Packages to Install: 0

Language

- Primary Language: English (US)

Figure 5-10 Installation Settings panel

11. Click **Partitioning** and continue to the panel shown in Figure 5-11 to configure the disk partitions.



For a root file system on SCSI disks, add a /boot partition on DASD to use for IPL.

The table to the right shows the current partitions on all your hard disks. **Nothing will be written to your hard disk** until you confirm the entire installation in the last installation dialog. Until that point, you can safely abort the installation.

Hard disks are designated like this:

/dev/dasda First DASD disk /dev/dasdb Second DASD disk /dev/dasdc Third DASD disk

- or -

/dev/sda First SCSI disk /dev/sdb Second SCSI disk /dev/sdc Third SCSI disk

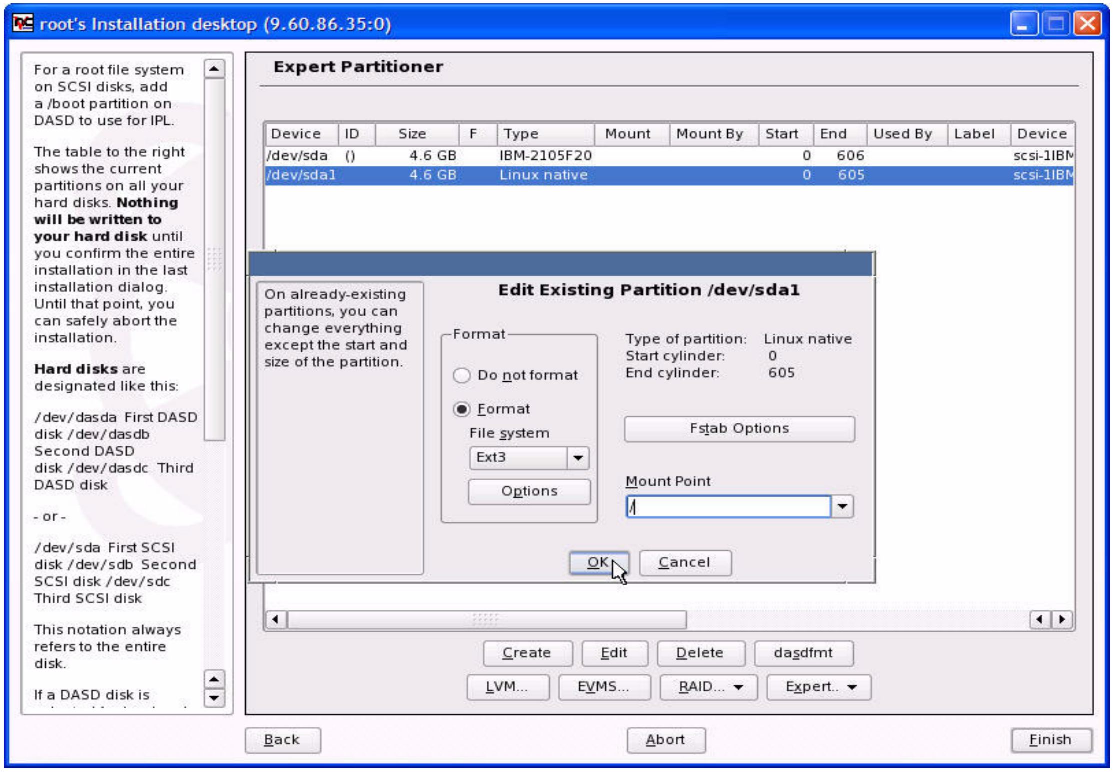
This notation always refers to the entire disk.

If a DASD disk is

Device	ID	Size	F	Type	Mount	Mount By	Start	End	Used By	Label	Device
/dev/sda	()	4.6 GB		IBM-2105F20			0	606			scsi-1IBM
/dev/sda1		4.6 GB		Linux native			0	605			scsi-1IBM

Figure 5-11 Expert Partitioner panel

12. Select a device and click **Create** to create a new partition, or select a partition and click **Edit** to change an existing partition's properties. Selecting **/dev/sda1** and clicking **Edit** displays the panel shown in Figure 5-12.



root's Installation desktop (9.60.86.35:0)

For a root file system on SCSI disks, add a /boot partition on DASD to use for IPL.

The table to the right shows the current partitions on all your hard disks. **Nothing will be written to your hard disk** until you confirm the entire installation in the last installation dialog. Until that point, you can safely abort the installation.

Hard disks are designated like this:

/dev/dasda First DASD disk /dev/dasdb Second DASD disk /dev/dasdc Third DASD disk

- or -

/dev/sda First SCSI disk /dev/sdb Second SCSI disk /dev/sdc Third SCSI disk

This notation always refers to the entire disk.

If a DASD disk is

Expert Partitioner

Device	ID	Size	F	Type	Mount	Mount By	Start	End	Used By	Label	Device
/dev/sda	()	4.6 GB		IBM-2105F20			0	606			scsi-1IBM
/dev/sda1		4.6 GB		Linux native			0	605			scsi-1IBM

Edit Existing Partition /dev/sda1

On already-existing partitions, you can change everything except the start and size of the partition.

Format

○ Do not format

● Format

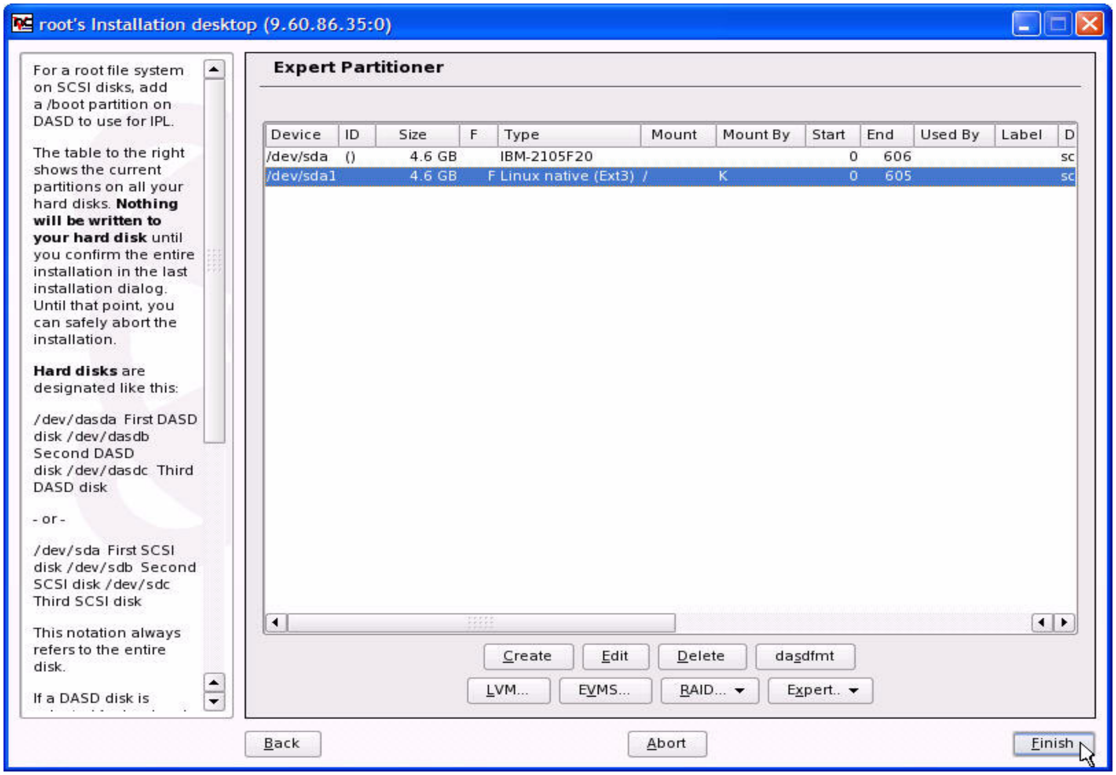
File system: Ext3

Type of partition: Linux native
Start cylinder: 0
End cylinder: 605

Mount Point: /

Figure 5-12 Edit Partition Properties panel

13. Set the partition properties. Choose the desired file system and mount point, then click **OK** to continue to the panel shown in Figure 5-13.



For a root file system on SCSI disks, add a /boot partition on DASD to use for IPL.

The table to the right shows the current partitions on all your hard disks. **Nothing will be written to your hard disk** until you confirm the entire installation in the last installation dialog. Until that point, you can safely abort the installation.

Hard disks are designated like this:

/dev/dasda First DASD disk /dev/dasdb Second DASD disk /dev/dasdc Third DASD disk

- or -

/dev/sda First SCSI disk /dev/sdb Second SCSI disk /dev/sdc Third SCSI disk

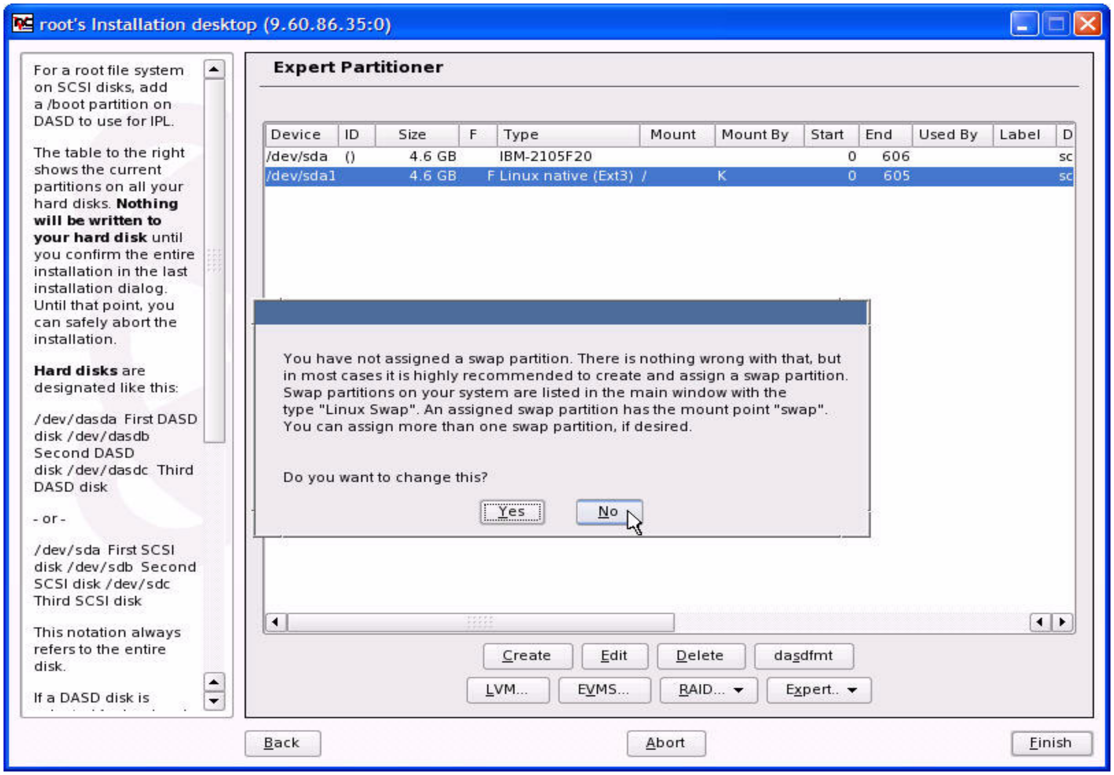
This notation always refers to the entire disk.

If a DASD disk is

Device	ID	Size	F	Type	Mount	Mount By	Start	End	Used By	Label	D
/dev/sda	()	4.6 GB		IBM-2105F20			0	606			sc
/dev/sda1		4.6 GB	F	Linux native (Ext3)	/	K	0	605			sc

Figure 5-13 Expert Partitioner panel

In this example the entire SLES10 file system is installed on a single 4.6 gig LUN (/dev/sda1) and mounted as '/', using the ext3 file system. Click **Finish** to continue to the panel shown in Figure 5-14.



For a root file system on SCSI disks, add a /boot partition on DASD to use for IPL.

The table to the right shows the current partitions on all your hard disks. **Nothing will be written to your hard disk** until you confirm the entire installation in the last installation dialog. Until that point, you can safely abort the installation.

Hard disks are designated like this:

/dev/dasda First DASD disk /dev/dasdb Second DASD disk /dev/dasdc Third DASD disk

- or -

/dev/sda First SCSI disk /dev/sdb Second SCSI disk /dev/sdc Third SCSI disk

This notation always refers to the entire disk.

If a DASD disk is

Device	ID	Size	F	Type	Mount	Mount By	Start	End	Used By	Label	D
/dev/sda	()	4.6 GB		IBM-2105F20			0	606			sc
/dev/sda1		4.6 GB	F	Linux native (Ext3)	/	K	0	605			sc

You have not assigned a swap partition. There is nothing wrong with that, but in most cases it is highly recommended to create and assign a swap partition. Swap partitions on your system are listed in the main window with the type "Linux Swap". An assigned swap partition has the mount point "swap". You can assign more than one swap partition, if desired.

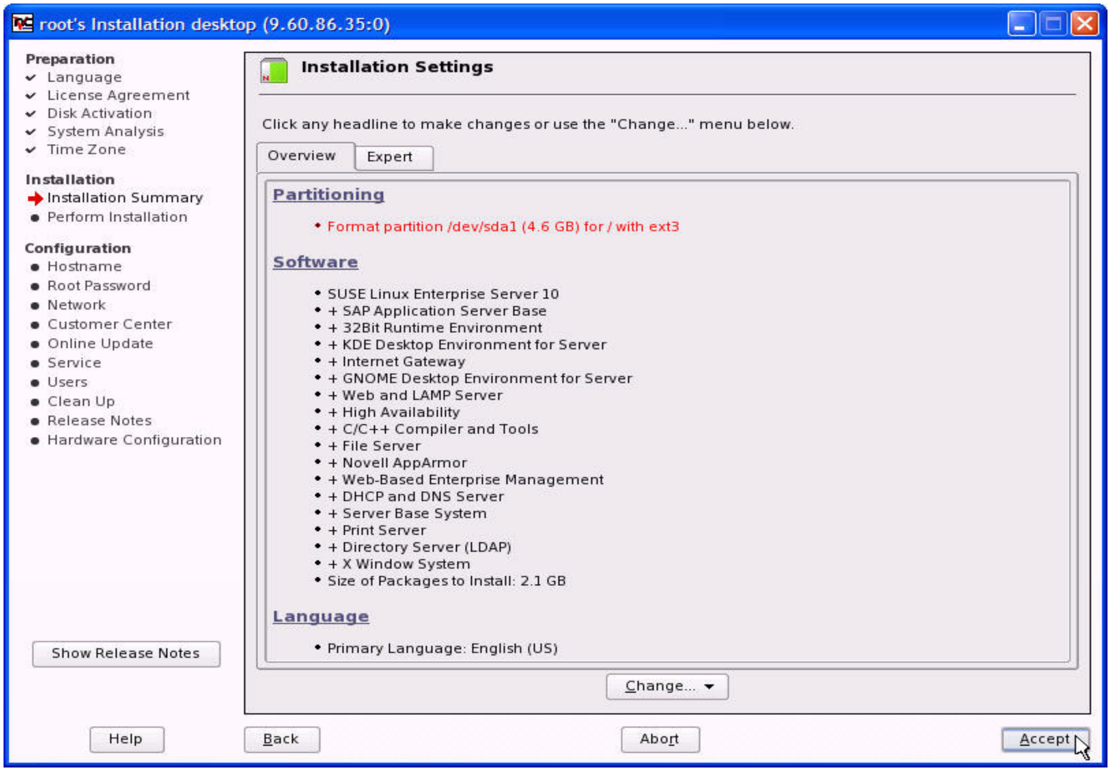
Do you want to change this?

Figure 5-14 No swap partition message

The steps are:

1. If no swap partition is required, click **No** to continue to Figure 5-15. Otherwise, click **Yes** to go back and create one.

> **Note:** In order to conserve DASD, this installation was performed without defining a Linux swap partition. However, we highly recommend creating a swap partition for any production Linux machine.



Installation Settings

Click any headline to make changes or use the "Change..." menu below.

Partitioning

- Format partition /dev/sda1 (4.6 GB) for / with ext3

Software

- SUSE Linux Enterprise Server 10
- + SAP Application Server Base
- + 32Bit Runtime Environment
- + KDE Desktop Environment for Server
- + Internet Gateway
- + GNOME Desktop Environment for Server
- + Web and LAMP Server
- + High Availability
- + C/C++ Compiler and Tools
- + File Server
- + Novell AppArmor
- + Web-Based Enterprise Management
- + DHCP and DNS Server
- + Server Base System
- + Print Server
- + Directory Server (LDAP)
- + X Window System
- Size of Packages to Install: 2.1 GB

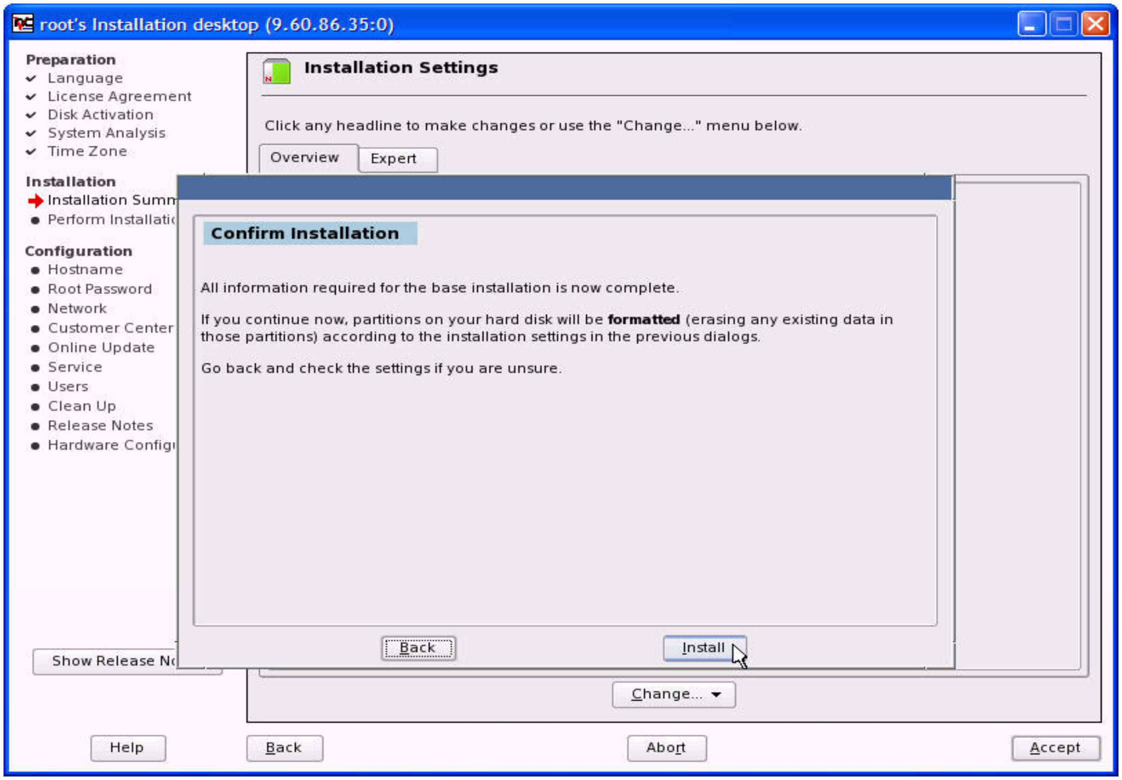
Language

- Primary Language: English (US)

Figure 5-15 Installation Settings panel

2. Click **Accept** to continue the installation process in Figure 5-16.

> **Note:** Remember to complete the remaining installation steps on this panel. Additional tasks can be found under the Expert tab. The images of the panels for the following steps were excluded from this example:
>
> 1. Software selection
> 2. Add-on products
> 3. Boot options
> 4. Default run level



Click any headline to make changes or use the "Change..." menu below.

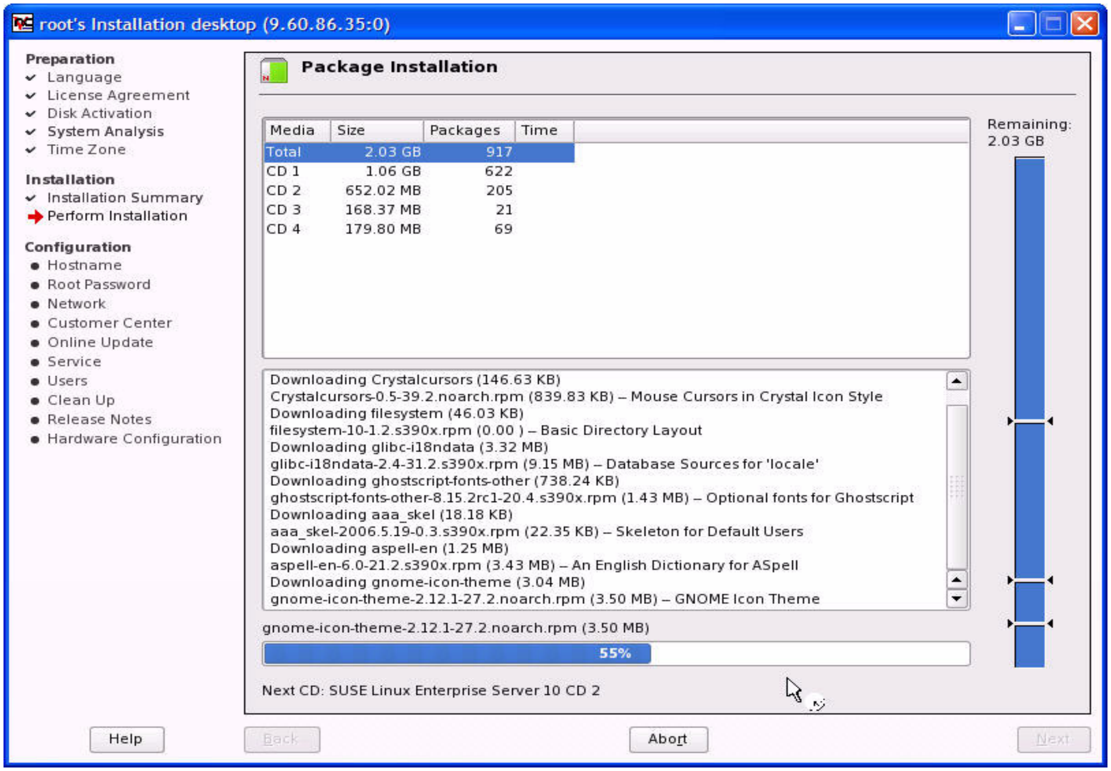
All information required for the base installation is now complete.

If you continue now, partitions on your hard disk will be **formatted** (erasing any existing data in those partitions) according to the installation settings in the previous dialogs.

Go back and check the settings if you are unsure.

Figure 5-16 Confirm installation

5. Click **Install** to start the installation and continue with the panel shown in Figure 5-17.



Media	Size	Packages	Time
Total	2.03 GB	917	
CD 1	1.06 GB	622	
CD 2	652.02 MB	205	
CD 3	168.37 MB	21	
CD 4	179.80 MB	69	

```
Downloading Crystalcursors (146.63 KB)
Crystalcursors-0.5-39.2.noarch.rpm (839.83 KB) – Mouse Cursors in Crystal Icon Style
Downloading filesystem (46.03 KB)
filesystem-10-1.2.s390x.rpm (0.00 ) – Basic Directory Layout
Downloading glibc-i18ndata (3.32 MB)
glibc-i18ndata-2.4-31.2.s390x.rpm (9.15 MB) – Database Sources for 'locale'
Downloading ghostscript-fonts-other (738.24 KB)
ghostscript-fonts-other-8.15.2rc1-20.4.s390x.rpm (1.43 MB) – Optional fonts for Ghostscript
Downloading aaa_skel (18.18 KB)
aaa_skel-2006.5.19-0.3.s390x.rpm (22.35 KB) – Skeleton for Default Users
Downloading aspell-en (1.25 MB)
aspell-en-6.0-21.2.s390x.rpm (3.43 MB) – An English Dictionary for ASpell
Downloading gnome-icon-theme (3.04 MB)
gnome-icon-theme-2.12.1-27.2.noarch.rpm (3.50 MB) – GNOME Icon Theme
```

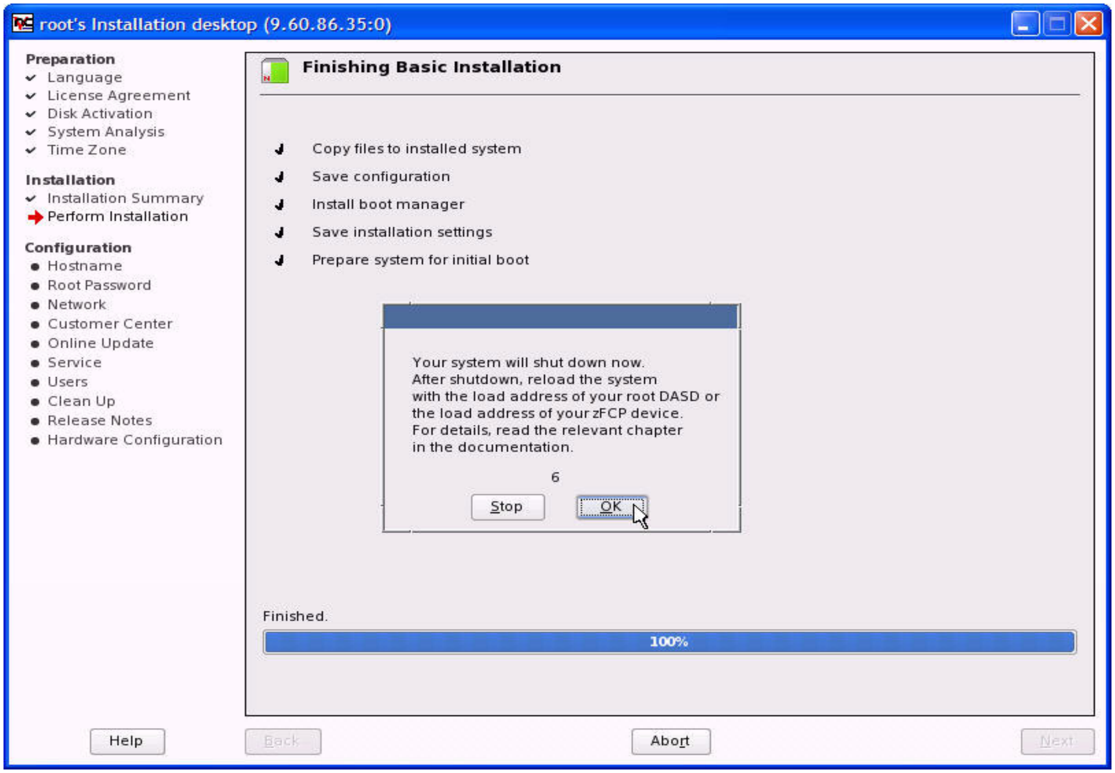
gnome-icon-theme-2.12.1-27.2.noarch.rpm (3.50 MB)

55%

Next CD: SUSE Linux Enterprise Server 10 CD 2

Figure 5-17 Package Installation panel

6. The install process can take quite a while, depending on the number of packages selected and the current load on the underlying z/VM system. After completion, the window shown in Figure 5-18 is displayed.



Your system will shut down now.
After shutdown, reload the system
with the load address of your root DASD or
the load address of your zFCP device.
For details, read the relevant chapter
in the documentation.

6

Figure 5-18 Finishing Basic Installation panel

7. Click **OK** to finish the basic installation. The Linux installation system is shut down and the VNC session is terminated. The virtual machine is placed in a CP disabled wait and must be IPLed from the FCP device to complete the installation.

8. The SET LOADDEV command is used to identify the location of a program to be loaded as a result of a z/VM guest IPL from SCSI disk. Specify the LOADDEV parameters using the PORTNAME and LUN that was entered in

Figure 5-5 on page 56 for the FCP device. The SET LOADDEV command can be issued dynamically, as shown in Figure 5-19, or automated by placing it in the VM directory of the Linux guest, as shown in Figure 5-20 on page 71.

> **Important:** The FCP device number, WWPN, and LUN are part of the Linux IPL records on disk. The Linux kernel panics if it is IPLed using a different FCP device, WWPN, or LUN than was used to install the Linux system.

```
SET LOADDEV PORTNAME 50050763 00C300AA LUN 50500000 00000000
I 6008
HCPLDI2816I Acquiring the machine loader from the processor controller.
HCPLDI2817I Load completed from the processor controller.
HCPLDI2817I Now starting the machine loader.
MLOEVL012I: Machine loader up and running (version 0.18).
MLOPDM003I: Machine loader finished, moving data to final storage location.
Linux version 2.6.16.21-0.8-default (geeko@buildhost) (gcc version 4.1.0 (SUSE L
inux)) #1 SMP Mon Jul 3 18:25:39 UTC 2006
We are running under VM (64 bit mode)
Detected 1 CPU's
Boot cpu address  0
Built 1 zonelists
Kernel command line: root=/dev/sda1   TERM=dumb
PID hash table entries: 4096 (order: 12, 131072 bytes)
Dentry cache hash table entries: 131072 (order: 8, 1048576 bytes)
Inode-cache hash table entries: 65536 (order: 7, 524288 bytes)
Memory: 501760k/524288k available (4298k kernel code, 0k reserved, 1401k data, 1
96k init)
Security Framework v1.0.0 initialized
Mount-cache hash table entries: 256
checking if image is initramfs... it is
```

Figure 5-19   IPL from FCP device

Note the syntax of the SET LOADDEV command. The spaces between the first 4 bytes and the second 4 bytes of both the PORTNAME and LUN parameters are required. The SLES10 system boots and the VNC client must reconnect to the system to continue with YaST and complete the installation of Linux. See Figure 5-20.

```
USER SLES10 SLES10 512M 512M  G
  MACHINE ESA
  IPL CMS
* IDENTIFY WWPN AND LUN OF SCSI IPL DEVICE
LOADDEV PORT 5005076300C300AA
LOADDEV LUN 5050000000000000
  CONSOLE 001F 3215 T
  SPOOL 000C 2540 READER *
  SPOOL 000D 2540 PUNCH A
  SPOOL 000E 1403 A
  LINK MAINT 0190 0190 RR
  LINK MAINT 019D 019D RR
  LINK MAINT 019E 019E RR
  LINK TCPMAINT 0592 0592 RR
* BEGIN DEDICATED FCP DEVICES
  DEDICATE 6008 6008
  DEDICATE 6009 6009
  DEDICATE 6210 6210
  DEDICATE 6211 6211
* END DEDICATED FCP DEVICES
  NICDEF 500 TYPE QDIO DEV 3 LAN SYSTEM TESTSWIT
  MDISK 191 3390 3150 50 510W02  MR READ     WRITE    MULTIPLE
*
```

Figure 5-20 VM directory entry

Note that the syntax of the LOADDEV statement in the VM directory entry differs from the syntax of the SET LOADDEV command in that it does not require spaces between the first 4 bytes and the second 4 bytes of the PORTNAME or LUN.

5.2 Adding FCP devices to an existing SLES 10 system

This section discusses two ways of adding FCP devices to an existing SLES 10 system.

5.2.1 Adding disks with YaST

For a graphical user interface to add FCP disks:

1. Start YaST from a bash shell using a VNC client/server, and the window shown in Figure 5-21 will be displayed.

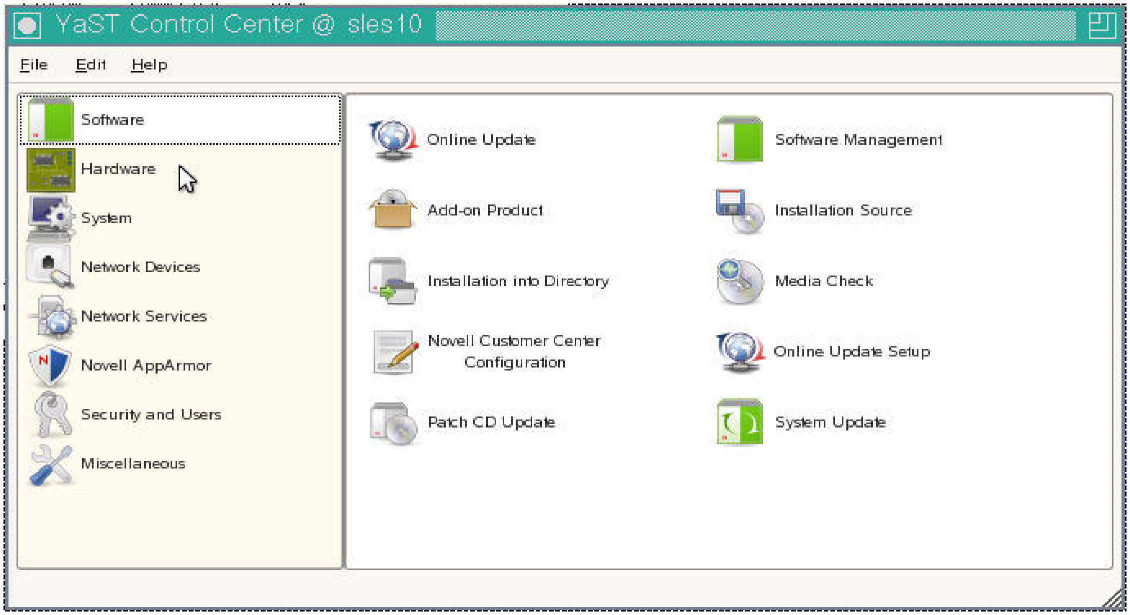


Figure 5-21 YaST Control Center Software panel

2. Click **Hardware** in the navigation section on the left side of the panel to go to Figure 5-22.

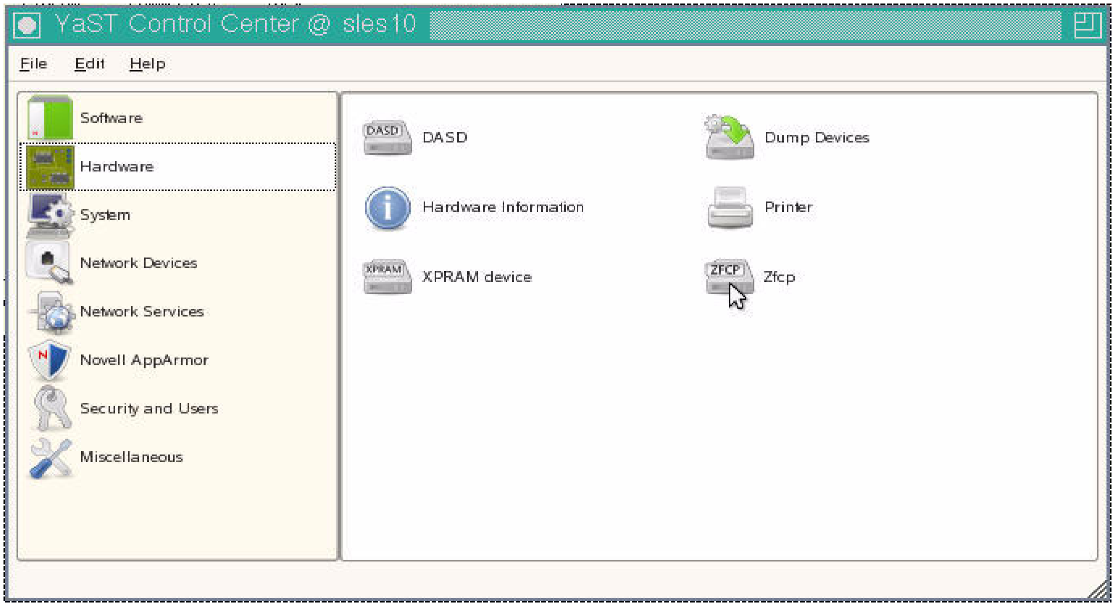
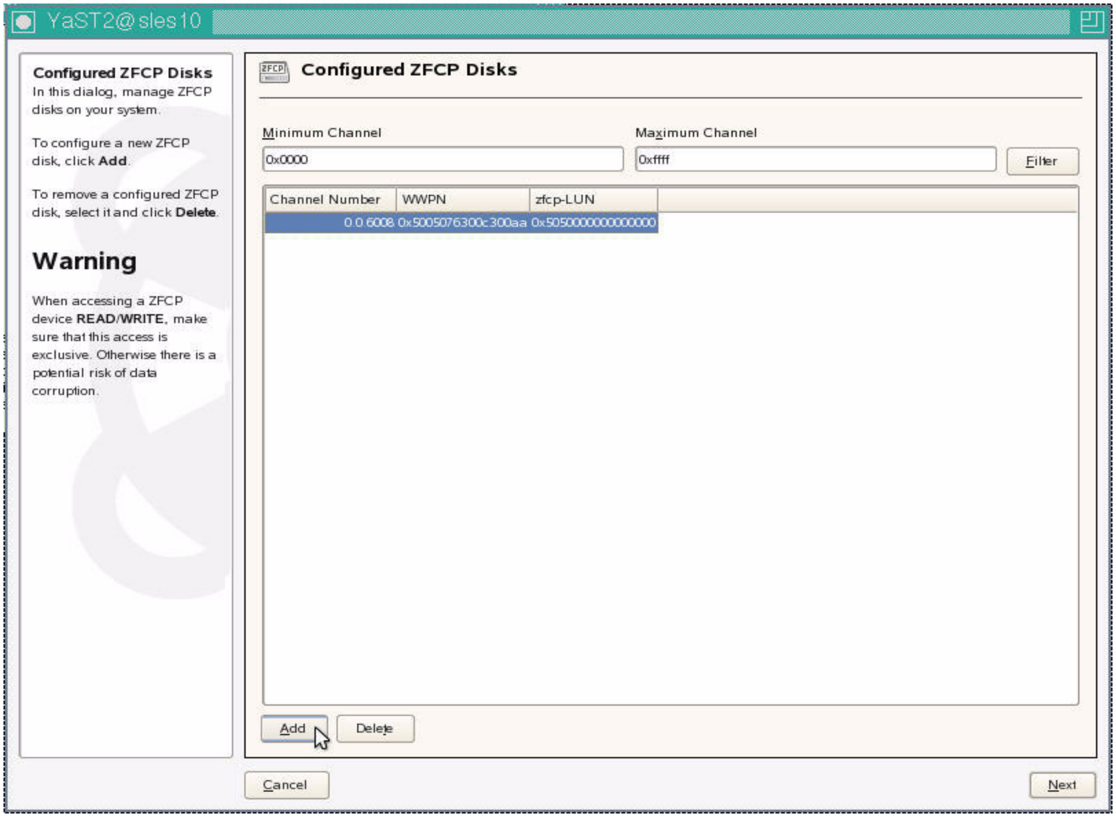


Figure 5-22 YaST Control Center Hardware panel

3. Click the ZFCP icon to go to the panel shown in Figure 5-23.



Configured ZFCP Disks
In this dialog, manage ZFCP disks on your system.

To configure a new ZFCP disk, click **Add**.

To remove a configured ZFCP disk, select it and click **Delete**.

Warning

When accessing a ZFCP device **READ/WRITE**, make sure that this access is exclusive. Otherwise there is a potential risk of data corruption.

Configured ZFCP Disks

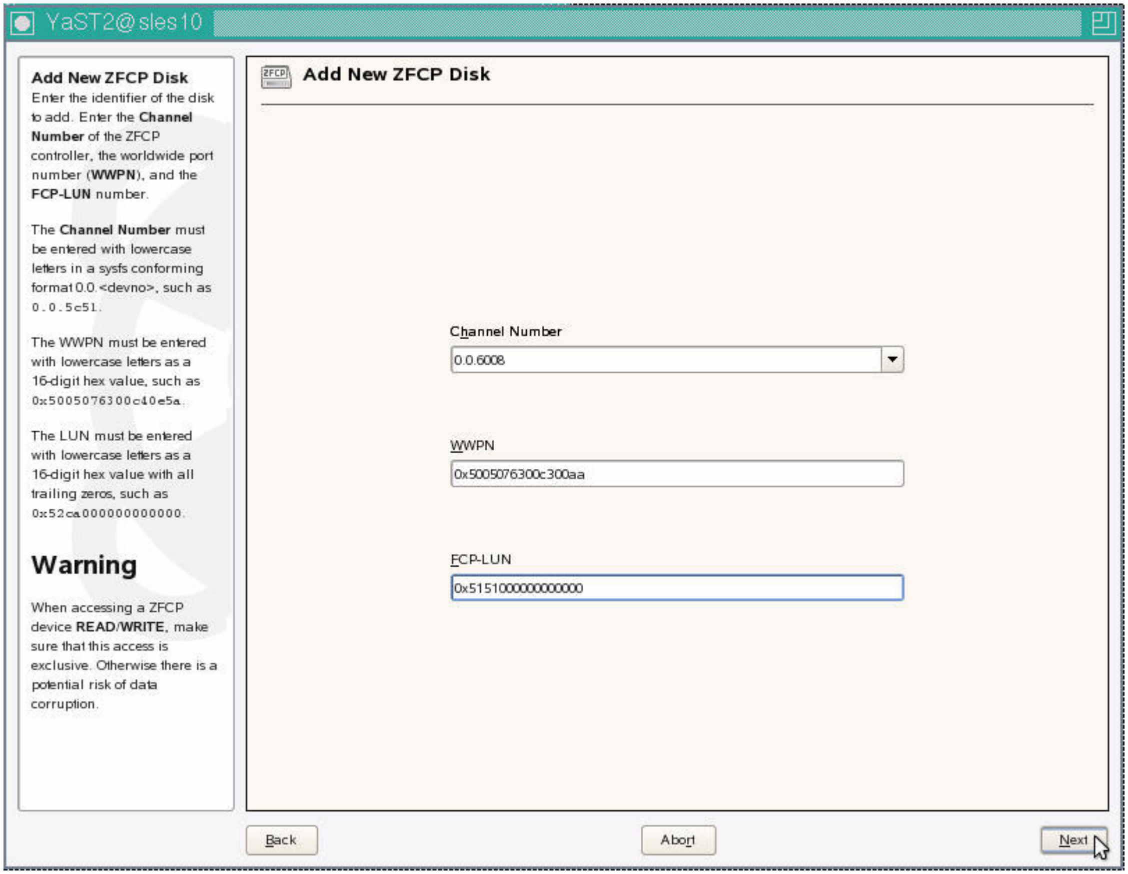
Minimum Channel: 0x0000

Maximum Channel: 0xffff

Channel Number	WWPN	zfcp-LUN
0.0.6008	0x5005076300c300aa	0x5050000000000000

Figure 5-23 Configured ZFCP disks

4. Currently defined FCP disks are displayed. Click **Add** to go to the panel shown in Figure 5-24.



Add New ZFCP Disk
Enter the identifier of the disk to add. Enter the **Channel Number** of the ZFCP controller, the worldwide port number (**WWPN**), and the **FCP-LUN** number.

The **Channel Number** must be entered with lowercase letters in a sysfs conforming format 0.0.<devno>, such as `0.0.5c51`.

The WWPN must be entered with lowercase letters as a 16-digit hex value, such as `0x5005076300c40e5a`.

The LUN must be entered with lowercase letters as a 16-digit hex value with all trailing zeros, such as `0x52ca000000000000`.

Warning

When accessing a ZFCP device **READ/WRITE**, make sure that this access is exclusive. Otherwise there is a potential risk of data corruption.

Add New ZFCP Disk

Channel Number: 0.0.6008

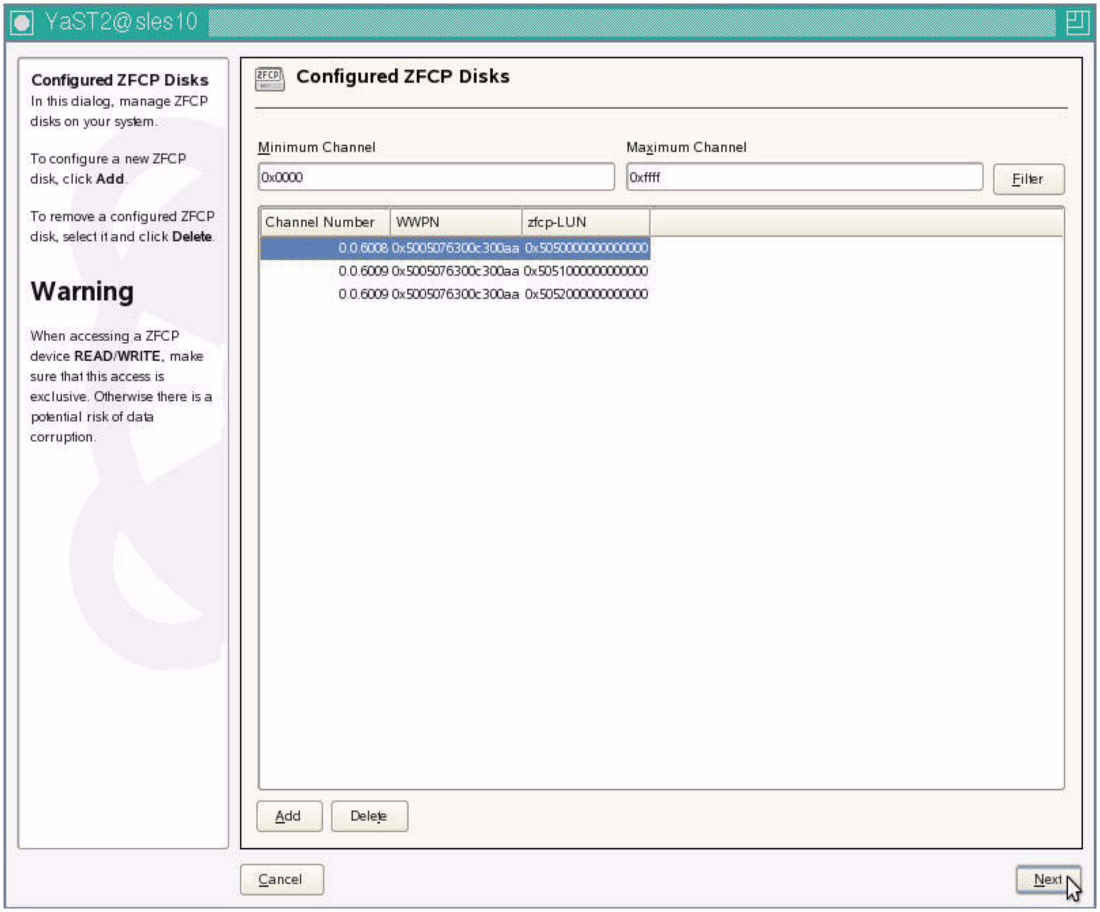
WWPN: 0x5005076300c300aa

FCP-LUN: 0x5151000000000000

Figure 5-24 Add New ZFCP disk

5. To add a new zfcp disk, select the FCP device from the Channel Number drop-down list. Enter the WWPN of the storage device that holds the LUN into the middle field. Enter the FCP-LUN number in the bottom field. Click **Next** to continue to the panel shown in Figure 5-25.

> **Caution:** The WWPN and FCP-LUN fields must be entered in lowercase letters as a 16-digit hex value with all trailing zeros and begin with the character string prefix 0x. Leaving this prefix off the WWPN field resulted in a Invalid WWPN pop-up window when **Next** is clicked to continue.



Configured ZFCP Disks
In this dialog, manage ZFCP disks on your system.

To configure a new ZFCP disk, click **Add**.

To remove a configured ZFCP disk, select it and click **Delete**.

Warning

When accessing a ZFCP device **READ/WRITE**, make sure that this access is exclusive. Otherwise there is a potential risk of data corruption.

Configured ZFCP Disks

Minimum Channel: 0x0000
Maximum Channel: 0xffff

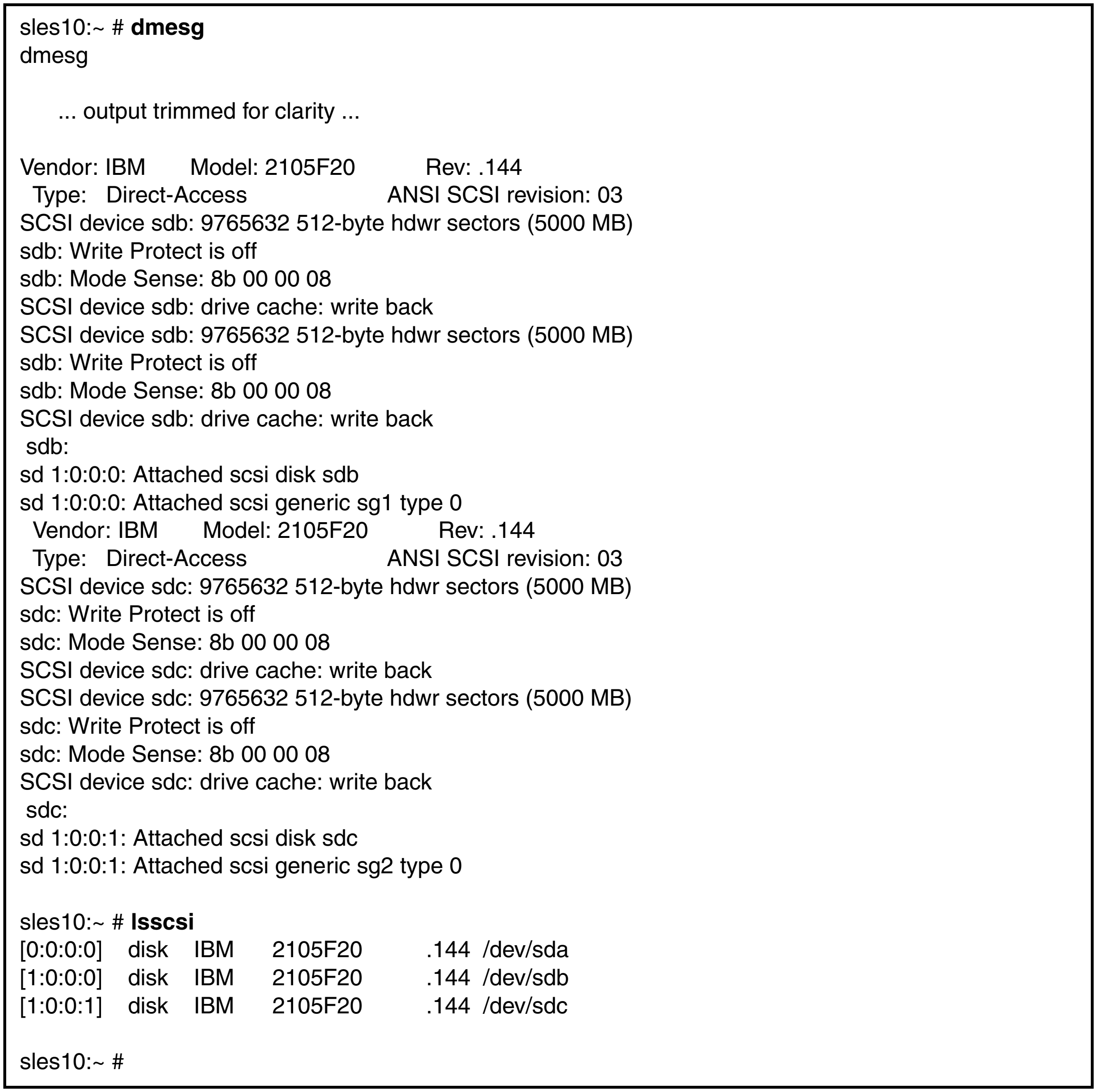
Channel Number	WWPN	zfcp-LUN
0.0.6008	0x5005076300c300aa	0x5050000000000000
0.0.6009	0x5005076300c300aa	0x5051000000000000
0.0.6009	0x5005076300c300aa	0x5052000000000000

Figure 5-25 Configured ZFCP disks

6. Continue using the **Add** button to add devices to the list until all required devices have been added. Unwanted devices can also be removed by

highlighting the device to be removed and clicking the **Delete** button. In this example two LUNs (0x5051000000000000 and 0x5052000000000000) were added behind WWPN 0x5005076300c300aa on FCP device number 6009. When all additions/deletions are complete, click **Next** to finish.

The **dmesg** command or **lsscsi command** (as shown in Figure 5-26) can be used to verify that the new FCP devices are available for use.

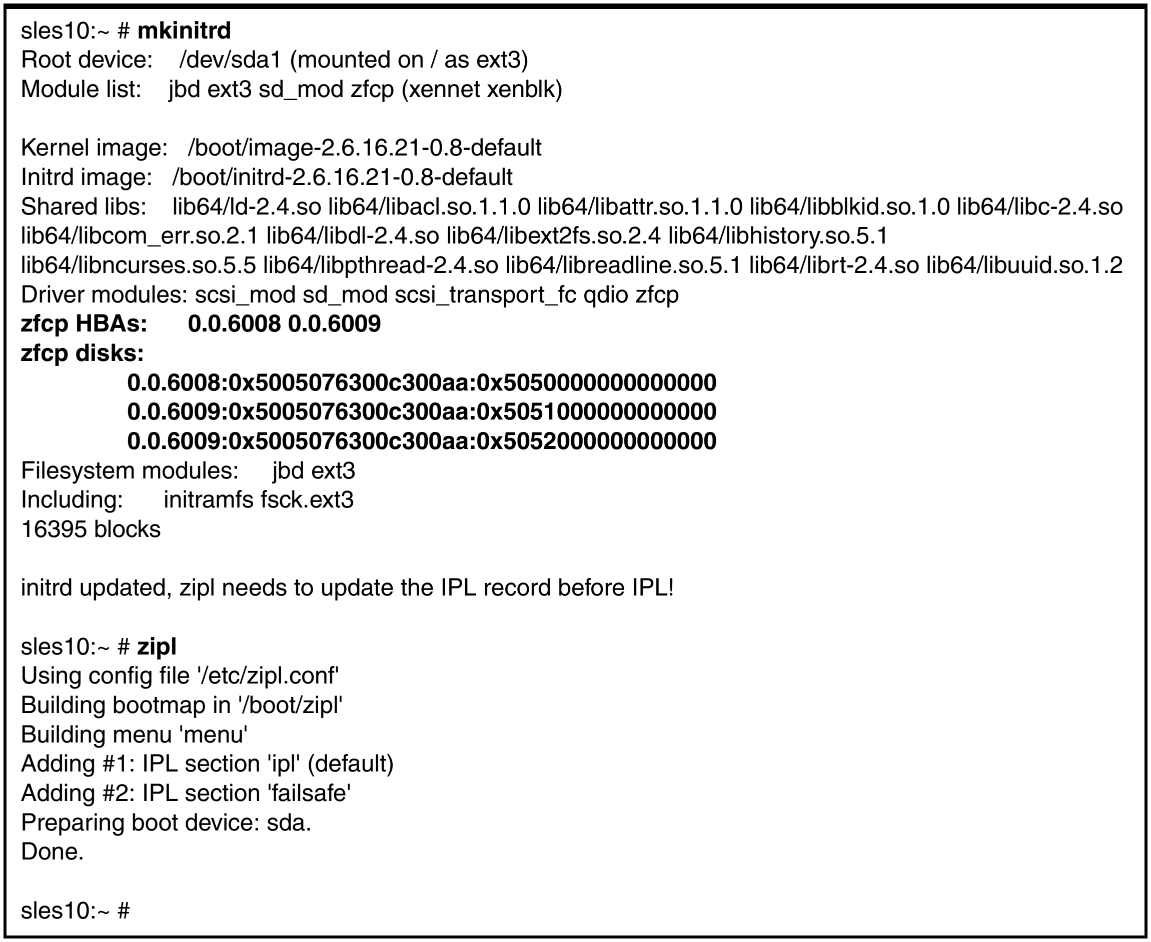


```
sles10:~ # dmesg
dmesg

    ... output trimmed for clarity ...

Vendor: IBM      Model: 2105F20         Rev: .144
  Type:   Direct-Access                  ANSI SCSI revision: 03
SCSI device sdb: 9765632 512-byte hdwr sectors (5000 MB)
sdb: Write Protect is off
sdb: Mode Sense: 8b 00 00 08
SCSI device sdb: drive cache: write back
SCSI device sdb: 9765632 512-byte hdwr sectors (5000 MB)
sdb: Write Protect is off
sdb: Mode Sense: 8b 00 00 08
SCSI device sdb: drive cache: write back
 sdb:
sd 1:0:0:0: Attached scsi disk sdb
sd 1:0:0:0: Attached scsi generic sg1 type 0
  Vendor: IBM      Model: 2105F20         Rev: .144
  Type:   Direct-Access                  ANSI SCSI revision: 03
SCSI device sdc: 9765632 512-byte hdwr sectors (5000 MB)
sdc: Write Protect is off
sdc: Mode Sense: 8b 00 00 08
SCSI device sdc: drive cache: write back
SCSI device sdc: 9765632 512-byte hdwr sectors (5000 MB)
sdc: Write Protect is off
sdc: Mode Sense: 8b 00 00 08
SCSI device sdc: drive cache: write back
 sdc:
sd 1:0:0:1: Attached scsi disk sdc
sd 1:0:0:1: Attached scsi generic sg2 type 0

sles10:~ # lsscsi
[0:0:0:0]   disk   IBM    2105F20       .144  /dev/sda
[1:0:0:0]   disk   IBM    2105F20       .144  /dev/sdb
[1:0:0:1]   disk   IBM    2105F20       .144  /dev/sdc

sles10:~ #
```

Figure 5-26 New device messages

Devices /dev/sdb and /dev/sdc can now be partitioned for use with **fdisk**.

Caution: At this point the new FCP devices have been added to the appropriate configuration files and made available for use. However, the new LUNs have not been added to the Linux initial ram disk that is loaded at boot time. Prior to shutdown, `mkinitrd` and `zipl` should be run in order to rebuild the initial ram disk and update the Linux boot record. The required commands and their output are shown in Figure 5-27.



```
sles10:~ # mkinitrd
Root device:    /dev/sda1 (mounted on / as ext3)
Module list:    jbd ext3 sd_mod zfcp (xennet xenblk)

Kernel image:   /boot/image-2.6.16.21-0.8-default
Initrd image:   /boot/initrd-2.6.16.21-0.8-default
Shared libs:    lib64/ld-2.4.so lib64/libacl.so.1.1.0 lib64/libattr.so.1.1.0 lib64/libblkid.so.1.0 lib64/libc-2.4.so
lib64/libcom_err.so.2.1 lib64/libdl-2.4.so lib64/libext2fs.so.2.4 lib64/libhistory.so.5.1
lib64/libncurses.so.5.5 lib64/libpthread-2.4.so lib64/libreadline.so.5.1 lib64/librt-2.4.so lib64/libuuid.so.1.2
Driver modules: scsi_mod sd_mod scsi_transport_fc qdio zfcp
zfcp HBAs:      0.0.6008 0.0.6009
zfcp disks:
                0.0.6008:0x5005076300c300aa:0x5050000000000000
                0.0.6009:0x5005076300c300aa:0x5051000000000000
                0.0.6009:0x5005076300c300aa:0x5052000000000000
Filesystem modules:     jbd ext3
Including:      initramfs fsck.ext3
16395 blocks

initrd updated, zipl needs to update the IPL record before IPL!

sles10:~ # zipl
Using config file '/etc/zipl.conf'
Building bootmap in '/boot/zipl'
Building menu 'menu'
Adding #1: IPL section 'ipl' (default)
Adding #2: IPL section 'failsafe'
Preparing boot device: sda.
Done.

sles10:~ #
```

Figure 5-27 mkinitrd and zipl command output

5.2.2 Adding disks manually

The SLES10 system maintains hardware configuration information in the /etc/sysconfig/hardware directory. By creating or editing files there and echoing values to strategic locations in the /sys file system, devices can be added to and deleted from the system without the use of YaST.

There are three cases of how one might want to add LUNs to an existing system:

- Add LUNs on a new FCP device.
- Add LUNs to an existing FCP device behind a new WWPN.
- Add LUNs to an existing FCP device behind an existing WWPN.

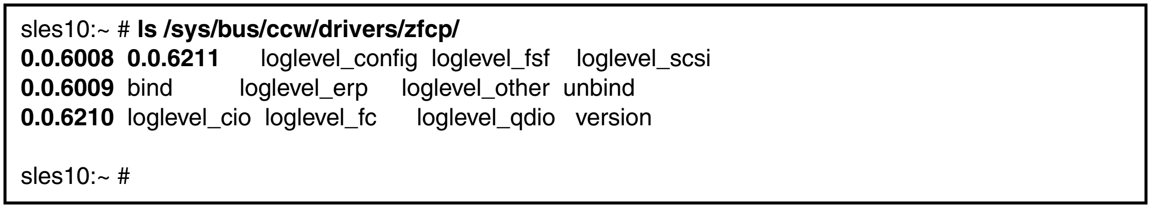
These three cases are examined in sequence.

Add LUNs on a new FCP device

Adding LUNs to a new FCP device requires the following steps:

1. Verify that the FCP device is available to the Linux system.
2. Create a hardware configuration file that describes the FCP device, WWPNs, and LUNs that will be added.
3. Bring the new FCP device online.

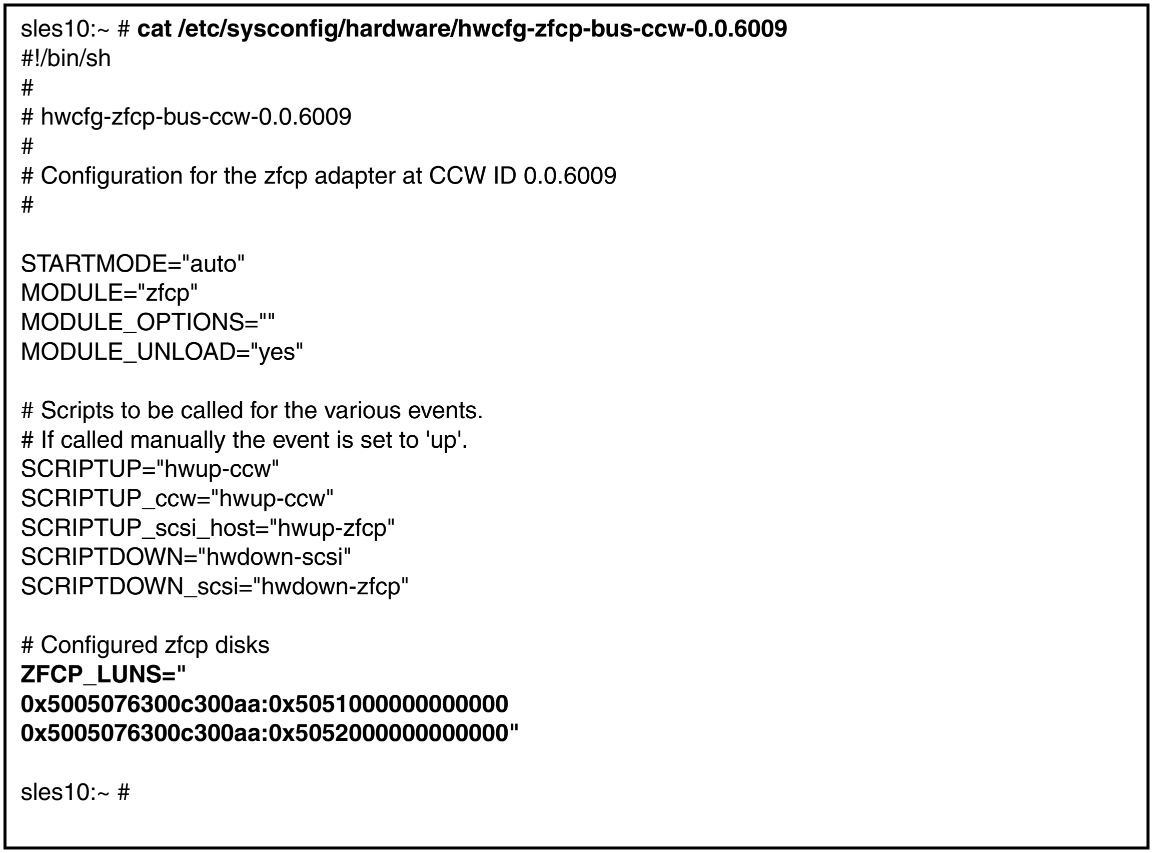
Verify that the FCP device is available to the Linux system by checking that its directory entry exists in the /sys file system. When the zfcp driver is loaded, it will detect all FCP devices that are available to Linux. In Figure 5-28 four devices have been detected and available: 6008, 6009, 6210, and 6211.



```
sles10:~ # ls /sys/bus/ccw/drivers/zfcp/
0.0.6008  0.0.6211     loglevel_config  loglevel_fsf    loglevel_scsi
0.0.6009  bind         loglevel_erp     loglevel_other  unbind
0.0.6210  loglevel_cio  loglevel_fc      loglevel_qdio   version

sles10:~ #
```

Figure 5-28 ZFCP devices

If the system already has an FCP device configured, then there will be a file in the /etc/sysconfig/hardware directory with a name in the form of hwcfg-zfcp-bus-ccw-0.0.*NNNN* (where *NNNN* is the FCP device number). To add a new FCP device, copy this file to a new file with the name hwcfg-zfcp-bus-ccw-0.0.*YYYY* (where *YYYY* is the FCP device number of the new FCP device). Then edit the ZFCP_LUNS parameter to reflect the new WWPNs and LUNs. In this example, the root file system was installed on FCP device 6008, so the /etc/sysconfig/hardware/hwcfg-zfcp-bus-ccw-0.0.6008 file already existed. This file was copied to /etc/sysconfig/hardware/hwcfg-zfcp-bus-ccw-0.0.6009 and modified. The contents of the file define the FCP device 6009, two LUNs (0x5051000000000000 and 0x5052000000000000), behind a WWPN of 0x5005076300c300aa. The file contents are shown in Figure 5-29.



```
sles10:~ # cat /etc/sysconfig/hardware/hwcfg-zfcp-bus-ccw-0.0.6009
#!/bin/sh
#
# hwcfg-zfcp-bus-ccw-0.0.6009
#
# Configuration for the zfcp adapter at CCW ID 0.0.6009
#

STARTMODE="auto"
MODULE="zfcp"
MODULE_OPTIONS=""
MODULE_UNLOAD="yes"

# Scripts to be called for the various events.
# If called manually the event is set to 'up'.
SCRIPTUP="hwup-ccw"
SCRIPTUP_ccw="hwup-ccw"
SCRIPTUP_scsi_host="hwup-zfcp"
SCRIPTDOWN="hwdown-scsi"
SCRIPTDOWN_scsi="hwdown-zfcp"

# Configured zfcp disks
ZFCP_LUNS="
0x5005076300c300aa:0x5051000000000000
0x5005076300c300aa:0x5052000000000000"

sles10:~ #
```

Figure 5-29 ZFCP hardware configuration file

Note that the ZFCP_LUNS parameter is simply a list of the WWPNs and LUN numbers. If the file is based on a copy from another FCP device, be sure to edit the comments so that they refer to the correct FCP device.

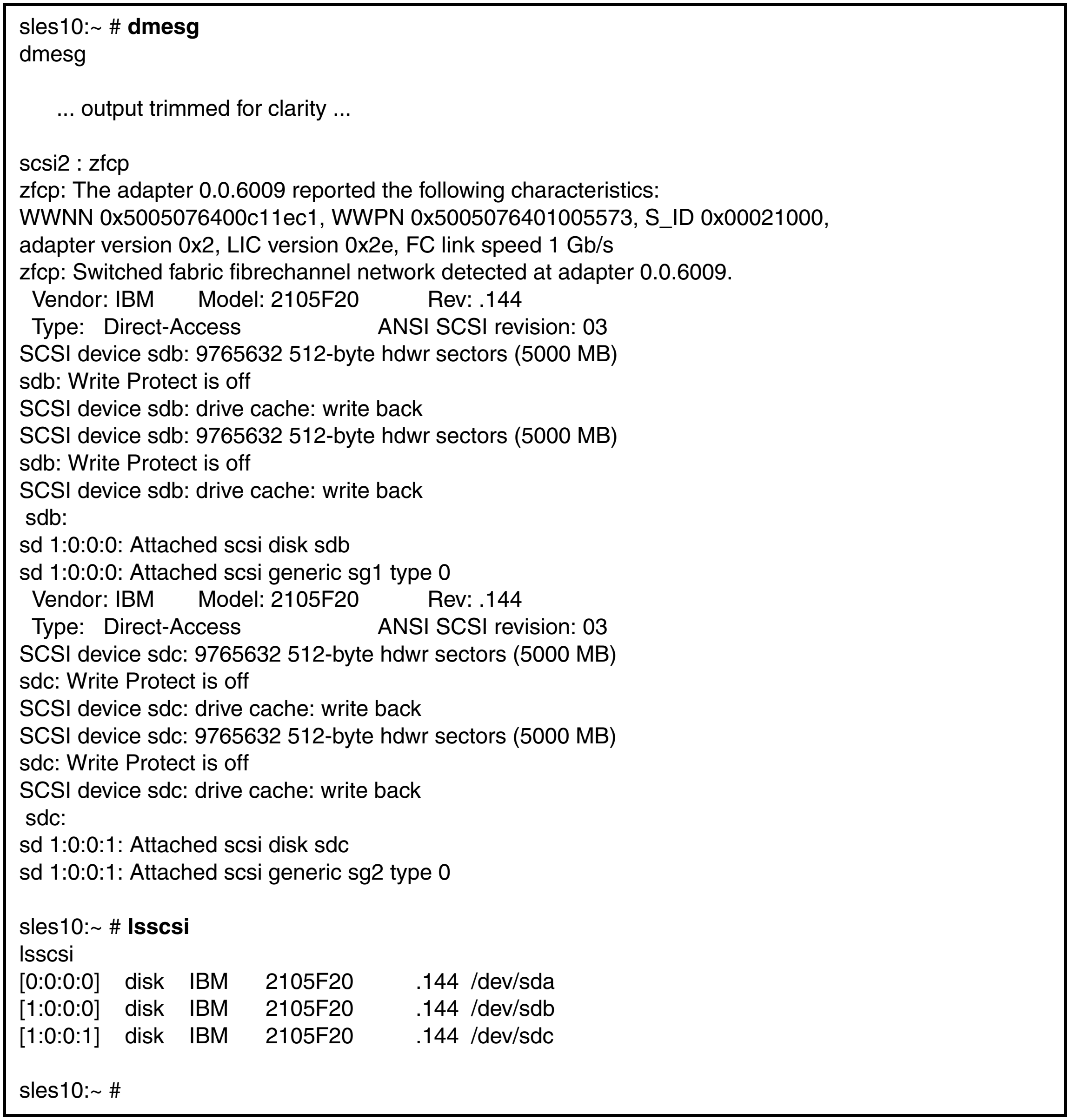
By default, FCP channels are offline. The `chccwdev` command can be used to set an FCP channel online or offline. Alternately, you can write a 1 to a FCP channels online attribute in the /sys file system to set it online or a 0 to set it offline. When the ZFCP hardware configuration file modifications are complete, save the file. The LUNs can be made available immediately by bringing the associated FCP device online using the **chccwdev** command:

```
# chccwdev -e 0.0.6009
```

or by echoing a 1 into the online attribute of the 6009 FCP device in the /sys file system:

```
# echo 1 > /sys/bus/ccw/drivers/zfcp/0.0.6009/online
```

Verify that the LUNs were detected by the FCP subsystem by checking the output of the `dmesg` command or `lsscsi` command, as shown in Figure 5-30.



```
sles10:~ # dmesg
dmesg

    ... output trimmed for clarity ...

scsi2 : zfcp
zfcp: The adapter 0.0.6009 reported the following characteristics:
WWNN 0x5005076400c11ec1, WWPN 0x5005076401005573, S_ID 0x00021000,
adapter version 0x2, LIC version 0x2e, FC link speed 1 Gb/s
zfcp: Switched fabric fibrechannel network detected at adapter 0.0.6009.
  Vendor: IBM      Model: 2105F20          Rev: .144
  Type:   Direct-Access                    ANSI SCSI revision: 03
SCSI device sdb: 9765632 512-byte hdwr sectors (5000 MB)
sdb: Write Protect is off
SCSI device sdb: drive cache: write back
SCSI device sdb: 9765632 512-byte hdwr sectors (5000 MB)
sdb: Write Protect is off
SCSI device sdb: drive cache: write back
 sdb:
sd 1:0:0:0: Attached scsi disk sdb
sd 1:0:0:0: Attached scsi generic sg1 type 0
  Vendor: IBM      Model: 2105F20          Rev: .144
  Type:   Direct-Access                    ANSI SCSI revision: 03
SCSI device sdc: 9765632 512-byte hdwr sectors (5000 MB)
sdc: Write Protect is off
SCSI device sdc: drive cache: write back
SCSI device sdc: 9765632 512-byte hdwr sectors (5000 MB)
sdc: Write Protect is off
SCSI device sdc: drive cache: write back
 sdc:
sd 1:0:0:1: Attached scsi disk sdc
sd 1:0:0:1: Attached scsi generic sg2 type 0

sles10:~ # lsscsi
lsscsi
[0:0:0:0]   disk   IBM    2105F20        .144  /dev/sda
[1:0:0:0]   disk   IBM    2105F20        .144  /dev/sdb
[1:0:0:1]   disk   IBM    2105F20        .144  /dev/sdc

sles10:~ #
```

Figure 5-30 New device messages

The new LUNs, /dev/sdb and /dev/sdc, can now be partitioned for use with **`fdisk`**.

Caution: At this point the new FCP devices have been added to the appropriate configuration files and made available for use. However, the new LUNs have not been added to the Linux initial ram disk that is loaded at boot time. Prior to shutdown, **`mkinitrd`** and **`zipl`** should be run in order to rebuild the initial ram disk and update the Linux boot record. The required commands and their output are shown in Figure 5-27 on page 79.

Add LUNs to an existing FCP device behind a new WWPN

Adding LUNs behind a new WWPN on an FCP device that is already in use requires the following steps:

1. Edit the hardware configuration file for the FCP device.
2. Echo the WWPN to the port_add attribute in /sys for the FCP device.
3. Echo the LUN to the unit_add attribute in /sys for the FCP device.

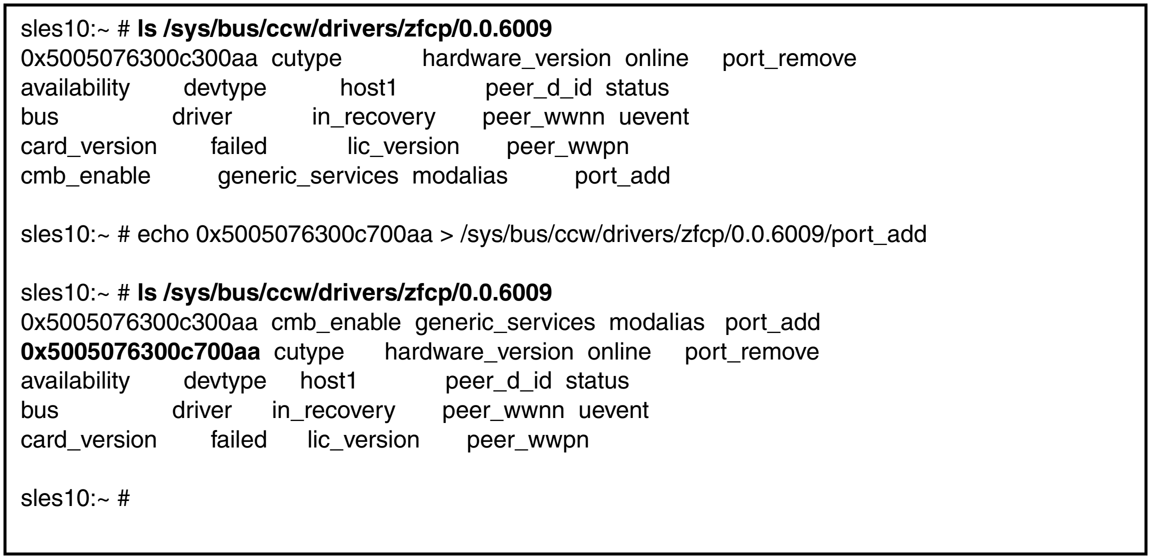
Do the following:

1. Edit the hardware configuration file in /etc/sysconfig/hardware to reflect the new WWPN and LUN being added. Figure 5-31 shows that WWPN 0x5005076300c700aa and LUN number 0x5053000000000000 have been added to the existing FCP device 6009.

```
sles10:~ # cat /etc/sysconfig/hardware/hwcfg-zfcp-bus-ccw-0.0.6009
#!/bin/sh
#
# hwcfg-zfcp-bus-ccw-0.0.6009
#
# Configuration for the zfcp adapter at CCW ID 0.0.6009
#

STARTMODE="auto"
MODULE="zfcp"
MODULE_OPTIONS=""
MODULE_UNLOAD="yes"

# Scripts to be called for the various events.
# If called manually the event is set to 'up'.
SCRIPTUP="hwup-ccw"
SCRIPTUP_ccw="hwup-ccw"
SCRIPTUP_scsi_host="hwup-zfcp"
SCRIPTDOWN="hwdown-scsi"
SCRIPTDOWN_scsi="hwdown-zfcp"

# Configured zfcp disks
ZFCP_LUNS="
0x5005076300c300aa:0x5051000000000000
0x5005076300c300aa:0x5052000000000000
0x5005076300c700aa:0x5053000000000000"

sles10:~ #
```

Figure 5-31 ZFCP hardware configuration file

2. After the file has been edited, **echo** the WWPN into the port_add attribute in the FCP device's /sys directory. Figure 5-32 shows the before and after contents of the FCP devices /sys file system entry.

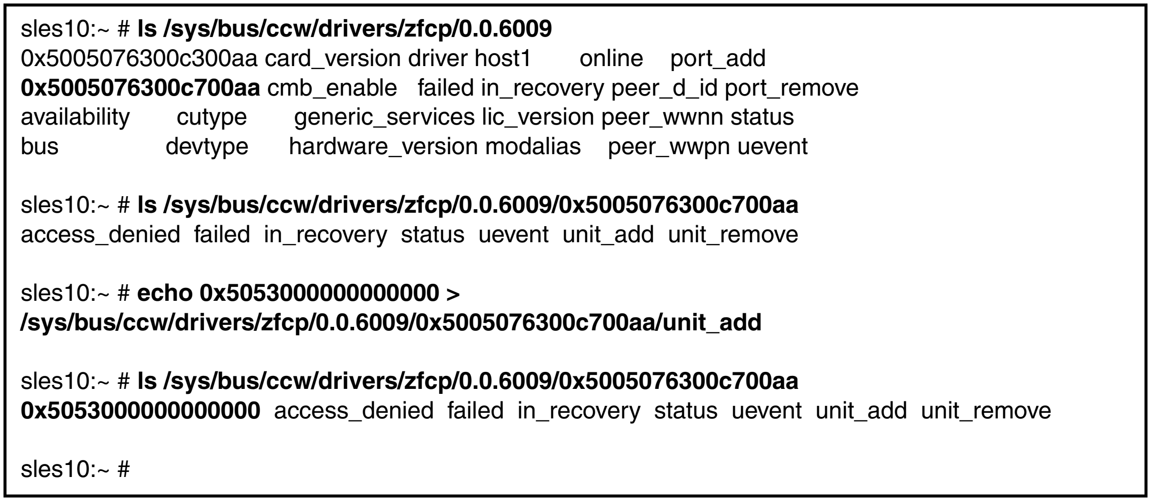


```
sles10:~ # ls /sys/bus/ccw/drivers/zfcp/0.0.6009
0x5005076300c300aa  cutype           hardware_version  online     port_remove
availability       devtype          host1            peer_d_id  status
bus               driver          in_recovery      peer_wwnn  uevent
card_version       failed           lic_version      peer_wwpn
cmb_enable         generic_services  modalias         port_add

sles10:~ # echo 0x5005076300c700aa > /sys/bus/ccw/drivers/zfcp/0.0.6009/port_add

sles10:~ # ls /sys/bus/ccw/drivers/zfcp/0.0.6009
0x5005076300c300aa  cmb_enable  generic_services  modalias   port_add
0x5005076300c700aa  cutype     hardware_version  online     port_remove
availability       devtype     host1            peer_d_id  status
bus               driver     in_recovery      peer_wwnn  uevent
card_version       failed     lic_version      peer_wwpn

sles10:~ #
```

Figure 5-32 The port_add file

Note that WWPNs can also be removed by echoing the WWPN into the port_remove attribute of an FCP devices /sys file system entry.

3. After the new WWPN has been added, **echo** the LUN number into the unit_add attribute in the FCP devices /sys directory. Figure 5-33 shows the before and after contents of the FCP devices /sys file system entry.

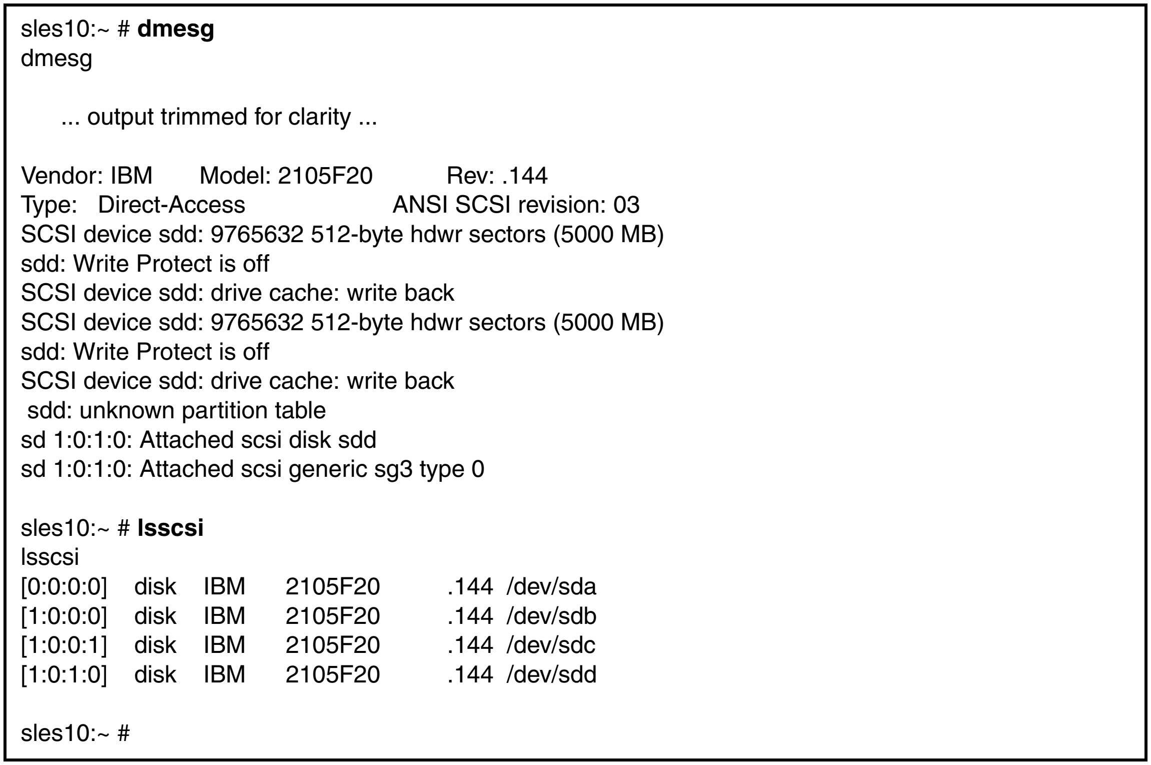


```
sles10:~ # ls /sys/bus/ccw/drivers/zfcp/0.0.6009
0x5005076300c300aa card_version driver host1       online    port_add
0x5005076300c700aa cmb_enable   failed in_recovery peer_d_id port_remove
availability       cutype       generic_services lic_version peer_wwnn status
bus                devtype      hardware_version modalias    peer_wwpn uevent

sles10:~ # ls /sys/bus/ccw/drivers/zfcp/0.0.6009/0x5005076300c700aa
access_denied  failed  in_recovery  status  uevent  unit_add  unit_remove

sles10:~ # echo 0x5053000000000000 >
/sys/bus/ccw/drivers/zfcp/0.0.6009/0x5005076300c700aa/unit_add

sles10:~ # ls /sys/bus/ccw/drivers/zfcp/0.0.6009/0x5005076300c700aa
0x5053000000000000  access_denied  failed  in_recovery  status  uevent  unit_add  unit_remove

sles10:~ #
```

Figure 5-33 The unit_add file

Note that LUNs can also be removed by echoing the LUN number into the unit_remove attribute of an FCP devices /sys file system entry.

4. Verify that the LUN was detected by the FCP subsystem by checking the output of the **dmesg** command or the **lsscsi** command, as shown in Figure 5-34.



```
sles10:~ # dmesg
dmesg

     ... output trimmed for clarity ...

Vendor: IBM      Model: 2105F20          Rev: .144
Type:   Direct-Access                    ANSI SCSI revision: 03
SCSI device sdd: 9765632 512-byte hdwr sectors (5000 MB)
sdd: Write Protect is off
SCSI device sdd: drive cache: write back
SCSI device sdd: 9765632 512-byte hdwr sectors (5000 MB)
sdd: Write Protect is off
SCSI device sdd: drive cache: write back
 sdd: unknown partition table
sd 1:0:1:0: Attached scsi disk sdd
sd 1:0:1:0: Attached scsi generic sg3 type 0

sles10:~ # lsscsi
lsscsi
[0:0:0:0]   disk   IBM    2105F20        .144  /dev/sda
[1:0:0:0]   disk   IBM    2105F20        .144  /dev/sdb
[1:0:0:1]   disk   IBM    2105F20        .144  /dev/sdc
[1:0:1:0]   disk   IBM    2105F20        .144  /dev/sdd

sles10:~ #
```

Figure 5-34 New device messages

The new LUN, /dev/sdd, can now be partitioned for use with **fdisk**.

> **Caution:** The new FCP devices have been added to the appropriate configuration files and made available for use. However, the new LUNs have not been added to the Linux initial ram disk that is loaded at boot time. Prior to shutdown, **mkinitrd** and **zipl** should be run in order to rebuild the initial ram disk and update the Linux boot record. The required commands and their output are shown in Figure 5-27 on page 79.

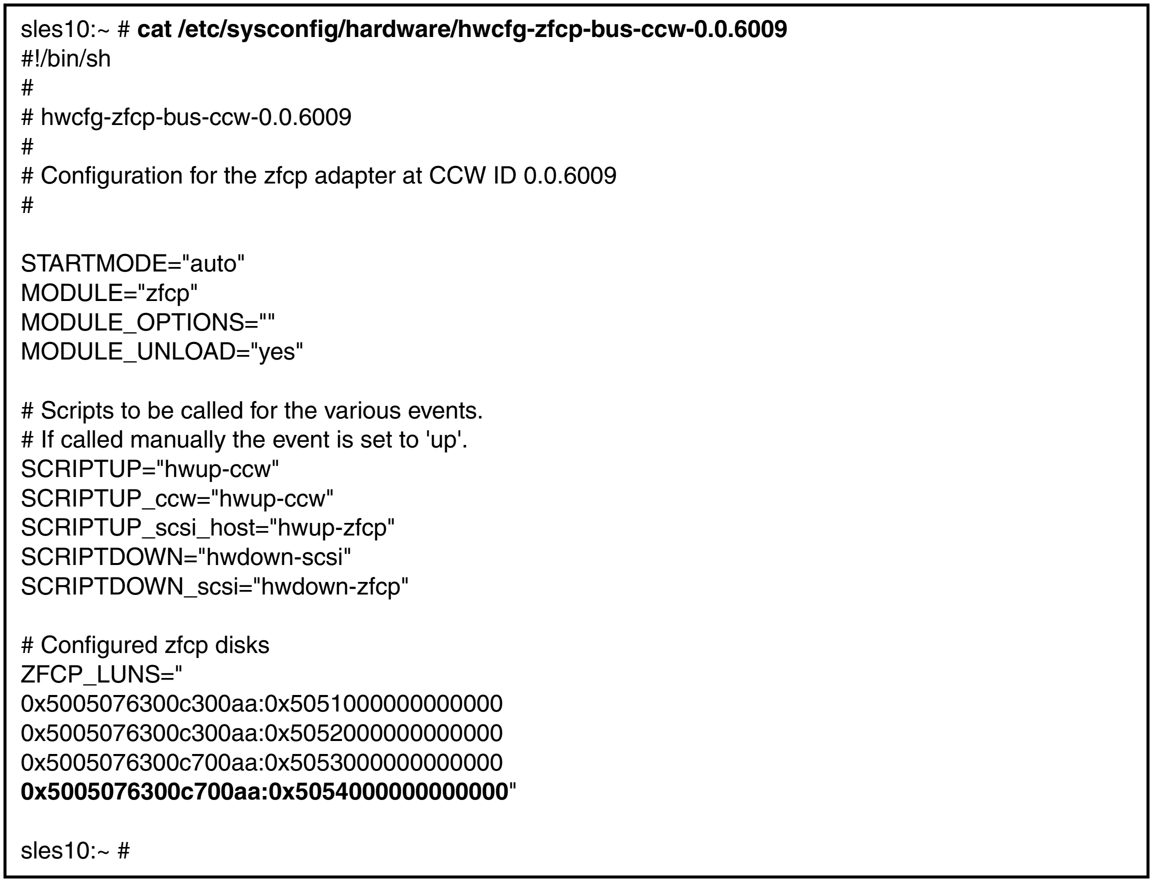
Add LUNs to an existing FCP device behind an existing WWPN

Adding LUNs behind an existing WWPN requires the following steps:

1. Edit the hardware configuration file for the FCP device.
2. Echo the LUN to the unit_add attribute in /sys for the FCP device.

Do the following:

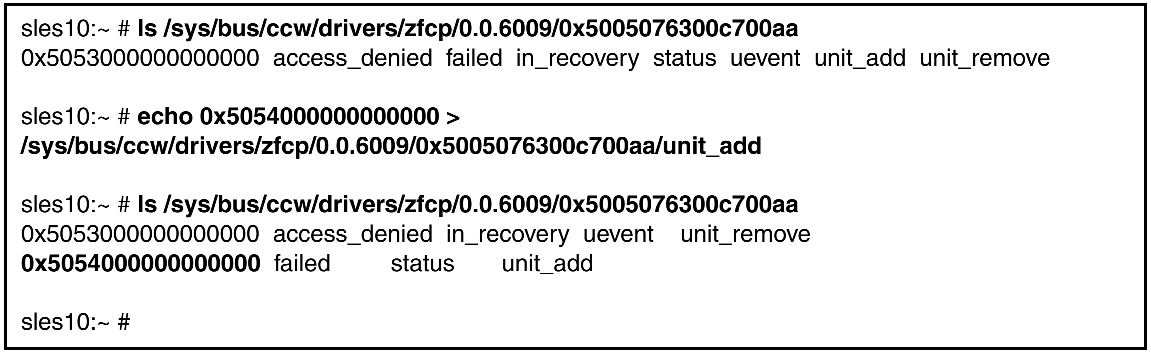
1. Edit the hardware configuration file in /etc/sysconfig/hardware to reflect the new LUN being added. Figure 5-35 shows that LUN number 0x5054000000000000 has been added to FCP device 6009, behind the previously defined WWPN 0x5005076300c700aa.



```
sles10:~ # cat /etc/sysconfig/hardware/hwcfg-zfcp-bus-ccw-0.0.6009
#!/bin/sh
#
# hwcfg-zfcp-bus-ccw-0.0.6009
#
# Configuration for the zfcp adapter at CCW ID 0.0.6009
#

STARTMODE="auto"
MODULE="zfcp"
MODULE_OPTIONS=""
MODULE_UNLOAD="yes"

# Scripts to be called for the various events.
# If called manually the event is set to 'up'.
SCRIPTUP="hwup-ccw"
SCRIPTUP_ccw="hwup-ccw"
SCRIPTUP_scsi_host="hwup-zfcp"
SCRIPTDOWN="hwdown-scsi"
SCRIPTDOWN_scsi="hwdown-zfcp"

# Configured zfcp disks
ZFCP_LUNS="
0x5005076300c300aa:0x5051000000000000
0x5005076300c300aa:0x5052000000000000
0x5005076300c700aa:0x5053000000000000
0x5005076300c700aa:0x5054000000000000"

sles10:~ #
```

Figure 5-35 ZFCP hardware configuration file

2. After the hardware file has been edited, `echo` the LUN number into the unit_add attribute in the FCP devices /sys directory. Figure 5-36 shows the before and after contents of the FCP devices /sys file system entry.

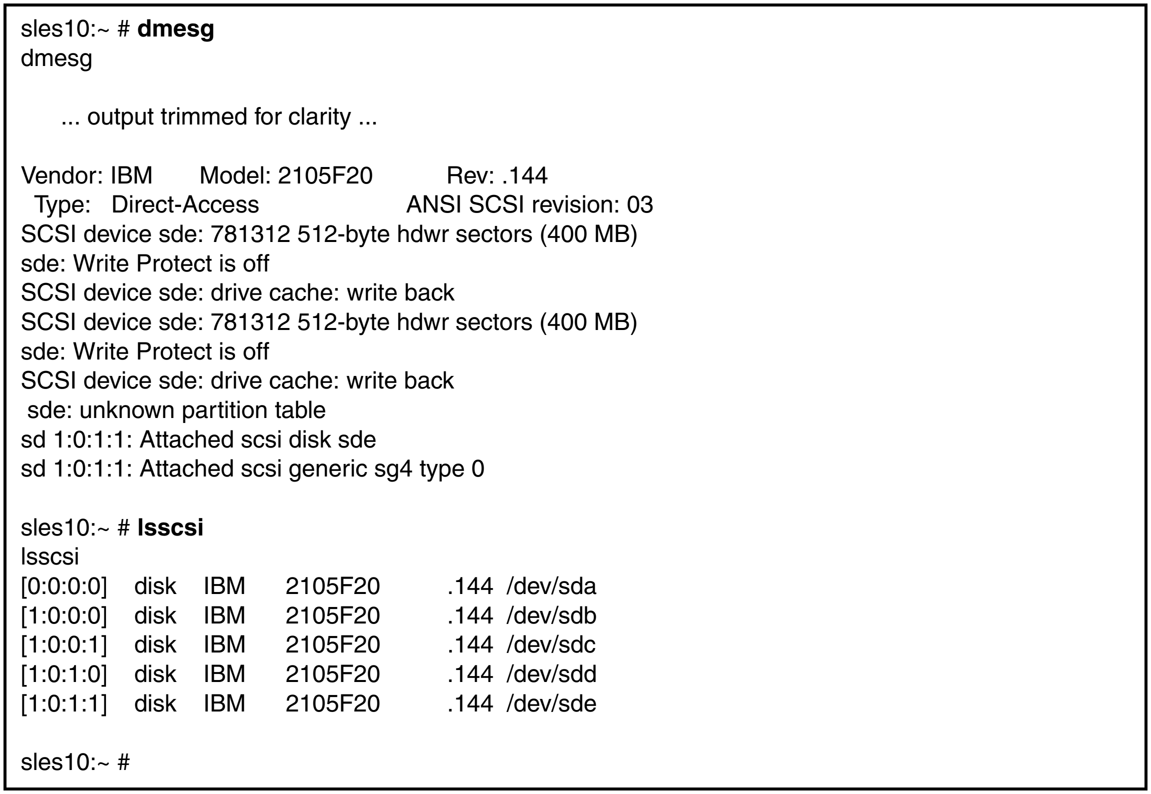


```
sles10:~ # ls /sys/bus/ccw/drivers/zfcp/0.0.6009/0x5005076300c700aa
0x5053000000000000  access_denied  failed  in_recovery  status  uevent  unit_add  unit_remove

sles10:~ # echo 0x5054000000000000 >
/sys/bus/ccw/drivers/zfcp/0.0.6009/0x5005076300c700aa/unit_add

sles10:~ # ls /sys/bus/ccw/drivers/zfcp/0.0.6009/0x5005076300c700aa
0x5053000000000000  access_denied  in_recovery  uevent    unit_remove
0x5054000000000000  failed         status       unit_add

sles10:~ #
```

Figure 5-36 The unit_add interface

Note that LUNs can also be removed by echoing the LUN number into the unit_remove attribute of an FCP device’s /sys file system entry.

3. Verify that the LUN was detected by the FCP subsystem by checking the output of the **dmesg** command or the **lsscsi** command, as shown in Figure 5-37.



```
sles10:~ # dmesg
dmesg

    ... output trimmed for clarity ...

Vendor: IBM      Model: 2105F20          Rev: .144
  Type:   Direct-Access                   ANSI SCSI revision: 03
SCSI device sde: 781312 512-byte hdwr sectors (400 MB)
sde: Write Protect is off
SCSI device sde: drive cache: write back
SCSI device sde: 781312 512-byte hdwr sectors (400 MB)
sde: Write Protect is off
SCSI device sde: drive cache: write back
 sde: unknown partition table
sd 1:0:1:1: Attached scsi disk sde
sd 1:0:1:1: Attached scsi generic sg4 type 0

sles10:~ # lsscsi
lsscsi
[0:0:0:0]   disk   IBM     2105F20        .144  /dev/sda
[1:0:0:0]   disk   IBM     2105F20        .144  /dev/sdb
[1:0:0:1]   disk   IBM     2105F20        .144  /dev/sdc
[1:0:1:0]   disk   IBM     2105F20        .144  /dev/sdd
[1:0:1:1]   disk   IBM     2105F20        .144  /dev/sde

sles10:~ #
```

Figure 5-37 New device messages

The new LUN, /dev/sde, can now be partitioned for use with **fdisk**.

> **Caution:** At this point the new FCP devices have been added to the appropriate configuration files and made available for use. However, the new LUNs have not been added to the Linux initial ram disk that is loaded at boot time. Prior to shutdown, **mkinitrd** and **zipl** should be run in order to rebuild the initial ram disk and update the Linux boot record. The required commands and their output are shown in Figure 5-27 on page 79.

6

FCP multipathing on SLES10

This chapter provides details for configuring FCP multipathing on SLES10. Topics include:

- Why implement multipathing for SCSI disks
- FCP multipath configuration utility options for SLES10
- Configuring FCP multipathing using multipath tools
- Creating a logical volume from multipathed FCP disks

6.1 Why implement multipathing for SCSI disks

In general, there are two reasons for establishing multiple paths to a device:

- High availability - Providing several physical paths to a device offers high availability in a failure scenario. If one path fails, other paths to the device are still available and processing can continue interrupted.
- High performance - Using multiple paths simultaneously to read from or to write to a device can significantly increase I/O performance.

Redundant paths for SCSI LUNs defined to Linux appear as two or more devices, each pointing to the same LUN. A diagram of the environment used in writing this chapter is depicted in Figure 6-1.

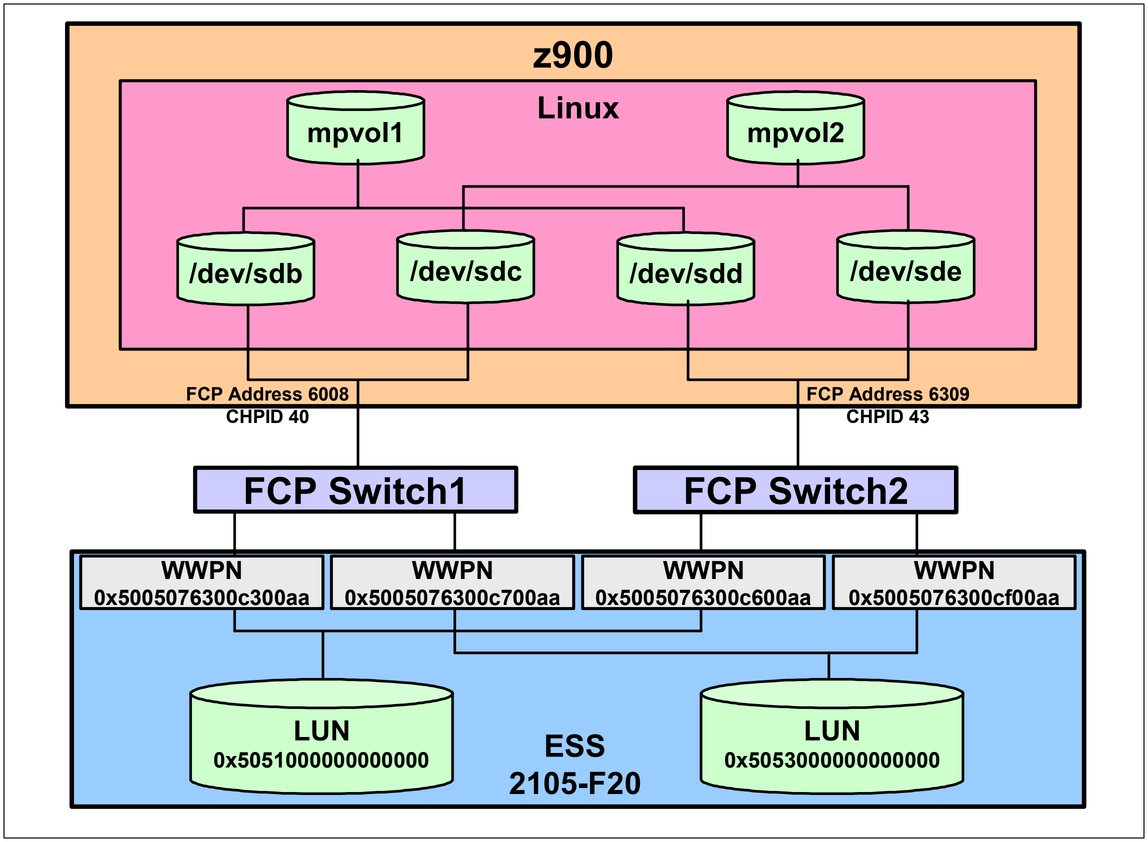


Figure 6-1 Multipath environment

As can be seen in Figure 6-1, two multipathed LUNs are defined on IBM 2105-F20 Enterprise Storage Subsystem (ESS). mpvol1 consists of /dev/sdb and

/dev/sdd, which each represent a separate path to LUN 0x5051000000000000. mpvol2 consists of /dev/sdc and /dev/sde, which each represent a separate path to LUN 0x5053000000000000. Each path is defined as an FCP device address/WWPN/LUN triplet, as follows:

```
/dev/sdb = 6008/0x5005076300c300aa/0x505100000000
/dev/sdc = 6008/0x5005076300c700aa/0x505300000000
/dev/sdd = 6309/0x5005076300c600aa/0x505100000000
/dev/sde = 6309/0x5005076300cf00aa/0x505300000000
```

Note that the two paths to each LUN are totally unique: two separate FCP device addresses on two separate FCP CHPIDs; two separate fibre switches; and four separate Host Bus Adapters (HBAs), each with its own WWPN. This configuration totally isolates each of the paths to a LUN and eliminates any single point of failure. Using only two separate HBAs or WWPNs would still have maintained the unique paths required to eliminate any single point of failure, but exploiting the four HBAs that were available allows the I/O to be spread across four WWPNs versus two, which can improve throughput depending on workload characteristics.

6.2 FCP multipath configuration utility options for SLES10

Unlike traditional Extended Count Key Data (ECKD) mainframe DASD, where multiple paths to individual devices is handled by the IOCP configuration and hardware itself, FCP multipathing requires software to manage multiple paths to a device. There are currently two options available in SLES10 to configure FCP device multipathing:

- Multipath tools - user space tools included with the SLES10 distribution. This package provides a command-line interface to manage multipathed devices with the device-mapper. Logical Volume Manager 2 (LVM2) can also be employed to create multipathed logical volumes for greater versatility and future expendability. For more information see:

  http://christophe.varoqui.free.fr/wiki/wakka.php?wiki=Home

- Enterprise Volume Management System (EVMS) - a storage management utility that provides a single framework for managing and administering system storage devices. It is included with the SLES10 distribution and provides command-line, ncurses text mode, and full GUI interfaces. For more information see:

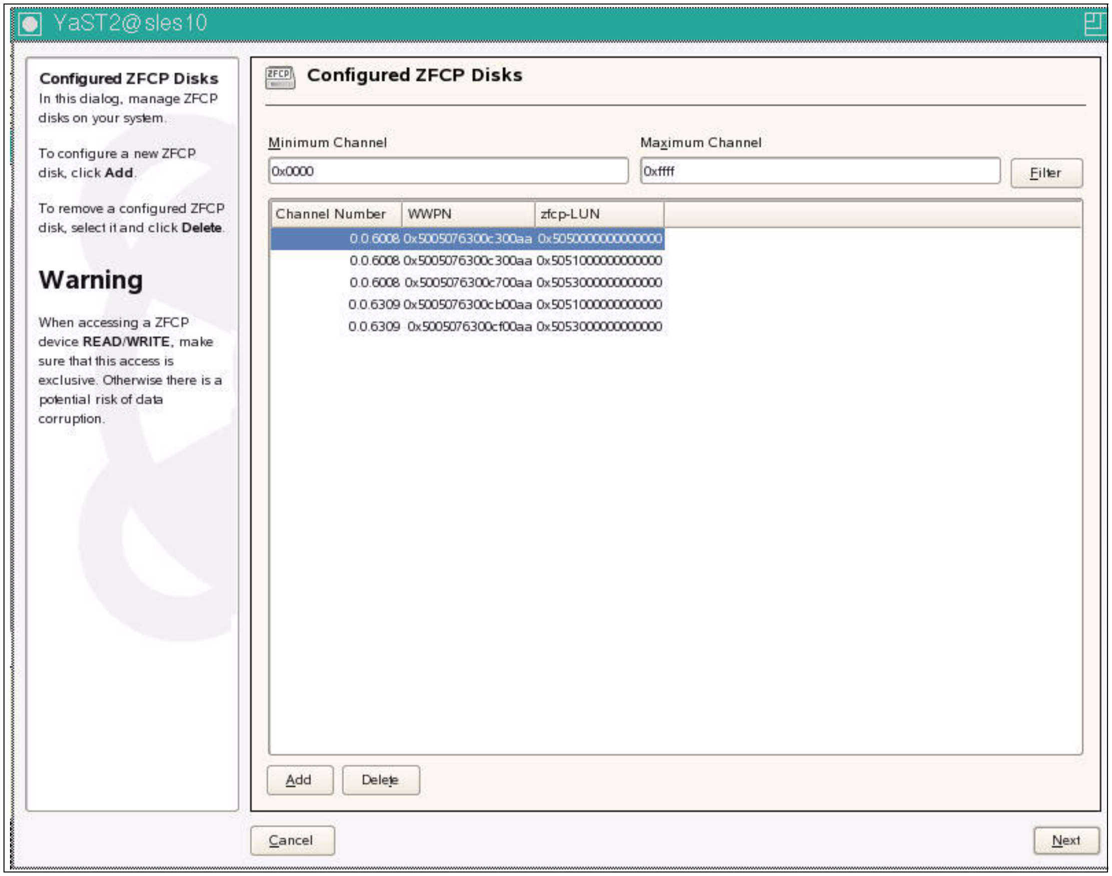
  http://evms.sourceforge.net/

Both utilities were employed during the writing of this book and both utilities are currently supported. However, it was decided to include only the multipath tools utility and LVM2 in this chapter for several reasons. First, EVMS terminology is very different from that of LVM and LVM2. It takes some time for most users to get comfortable with. This is particularly true for users who have previous experience with LVM or LVM2. Second, use of a Logical Volume Manager is mutually exclusive. LVM is the most commonly used Logical Volume Manager, and, in fact, is loaded by default in SLES9 and SLES10. Systems currently exploiting LVM would need to convert their existing logical volumes to EVMS managed volumes. Third, testing and measurement results in the lab have shown that in failover and multibus mode, EVMS-created logical volumes show significantly higher CPU costs than LVM2 created logical volumes. In single path mode, EVMS performs well with one process. With more than one process the difference in CPU costs is less significant. Measurement details can be found here:

http://www.ibm.com/developerworks/linux/linux390/perf/tuning_res_dasd_multipath.html

6.3 Configuring FCP multipathing using multipath tools

In this section, configuring FCP multipathing using device-mapper and multipath tools on a SLES10 Linux guest of VM is illustrated. Figure 6-3 on page 98 shows the YaST Configured ZFCP Disks panel displaying the channel number, WWPN, and FCP_LUN of the ZFCP disks that have been defined to the Linux guest.



Configured ZFCP Disks
In this dialog, manage ZFCP disks on your system.

To configure a new ZFCP disk, click **Add**.

To remove a configured ZFCP disk, select it and click **Delete**.

Warning

When accessing a ZFCP device **READ/WRITE**, make sure that this access is exclusive. Otherwise there is a potential risk of data corruption.

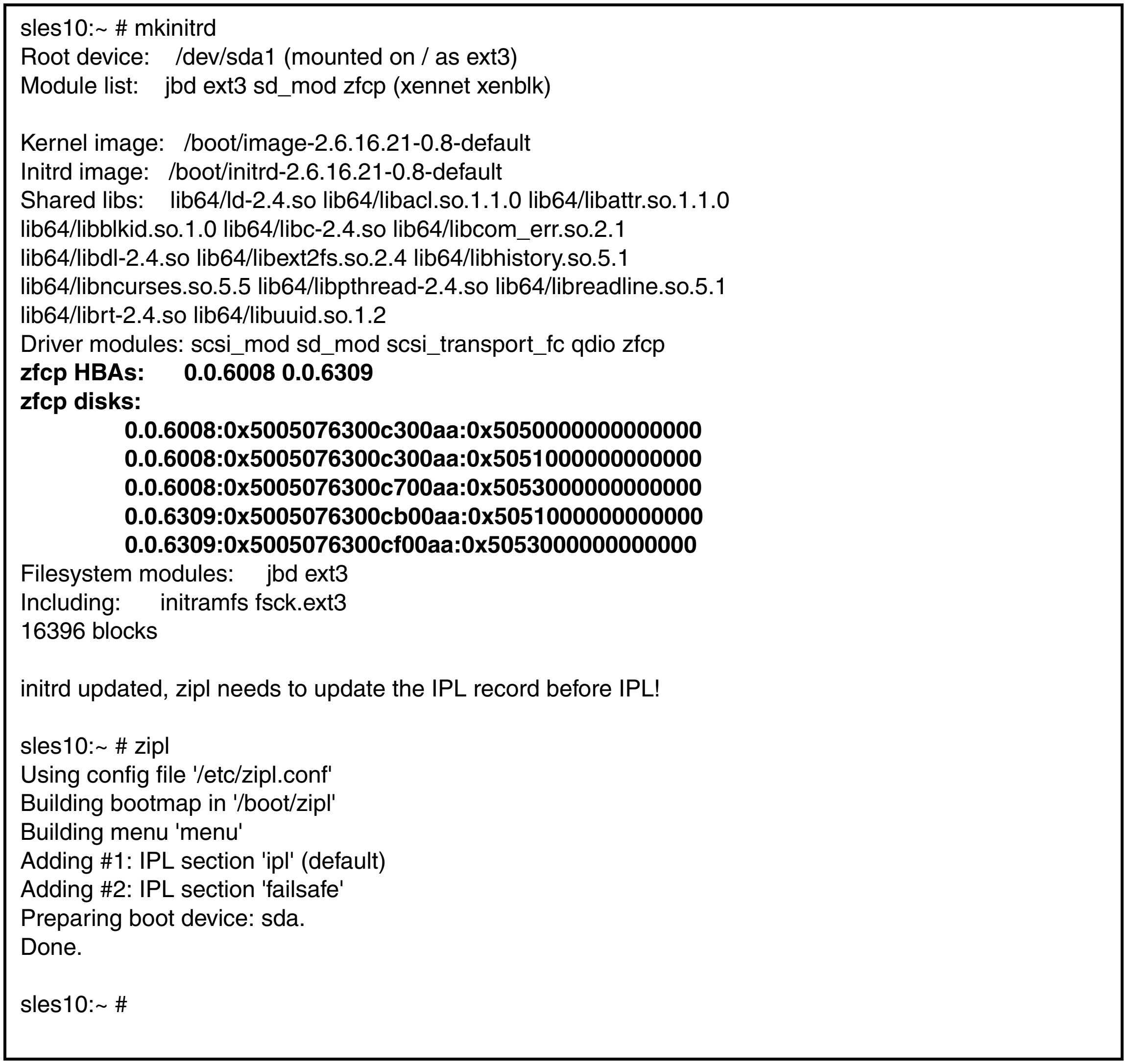
Minimum Channel: 0x0000
Maximum Channel: 0xffff

Channel Number	WWPN	zfcp-LUN
0.0.6008	0x5005076300c300aa	0x5050000000000000
0.0.6008	0x5005076300c300aa	0x5051000000000000
0.0.6008	0x5005076300c700aa	0x5053000000000000
0.0.6309	0x5005076300cb00aa	0x5051000000000000
0.0.6309	0x5005076300cf00aa	0x5053000000000000

Figure 6-2 YaST Configured ZFCP Disks panel

LUN 0x505000000000 is the boot device for this Linux system that was installed in a previous chapter. Note that two devices point to LUN 0x5051000000000000 and two devices point to LUN 0x5053000000000000, as per the configuration diagram in Figure 6-1 on page 94. These FCP devices can be added to the system using the YaST GUI or manually.

After the FCP devices have been added the Linux system, they are available for use. However, the new LUNs have not been added to the Linux initial ram disk that is loaded at boot time. In order for the LUNs to be made available at boot time, the initial ram disk needs to be rebuilt and the Linux boot record updated. This is accomplished with the **mkinitrd** and **zipl** commands. The commands and their output are displayed in Figure 6-3.

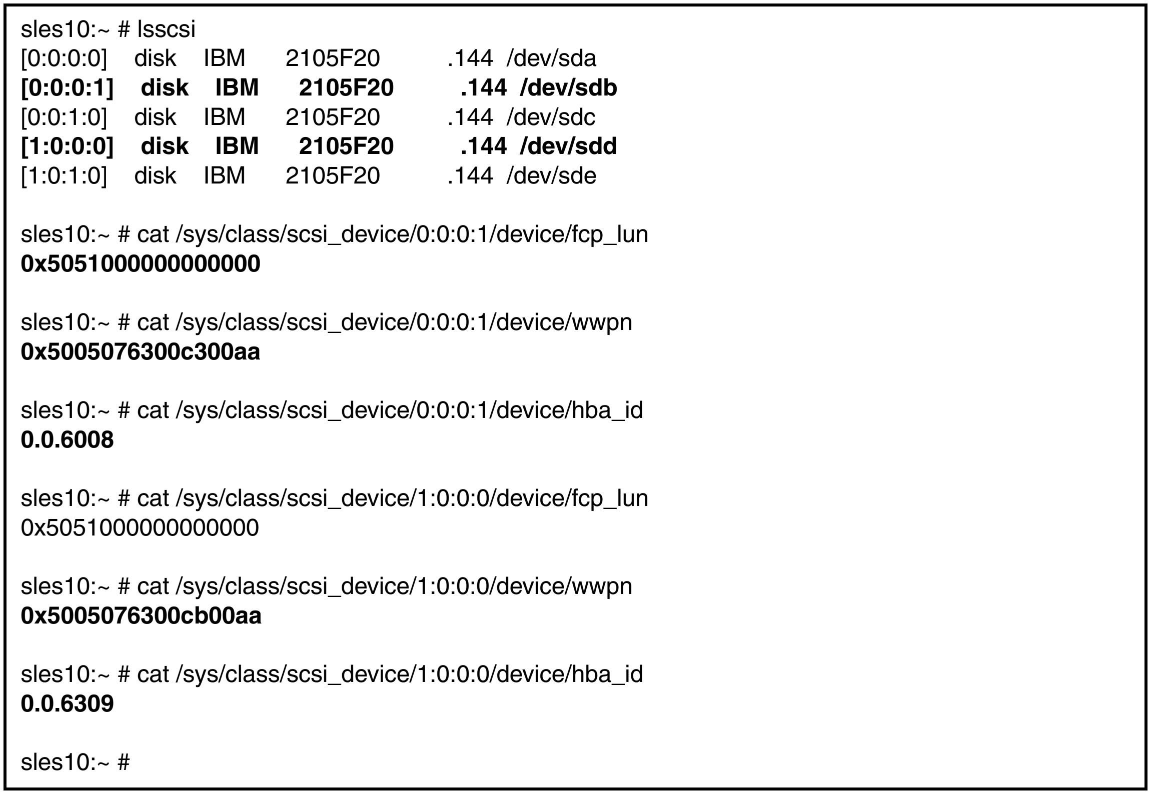


```
sles10:~ # mkinitrd
Root device:    /dev/sda1 (mounted on / as ext3)
Module list:    jbd ext3 sd_mod zfcp (xennet xenblk)

Kernel image:   /boot/image-2.6.16.21-0.8-default
Initrd image:   /boot/initrd-2.6.16.21-0.8-default
Shared libs:    lib64/ld-2.4.so lib64/libacl.so.1.1.0 lib64/libattr.so.1.1.0
lib64/libblkid.so.1.0 lib64/libc-2.4.so lib64/libcom_err.so.2.1
lib64/libdl-2.4.so lib64/libext2fs.so.2.4 lib64/libhistory.so.5.1
lib64/libncurses.so.5.5 lib64/libpthread-2.4.so lib64/libreadline.so.5.1
lib64/librt-2.4.so lib64/libuuid.so.1.2
Driver modules: scsi_mod sd_mod scsi_transport_fc qdio zfcp
zfcp HBAs:      0.0.6008 0.0.6309
zfcp disks:
            0.0.6008:0x5005076300c300aa:0x5050000000000000
            0.0.6008:0x5005076300c300aa:0x5051000000000000
            0.0.6008:0x5005076300c700aa:0x5053000000000000
            0.0.6309:0x5005076300cb00aa:0x5051000000000000
            0.0.6309:0x5005076300cf00aa:0x5053000000000000
Filesystem modules:     jbd ext3
Including:      initramfs fsck.ext3
16396 blocks

initrd updated, zipl needs to update the IPL record before IPL!

sles10:~ # zipl
Using config file '/etc/zipl.conf'
Building bootmap in '/boot/zipl'
Building menu 'menu'
Adding #1: IPL section 'ipl' (default)
Adding #2: IPL section 'failsafe'
Preparing boot device: sda.
Done.

sles10:~ #
```

Figure 6-3 mkinitrd and zipl command output

The **lsscsi command** can be used to verify device availability and responds with a list of active SCSI devices found in the 2.6 Linux kernels /sys file system, as discussed in Chapter 4, “FCP devices in the 2.6 Linux kernel” on page 31.

It is possible to verify which target LUN, WWPN, and FCP device address are associated with any given block device name displayed in the **lsscsi** command output by using the **cat** command to display the SCSI device attributes in the /sys file system. In Figure 6-4, the **cat** command is used to verify that SCSI device [0:0:0:1] or /dev/sdb and SCSI device [1:0:0:0] or /dev/sdd point to the same LUN (0x5051000000000000), using a different WWPN and FCP device address. Although not shown here, SCSI device [0:0:1:0] or /dev/sdc and SCSI device [1:0:1:0] or /dev/sde also point to the same LUN (0x5053000000000000).



```
sles10:~ # lsscsi
[0:0:0:0]   disk   IBM    2105F20       .144  /dev/sda
[0:0:0:1]   disk   IBM    2105F20       .144  /dev/sdb
[0:0:1:0]   disk   IBM    2105F20       .144  /dev/sdc
[1:0:0:0]   disk   IBM    2105F20       .144  /dev/sdd
[1:0:1:0]   disk   IBM    2105F20       .144  /dev/sde

sles10:~ # cat /sys/class/scsi_device/0:0:0:1/device/fcp_lun
0x5051000000000000

sles10:~ # cat /sys/class/scsi_device/0:0:0:1/device/wwpn
0x5005076300c300aa

sles10:~ # cat /sys/class/scsi_device/0:0:0:1/device/hba_id
0.0.6008

sles10:~ # cat /sys/class/scsi_device/1:0:0:0/device/fcp_lun
0x5051000000000000

sles10:~ # cat /sys/class/scsi_device/1:0:0:0/device/wwpn
0x5005076300cb00aa

sles10:~ # cat /sys/class/scsi_device/1:0:0:0/device/hba_id
0.0.6309

sles10:~ #
```

Figure 6-4 lsscsi command output and contents of SCSI attributes in the sysfs

Linux interprets each path to a SCSI LUN as though it were leading to a separate device. The device-mapper and multipath device-mapper target modules can detect multiple paths that lead to the same LUN and represent them as a single multipath I/O device (MPIO). MPIO devices can be partitioned for use or used as physical volumes for LVM or software RAID. The user-space tool (multipath) detects multiple paths, configure, list, and remove MPIO devices. The multipath daemon (multipathd) monitors paths, tests MPIO devices for path failures, and reactivates failed paths when they become available.

To invoke the user-space tools, both the dm-mod and dm-multipath kernel modules must be loaded. The dm-mod module is loaded by default in SLES10, but the dm-multipath module is not. The dm-multipath module can be loaded with the **modprobe** command, as displayed in Figure 6-5.

Once the required kernel modules are loaded, the user-space tool (multipath) can be invoked. The response from the **multipath** command (also displayed in the Figure 6-5) shows that two new devices were created, each being 4 GB in size. The names of the new devices are a long string of numbers and characters that represent the Universal Unique IDentifier (UUID), as returned by each physical LUN, along with the storage subsystem type (IBM 2105). The response indicates that the first device is composed of the block devices /dev/sdb and /dev/sdd. The second device is composed of the block /dev/sdc and /dev/sde. Note that the response also indicates that a round-robin scheme will be used to distribute I/O requests evenly across the defined paths.

```
sles10:~ # modprobe dm-multipath

sles10:~ # multipath
create: 1IBM_2105_05112320IBM,2105F20
[size=4G][features=1 queue_if_no_path][hwhandler=0]
\_ round-robin 0 [prio=2][undef]
 \_ 0:0:0:1 sdb 8:16  [undef][ready]
 \_ 1:0:0:0 sdd 8:48  [undef][ready]
create: 1IBM_2105_05312320IBM,2105F20
[size=4G][features=1 queue_if_no_path][hwhandler=0]
\_ round-robin 0 [prio=2][undef]
 \_ 0:0:1:0 sdc 8:32  [undef][ready]
 \_ 1:0:1:0 sde 8:64  [undef][ready]

sles10:~ #
```

Figure 6-5 Multipath command output

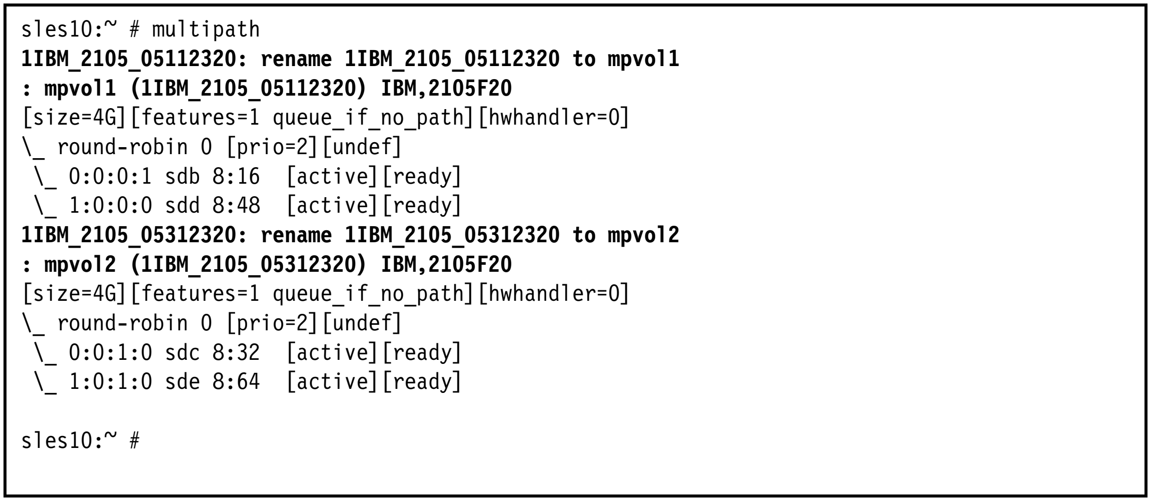
At this point the new multipath FCP devices are ready for use. However, the volumes would have to be referenced by their long UUID string. The /etc/multipath.conf file can be configured to assign aliases to each multipath FCP device, which would allow them to be referenced by meaningful short names. In addition, if the Linux system owns any FBA or ECKD volumes managed by the DASD device driver, multipath will attempt to create multipath devices out of these volumes, which, although not critical, results in error messages. The multipath.conf file can also be configured to prevent attempts to create multipath devices out of DASD disk devices.

The entries required in the /etc/multipath.conf file are shown in Figure 6-6. A multipathed volume (1IBM_2105_05112320) is assigned the alias mpvol1 and a multipathed volume (1IBM_2105_05312320) is assigned the alias mpvol2. The blacklist stanza directs multipath to ignore all DASD devices.

```
sles10:~ # cat /etc/multipath.conf
multipaths {
multipath {
wwid 1IBM_2105_05112320
alias mpvol1
}
multipath {
wwid 1IBM_2105_05312320
alias mpvol2
}
}
blacklist {
devnode "^(dasd)[0-9]*"
}

sles10:~ #
```

Figure 6-6 multipath.conf file

After the /etc/multipath.conf file is created, invoking the **multipath** command causes the device-mapper tables to be rebuilt according to the criteria specified in the updated configuration file. The response from this command, depicted in Figure 6-7, shows that the two FCP multipath devices have been renamed using the aliases mpvol1 and mpvol2.

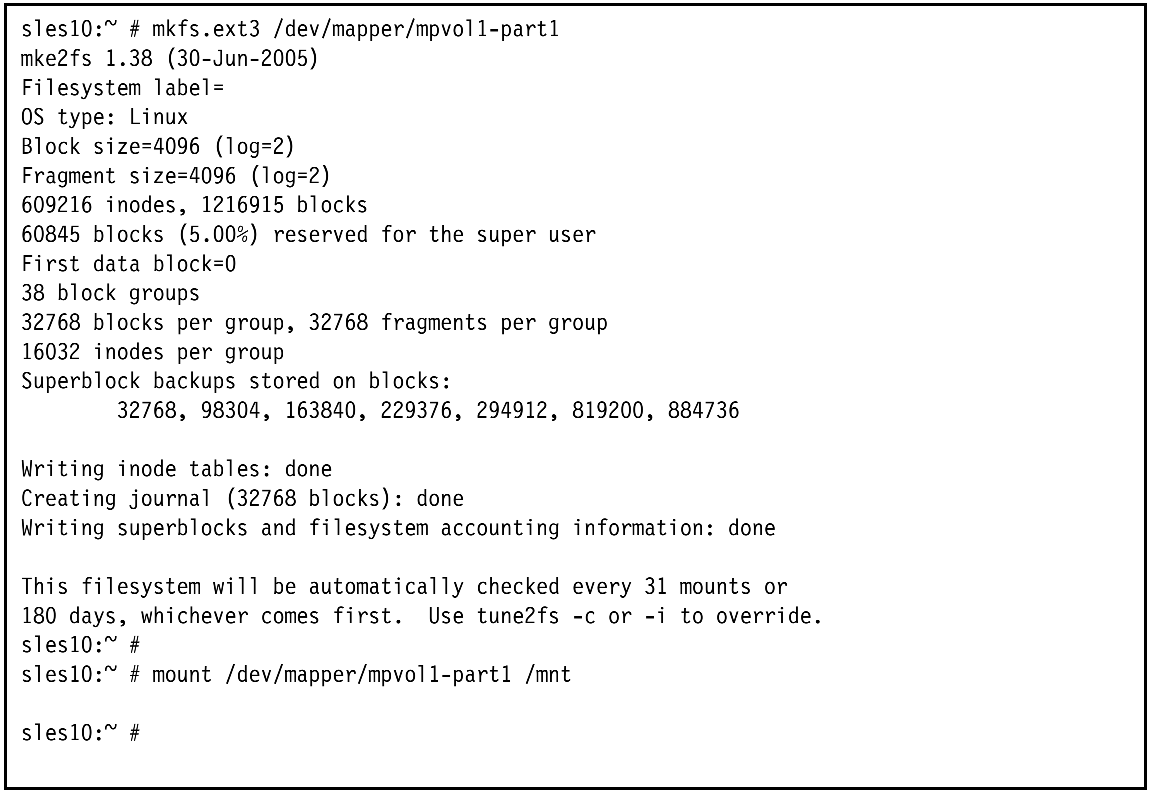


```
sles10:~ # multipath
1IBM_2105_05112320: rename 1IBM_2105_05112320 to mpvol1
: mpvol1 (1IBM_2105_05112320) IBM,2105F20
[size=4G][features=1 queue_if_no_path][hwhandler=0]
\_ round-robin 0 [prio=2][undef]
 \_ 0:0:0:1 sdb 8:16  [active][ready]
 \_ 1:0:0:0 sdd 8:48  [active][ready]
1IBM_2105_05312320: rename 1IBM_2105_05312320 to mpvol2
: mpvol2 (1IBM_2105_05312320) IBM,2105F20
[size=4G][features=1 queue_if_no_path][hwhandler=0]
\_ round-robin 0 [prio=2][undef]
 \_ 0:0:1:0 sdc 8:32  [active][ready]
 \_ 1:0:1:0 sde 8:64  [active][ready]

sles10:~ #
```

Figure 6-7 Multipath command output

The block devices were previously partitioned using **fdisk** to have one primary partition. When device-mapper devices are created, existing partitions are detected, and code known as kpartx is invoked to create device-mapper logical devices for both the base volume and its partitions. kpartx assigns a unique name to the partitions, which is the concatenation of the UUID or alias and the character string -part*x* (where *x* represents the partition number).

Before mpvol1 or mpvol2 can be written to, they must have a file system created on them and mounted. A file system is created on the partition and not the base volume. To create a file system and mount the devices, they must be referenced by their kpartx assigned names. Figure 6-8 shows the results of creating an ext3 file system on /dev/mapper/mpvol1-part1 and then mounting /dev/mapper/mpvol1-part1 at the /mnt mount point. Although not shown here, the same could be done with /dev/mapper/mpvol2-part1 when ready for use.



```
sles10:~ # mkfs.ext3 /dev/mapper/mpvol1-part1
mke2fs 1.38 (30-Jun-2005)
Filesystem label=
OS type: Linux
Block size=4096 (log=2)
Fragment size=4096 (log=2)
609216 inodes, 1216915 blocks
60845 blocks (5.00%) reserved for the super user
First data block=0
38 block groups
32768 blocks per group, 32768 fragments per group
16032 inodes per group
Superblock backups stored on blocks:
        32768, 98304, 163840, 229376, 294912, 819200, 884736

Writing inode tables: done
Creating journal (32768 blocks): done
Writing superblocks and filesystem accounting information: done

This filesystem will be automatically checked every 31 mounts or
180 days, whichever comes first.  Use tune2fs -c or -i to override.
sles10:~ #
sles10:~ # mount /dev/mapper/mpvol1-part1 /mnt

sles10:~ #
```

Figure 6-8 mkfs and mount command output

Figure 6-9 shows the results of running the file system benchmark utility, bonnie, against the multipathed volume mpvol1. /mnt was specified as the target directory with a file size of 1024 M. After the benchmark completes, the output of the Linux **iostat** command can be used to verify that reads and writes to the /mnt directory were spread equally across the two FCP devices (sdb and sdd), as well as the device-mapper devices (dm-1 and dm-3) that make up the multipathed volume mpvol1.

```
sles10:~ # bonnie -d /mnt -s 1024
Bonnie 1.4: File '/mnt/Bonnie.3231', size: 1073741824, volumes: 1
Writing with putc()...        done:   5704 kB/s  95.5 %CPU
Rewriting...                  done:  26505 kB/s  18.3 %CPU
Writing intelligently...      done:  42285 kB/s  45.5 %CPU
Reading with getc()...        done:   5527 kB/s  98.0 %CPU
Reading intelligently...      done:  54421 kB/s  16.8 %CPU
Seeker 1...Seeker 2...Seeker 3...start 'em...done...done...done...
              ---Sequential Output (nosync)--- ---Sequential Input-- --Rnd Seek-
              -Per Char- --Block--- -Rewrite-- -Per Char- --Block--- --04k (03)-
Machine    MB K/sec %CPU K/sec %CPU K/sec %CPU K/sec %CPU K/sec %CPU   /sec %CPU
sles10 1*1024  5704 95.5 42285 45.5 26505 18.3  5527 98.0 54421 16.8 3606.0 28.3

sles10:~ # iostat
Linux 2.6.16.21-0.8-default (sles10)    04/27/07

avg-cpu:  %user   %nice %system %iowait  %steal   %idle
           9.49    0.00    1.77    1.05    0.10   87.59

Device:            tps   Blk_read/s   Blk_wrtn/s   Blk_read   Blk_wrtn
sda               1.86        66.90        19.95     321592      95904
sdb               1.94       635.55       647.02    3054925    3110088
sdc               0.02         0.35         0.00       1677          0
sdd               1.96       691.11       647.18    3321991    3110816
sde               0.02         0.27         0.00       1286          0
dm-1            164.00      1326.12      1294.20    6374360    6220912
dm-2              0.05         0.08         0.00        407          0
dm-3            163.99      1326.08      1294.20    6374152    6220912
dm-4              0.04         0.04         0.00        199          0

sles10:~ #
```

Figure 6-9 bonnie and iostat command output

6.4 Creating a logical volume from multipathed FCP disks

The multipathed volumes created in the previous section are ready for use and at this point could be placed in the Linux File System Table (fstab) to be automatically mounted at boot time. However, as with any other type of disks, many users opt to exploit the use of logical volumes. LVM is the most commonly used volume manager package, and the current version (LVM2) is included with the SLES10 distribution. The primary benefits of using a volume manager is that much larger logical volumes can be created from smaller physical volumes. The use of several conveniently selected disks instead of one single disk can dramatically improve sequential read/write performance and allows the volume to be extended in the future. Only the basic commands required to implement a logical volume with multipathed FCP disks using LVM2 will be covered here. For detailed documentation on LVM visit the following Web site:

http://www.tldp.org/HOWTO/LVM-HOWTO/

Before using multipath devices in a logical volume configuration, changes are required to the lvm configuration file lvm.conf. By default, lvm.conf does not allow lvm to recognize device-mapper created devices, so it would be impossible to create a physical volume out of a multipath device or to create a volume group. The modifications required to the /etc/lvm/lvm.conf file allow device-mapper devices to be used with lvm are shown in Figure 6-10 on page 106.

```
sles10:~ # cat /etc/lvm/lvm.conf
# This is an example configuration file for the LVM2 system.
# It contains the default settings that would be used if there was no
# /etc/lvm/lvm.conf file.
#
# Refer to 'man lvm.conf' for further information including the file layout.

# This section allows you to configure which block devices should
# be used by the LVM system.
devices {

    # Where do you want your volume groups to appear ?
    dir = "/dev"

    # An array of directories that contain the device nodes you wish
    # to use with LVM2.
    scan = [ "/dev", "/dev/mapper" ]

    # A filter that tells LVM2 to only use a restricted set of devices.
    # The filter consists of an array of regular expressions.  These
    # expressions can be delimited by a character of your choice, and
    # prefixed with either an 'a' (for accept) or 'r' (for reject).
    # The first expression found to match a device name determines if
    # the device will be accepted or rejected (ignored).  Devices that
    # don't match any patterns are accepted.

    # Remember to run vgscan after you change this parameter to ensure
    # that the cache file gets regenerated (see below).

    # By default we accept every block device except udev names:
    filter = [ "a|/dev/mapper/.*|", "r|.*|" ]
...
...output trimmed for clarity...
...
    # List of pairs of additional acceptable block device types found
    # in /proc/devices with maximum (non-zero) number of partitions.
    types = [ "device-mapper", 16 ]
...
...output trimmed for clarity...
...

sles10:~ #
```

Figure 6-10   lvm.conf file

Once the updates have been made to the lvm.conf file, multipath I/O devices are recognized by lvm and can be used to create logical volumes. In this section a single logical volume is created from the two FCP multipath I/O devices that were created in the previous section by the device-mapper.

The first step is to designate the two FCP multipath I/O devices as lvm physical volumes using the **pvcreate** command. Note that the devices are referenced by their alias concatenated with their kpartx assigned names, as discussed on page 103.

The next step is to create an lvm volume group consisting of these two physical devices using the **vgcreate** command. In this example a volume group named mympvg is created that consists of the two physical volumes /dev/mapper/mpvol1-part1 and /dev/mapper/mpvol2-part1.

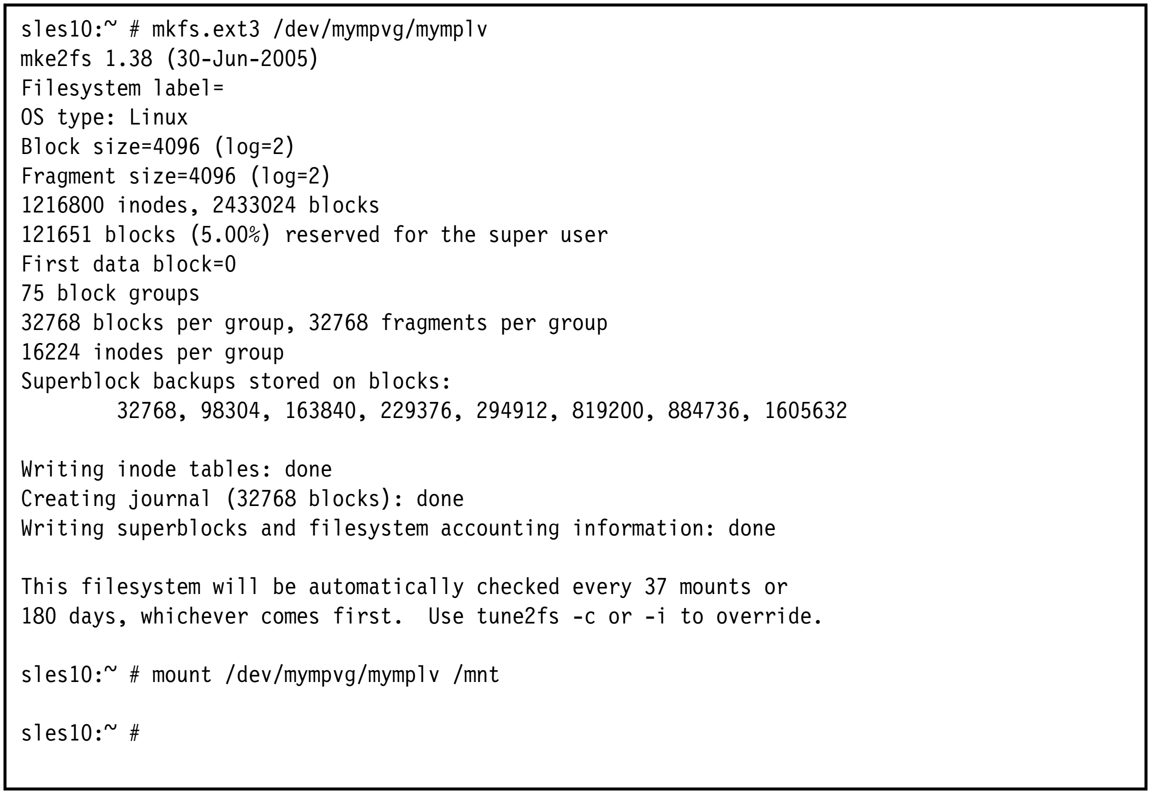
The **vgdisplay** command can be used to identify the characteristics of a volume group, including the allocated and free extent size. One or more logical volumes can be created in a volume group. Since the size of mympvg is relatively small, a single logical volume will be created that consumes all of the space in the volume group using the **lvcreate** command.

Note that in this example a logical volume named mymplv is created. The logical volume is 9.28 GB in size, consuming all free space in the volume group. The logical volume is striped across the two physical volumes to distribute the I/O load across the two multipath I/O devices, using a stripsize of 32. The device-mapper (using the round-robin scheduler) makes sure that I/O is distributed equally across each of the two paths for the underlying FCP LUNs. The lvm commands issued and their responses are depicted in Figure 6-11.

```
sles10:~ # pvcreate /dev/mapper/mpvol1-part1
  Physical volume "/dev/mapper/mpvol1-part1" successfully created

sles10:~ # pvcreate /dev/mapper/mpvol2-part1
  Physical volume "/dev/mapper/mpvol2-part1" successfully created

sles10:~ # vgcreate mympvg /dev/mapper/mpvol[1-2]-part1
  Volume group "mympvg" successfully created

sles10:~ # vgdisplay mympvg
  --- Volume group ---
  VG Name               mympvg
  System ID
  Format                lvm2
  Metadata Areas        2
  Metadata Sequence No  1
  VG Access             read/write
  VG Status             resizable
  MAX LV                0
  Cur LV                0
  Open LV               0
  Max PV                0
  Cur PV                2
  Act PV                2
  VG Size               9.28 GB
  PE Size               4.00 MB
  Total PE              2376
  Alloc PE / Size       0 / 0
  Free  PE / Size       2376 / 9.28 GB
  VG UUID               NTjgbr-zBJY-WMwQ-xOuw-cbP1-Z4mo-U8j5LE

sles10:~ # lvcreate --name mymplv --size 9.28G --stripes 2 --stripesize 32 mympvg
  Rounding up size to full physical extent 9.28 GB
  Logical volume "mymplv" created

sles10:~ #
```

Figure 6-11 lvm commands to create the logical volume

Before the logical volume mymplv can be used, it must have a file system created on it and then mounted. Figure 6-12 shows the commands and response messages received as a result of creating an ext3 file system on /dev/mympvg/mymplv and then mounting /dev/mympvg/mymplv at the /mnt mount point.

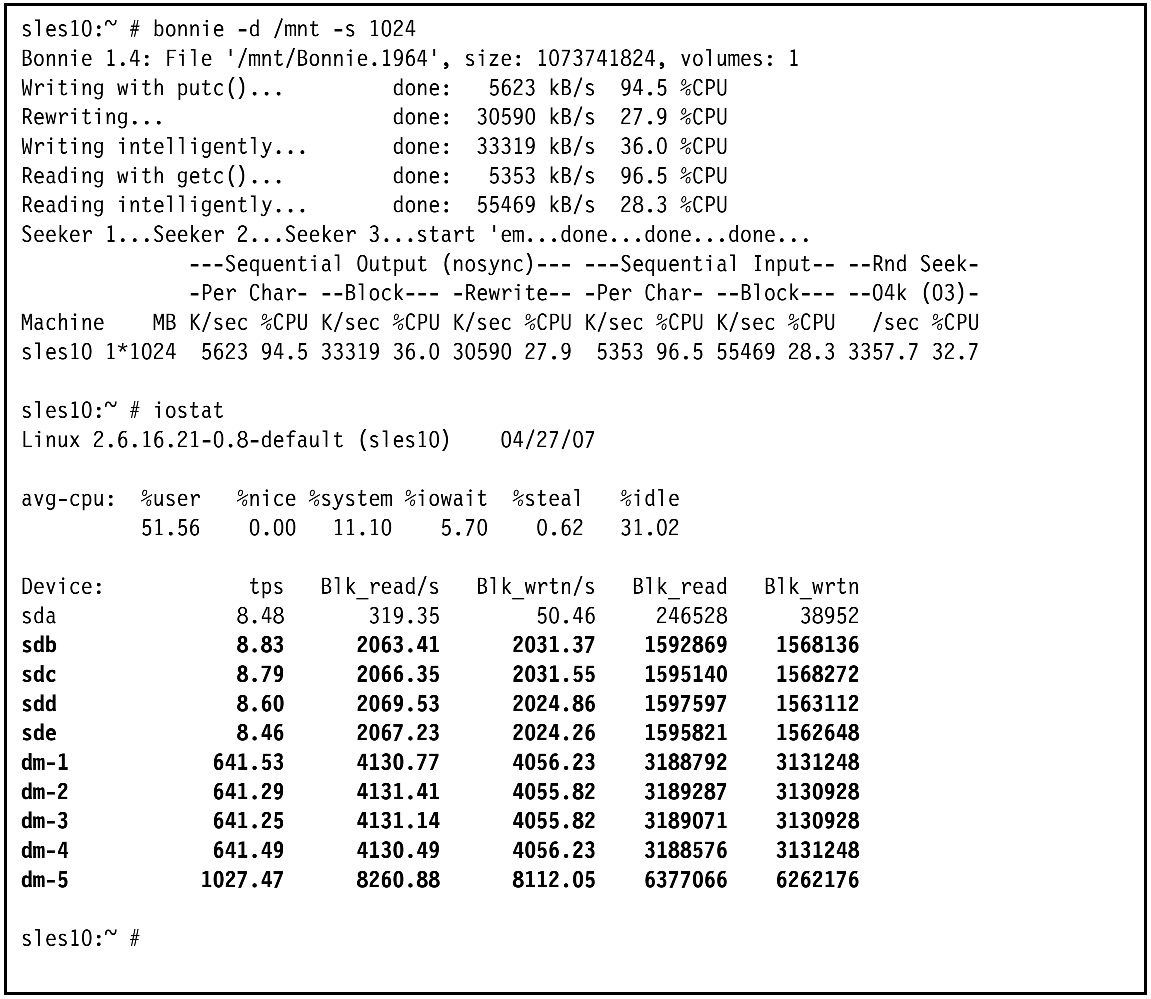


```
sles10:~ # mkfs.ext3 /dev/mympvg/mymplv
mke2fs 1.38 (30-Jun-2005)
Filesystem label=
OS type: Linux
Block size=4096 (log=2)
Fragment size=4096 (log=2)
1216800 inodes, 2433024 blocks
121651 blocks (5.00%) reserved for the super user
First data block=0
75 block groups
32768 blocks per group, 32768 fragments per group
16224 inodes per group
Superblock backups stored on blocks:
        32768, 98304, 163840, 229376, 294912, 819200, 884736, 1605632

Writing inode tables: done
Creating journal (32768 blocks): done
Writing superblocks and filesystem accounting information: done

This filesystem will be automatically checked every 37 mounts or
180 days, whichever comes first.  Use tune2fs -c or -i to override.

sles10:~ # mount /dev/mympvg/mymplv /mnt

sles10:~ #
```

Figure 6-12 mkfs and mount command output

Caution: A logical volume must have a file system created on it before it can be used. If any of the physical volumes included in the volume group from which the logical volume is carved contain file systems, the data in those file systems will be lost when creating a file system on the logical volume.

Figure 6-13 shows the results of running the file system benchmark utility, bonnie, against the logical volume mymplv. /mnt was specified as the target directory with a file size of 1024 M. After the benchmark completes, the output of the Linux **iostat** command can be used to verify that reads and writes to the /mnt directory were spread equally across the four FCP devices (sdb, sdc, sdd, and sde), as well as the device-mapper devices (dm-1 through dm-4) that make up the logical volume mymplv.



```
sles10:~ # bonnie -d /mnt -s 1024
Bonnie 1.4: File '/mnt/Bonnie.1964', size: 1073741824, volumes: 1
Writing with putc()...         done:   5623 kB/s  94.5 %CPU
Rewriting...                   done:  30590 kB/s  27.9 %CPU
Writing intelligently...       done:  33319 kB/s  36.0 %CPU
Reading with getc()...         done:   5353 kB/s  96.5 %CPU
Reading intelligently...       done:  55469 kB/s  28.3 %CPU
Seeker 1...Seeker 2...Seeker 3...start 'em...done...done...done...
              ---Sequential Output (nosync)--- ---Sequential Input-- --Rnd Seek-
              -Per Char- --Block--- -Rewrite-- -Per Char- --Block--- --04k (03)-
Machine    MB K/sec %CPU K/sec %CPU K/sec %CPU K/sec %CPU K/sec %CPU   /sec %CPU
sles10 1*1024  5623 94.5 33319 36.0 30590 27.9  5353 96.5 55469 28.3 3357.7 32.7

sles10:~ # iostat
Linux 2.6.16.21-0.8-default (sles10)    04/27/07

avg-cpu:  %user   %nice %system %iowait  %steal   %idle
          51.56    0.00   11.10    5.70    0.62   31.02

Device:            tps   Blk_read/s   Blk_wrtn/s   Blk_read   Blk_wrtn
sda               8.48       319.35        50.46     246528      38952
sdb               8.83      2063.41      2031.37    1592869    1568136
sdc               8.79      2066.35      2031.55    1595140    1568272
sdd               8.60      2069.53      2024.86    1597597    1563112
sde               8.46      2067.23      2024.26    1595821    1562648
dm-1            641.53      4130.77      4056.23    3188792    3131248
dm-2            641.29      4131.41      4055.82    3189287    3130928
dm-3            641.25      4131.14      4055.82    3189071    3130928
dm-4            641.49      4130.49      4056.23    3188576    3131248
dm-5           1027.47      8260.88      8112.05    6377066    6262176

sles10:~ #
```

Figure 6-13 bonnie and iostat command output

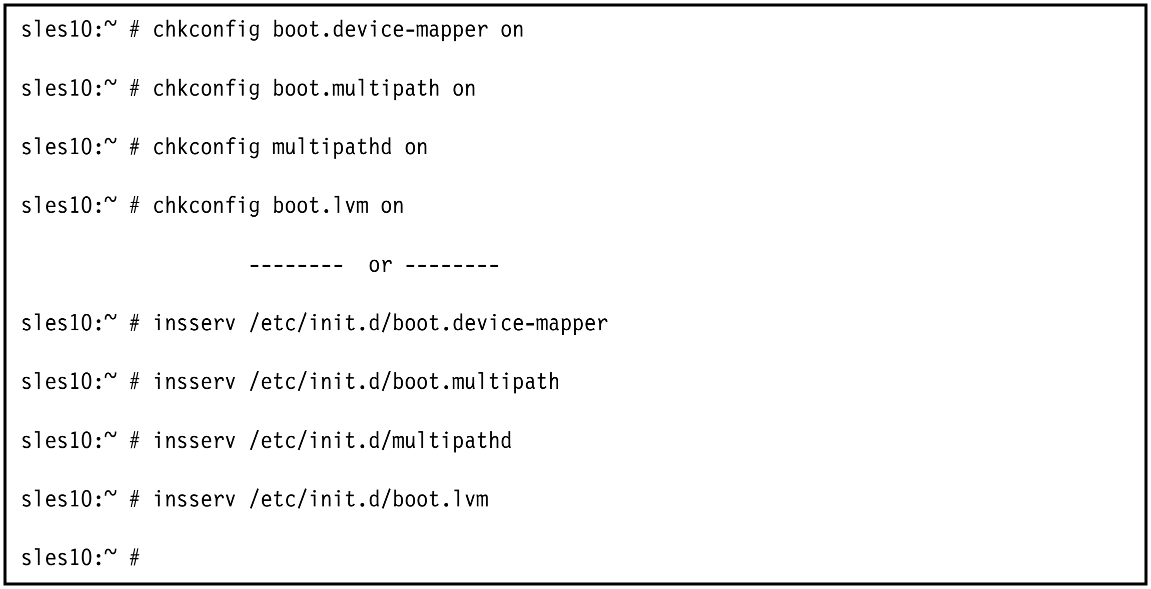
The Linux file system table can be updated to automatically have the logical volume mounted at boot time. The required updates to /etc/fstab are displayed via the **cat** command in Figure 6-14.

```
sles10:~ # cat /etc/fstab
/dev/sda1           /                   ext3       acl,user_xattr     1 1
/dev/mympvg/mymplv  /mnt                ext3       acl,user_xattr     1 2
proc                /proc               proc       defaults           0 0
sysfs               /sys                sysfs      noauto             0 0
debugfs             /sys/kernel/debug   debugfs    noauto             0 0
devpts              /dev/pts            devpts     mode=0620,gid=5    0 0

sles10:~ #
```

Figure 6-14 Linux file system table

Before rebooting Linux, several boot scripts need to be activated so that the device-mapper is loaded and the multipath module is executed early in the boot process before the lvm code scans for available logical volumes. The multipathd daemon is also required to monitor path status and dynamically make changes to the device-mapper tables in response to events such as a path failure.

The required boot scripts can be activated using either the **chkconfig** or **insserv** command. Although the device-mapper and lvm modules are loaded by default on a SLES10 system, they are included here for completeness. The syntax of the **chkconfig** and **insserv** commands is displayed in Figure 6-15.



```
sles10:~ # chkconfig boot.device-mapper on

sles10:~ # chkconfig boot.multipath on

sles10:~ # chkconfig multipathd on

sles10:~ # chkconfig boot.lvm on

                   --------  or --------

sles10:~ # insserv /etc/init.d/boot.device-mapper

sles10:~ # insserv /etc/init.d/boot.multipath

sles10:~ # insserv /etc/init.d/multipathd

sles10:~ # insserv /etc/init.d/boot.lvm

sles10:~ #
```

Figure 6-15 chkconfig and insserv command output

7

Installing Red Hat Enterprise Linux 5 on FCP-attached SCSI

This chapter describes how to install and configure Red Hat Enterprise Linux 5 (RHEL 5) on SCSI disk. We assume that the reader is familiar with the Red Hat Enterprise Linux 5.0 Installation Guide and the Anaconda installer. Topics include:

- The Linux guest user directory entry
- Starting the installation
- Adding an additional SCSI FCP LUN

7.1 The Linux guest user directory entry

For a fresh install, the user directory entry for the Linux guest virtual machine must define the DASD, FCP, and communication channels. For this guest there are three dedicated QDIO OSA addresses for communications, and one dedicated FCP path address (6016) for the Fibre Channel, as shown in Figure 7-1.

```
USER JNS3 JNS3   512M 1024M G
   MACHINE ESA
   LOADDEV PORT 5005076300c300aa 1
   LOADDEV LUN 5023000000000000 1
   CONSOLE 001F 3215 T
   SPOOL 000C 2540 READER *
   SPOOL 000D 2540 PUNCH A
   SPOOL 000E 1403 A
   LINK MAINT 0190 0190 RR
   LINK MAINT 019D 019D RR
   LINK MAINT 019E 019E RR
   LINK TCPIP 0592 0592 RR
   DEDICATE 2708 2708
   DEDICATE 2709 2709
   DEDICATE 270A 270A
   MDISK 191 3390 1878 100 J20W02  MR READ     WRITE    MULTIPLE
   DEDICATE 6016 6016 2
```

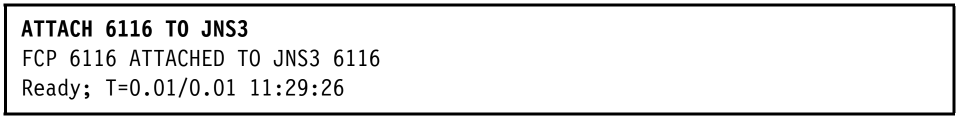
Figure 7-1 The JNS3 VM user directory entry

The LOADDEV PORT and LOADDEV LUN parameters define the IPL WWPN and LUN.

The address 6016 defines an FCP path to the SCSI LUNS.

Note: A virtual machine size of at least 512 MB is required to install RHEL 5, and 1 GB is recommended by RedHat. After installation, RHEL 5 itself can run with much less memory.

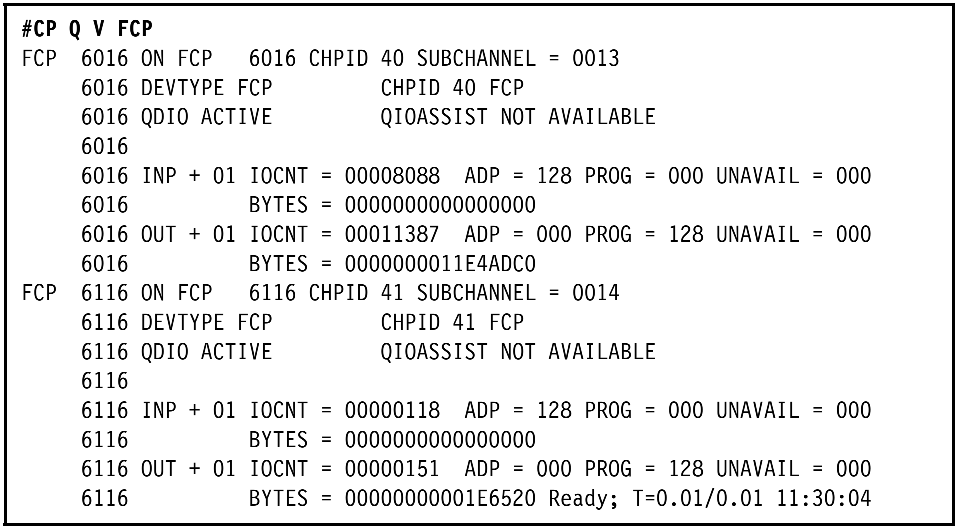
If not defined in the user directory entry, an FCP device can be manually attached to the guest using the CP ATTACH command. Figure 7-2.



```
ATTACH 6116 TO JNS3
FCP 6116 ATTACHED TO JNS3 6116
Ready; T=0.01/0.01 11:29:26
```

Figure 7-2 Attach an FCP device to a Linux guest

Use the QUERY V FCP command to view the FCP devices attached to this guest (Figure 7-3).



```
#CP Q V FCP
FCP  6016 ON FCP   6016 CHPID 40 SUBCHANNEL = 0013
     6016 DEVTYPE FCP         CHPID 40 FCP
     6016 QDIO ACTIVE         QIOASSIST NOT AVAILABLE
     6016
     6016 INP + 01 IOCNT = 00008088  ADP = 128 PROG = 000 UNAVAIL = 000
     6016          BYTES = 0000000000000000
     6016 OUT + 01 IOCNT = 00011387  ADP = 000 PROG = 128 UNAVAIL = 000
     6016          BYTES = 0000000011E4ADC0
FCP  6116 ON FCP   6116 CHPID 41 SUBCHANNEL = 0014
     6116 DEVTYPE FCP         CHPID 41 FCP
     6116 QDIO ACTIVE         QIOASSIST NOT AVAILABLE
     6116
     6116 INP + 01 IOCNT = 00000118  ADP = 128 PROG = 000 UNAVAIL = 000
     6116          BYTES = 0000000000000000
     6116 OUT + 01 IOCNT = 00000151  ADP = 000 PROG = 128 UNAVAIL = 000
     6116          BYTES = 00000000001E6520 Ready; T=0.01/0.01 11:30:04
```

Figure 7-3 Query FCP devices attached to a Linux guest

7.2 Starting the installation

With RHEL 5 the Anaconda installer now supports direct installation to FCP-attached SCSI disk.

This document assumes that the reader is familiar with the process of installing Linux on zSeries® as a guest of z/VM and skips some of the basic installation steps. The focus of this chapter is on the specific tasks required to install onto FCP devices.

7.2.1 The RHEL 5 PARM file

Figure 7-4 shows an example of the RHEL 5 parmfile. Since Red Hat has a limit of 32 total parameters in the parameter file, in order to stay within these limitations, a new configuration file on a CMS DASD should be created to configure the initial network setup and the DASD definitions. You still must use a .parm file for the real kernel parameters, such as root=/dev/ram0 or ip=off ramdisk_size=40000, and also single parameters that are not assigned to variables, such as vnc.

Two parameters are used in z/VM installs to point the installation program at the new CMS configuration file:

```
CMSDASD=191 CMSCONFFILE=redhat.conf
```

The CMSDASD parameter defines the z/VM device ID of the CMS-formatted DASD, which contains the configuration file. CMSDASD often points to the user's A disk (usually disk 191) of the z/VM guest account. The name of the configuration file must be set with CMSCONFFILE, and the file name needs to be lowercase. See Figure 7-4.

```
root=/dev/ram0 ro ip=off ramdisk_size=40000
CMSDASD=191 CMSCONFFILE=rhel5.conf
vnc
```

Figure 7-4 The RHEL5.PARM file

A sample of the rhel5.conf file is in Figure 7-5. This file sets up the DASD, FCP LUNS, and the OSA Network connection. Check the Red Hat RHEL5 Installation Guide, Part:III, for further explanation of this file.

```
DASD="999"
FCP_1="0.0.6016 0x01 0x5005076300c300aa 0x1 5020000000000000"
HOSTNAME="jns3.endicott.ibm.com" NETTYPE="qeth" IPADDR="9.60.86.g52"
SUBCHANNELS="0.0.2708,0.0.2709,0.0.270A" PORTNAME="JNS"
NETWORK="9.60.86.0" MTU="1500"
NETMASK="255.255.255.128" BROADCAST="9.60.86.127" GATEWAY="9.60.86.1"
DNS="9.0.3.1:9.0.2.11" SEARCHDNS="ibm.com:pok.ibm.com"
```

Figure 7-5 The RHEL5 CONF file

Note: Because of a limitation with Anaconda the RHEL5 Installer, there must be a DASD device defined, even though this installation will not use any ECKD devices. (It is possible to use a non-existent device address, such as DASD="999", to satisfy this limitation.)

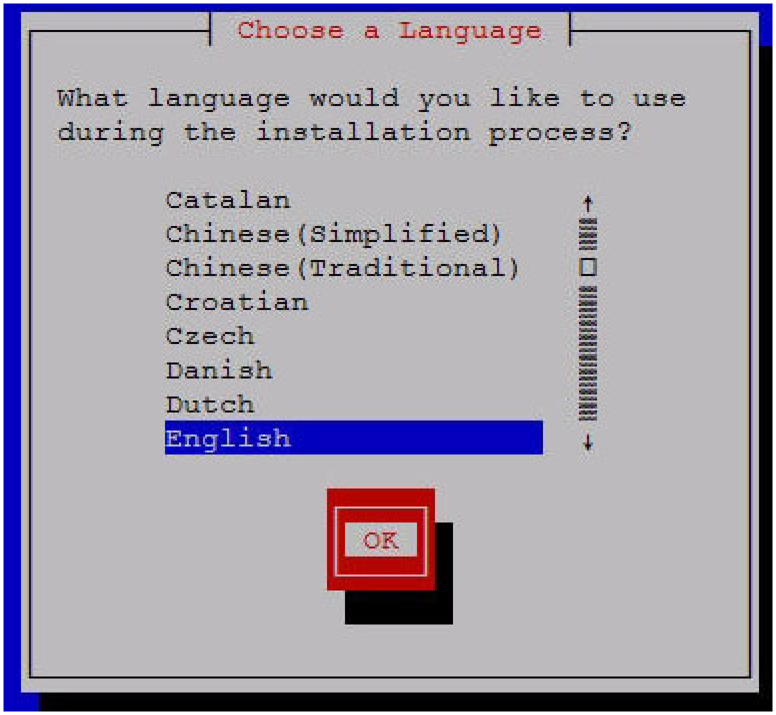
7.2.2 Completing the installation

Open an SSH session to the Linux guest. We use the PUTTY SSH client for our examples here.

Note: The PUTTY SSH client can be downloaded from:

http://www.chiark.greenend.org.uk/~sgtatham/putty/download.html

1. The Anaconda installer automatically starts the installation process, as shown in Figure 7-6. Select your language, then click **OK** and press Enter.



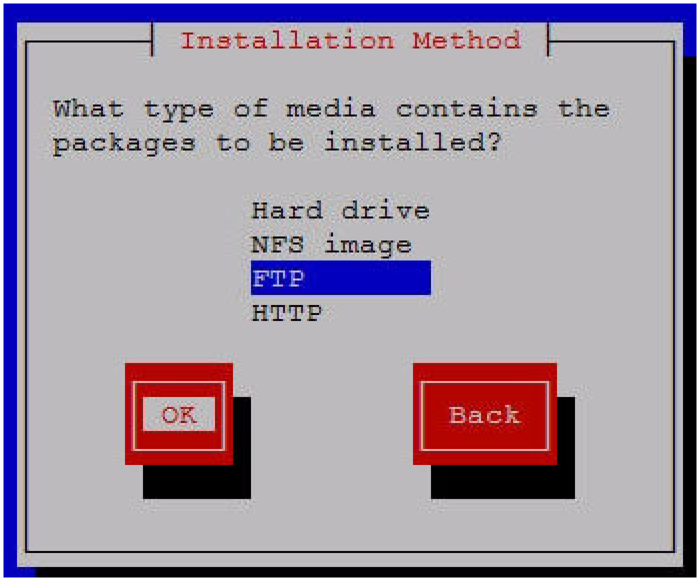
Choose a Language

What language would you like to use during the installation process?

Figure 7-6 Select the language

Note: The Installation is self-explanatory, with dialogs to help you select the desired features.

2. Select an installation method (NFS, FTP, HTTP, or Internal disk). We choose an FTP server (Figure 7-7).

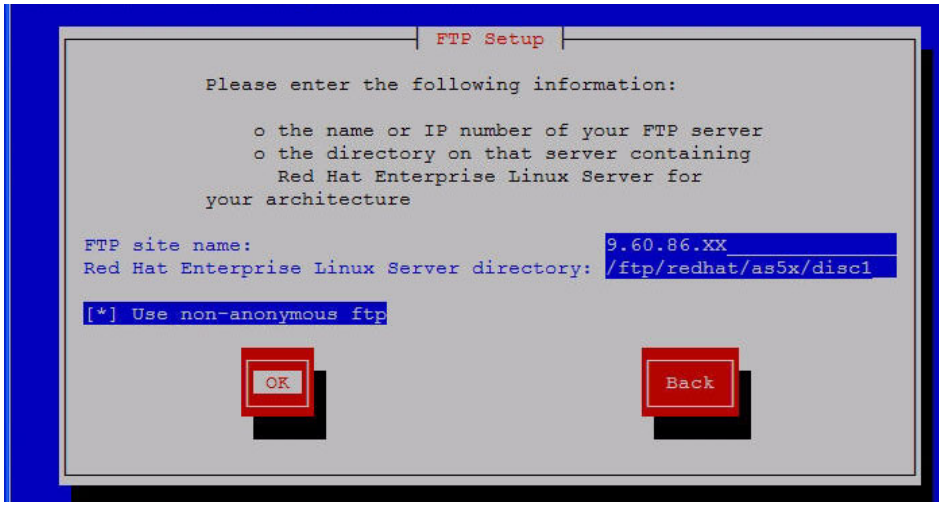


Installation Method

What type of media contains the packages to be installed?

Figure 7-7 Select an installation method

3. Enter your FTP server address, the directory path to your RHEL 5 disc1 image, and since our FTP server requires a user ID and password, we check the box to use a non-anonymous FTP before tabbing to OK. See Figure 7-8.



FTP Setup

Please enter the following information:

- o the name or IP number of your FTP server
- o the directory on that server containing Red Hat Enterprise Linux Server for your architecture

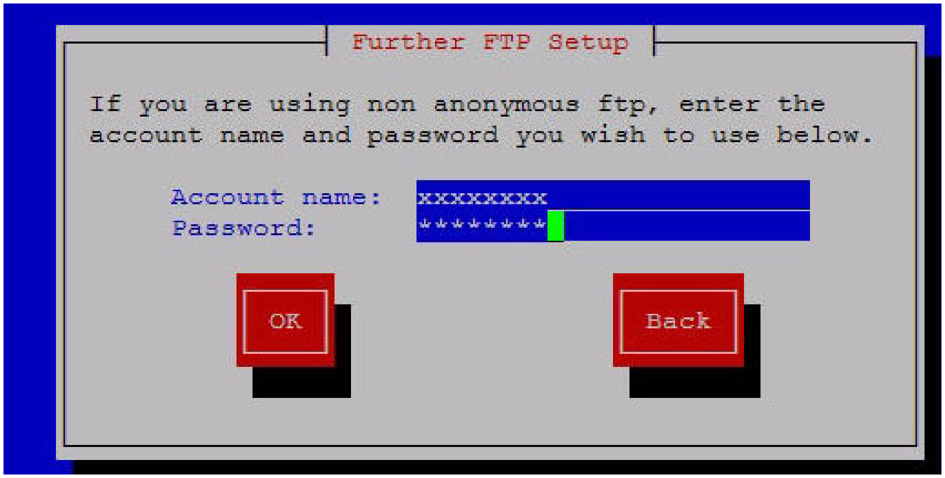
FTP site name: 9.60.86.XX

Red Hat Enterprise Linux Server directory: /ftp/redhat/as5x/disc1

[*] Use non-anonymous ftp

Figure 7-8 Enter your FTP Server Information

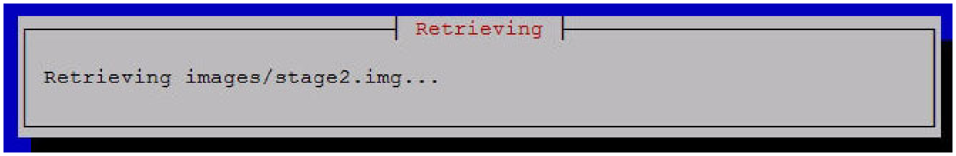
4. Enter your FTP server user ID and password. Tab to OK (Figure 7-9).



```
Further FTP Setup

If you are using non anonymous ftp, enter the
account name and password you wish to use below.

     Account name:   xxxxxxxx
     Password:       ********
```

Figure 7-9 Enter your FTP server user ID and password

The installation process now accesses the FTP server to load the files necessary to continue the installation process (Figure 7-10).

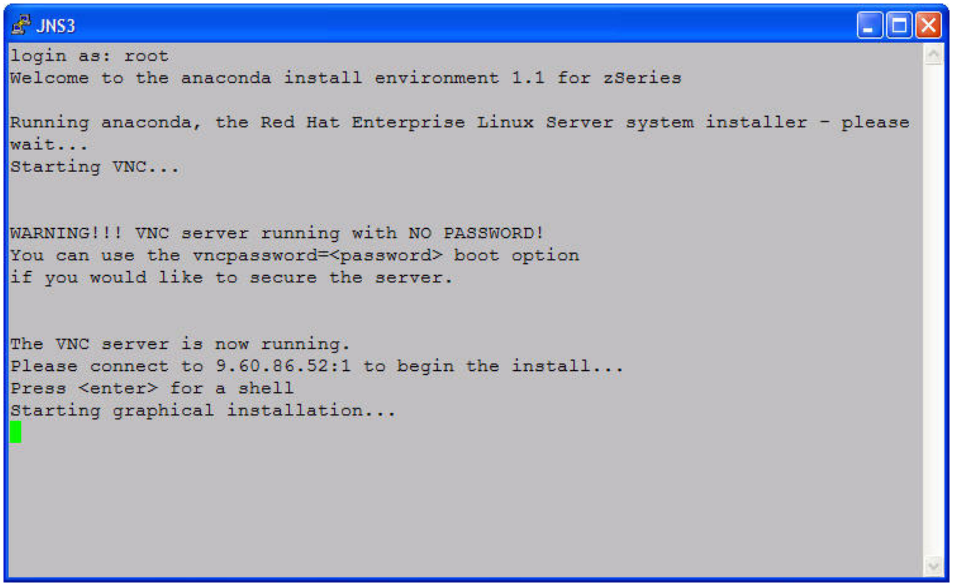


```
Retrieving

Retrieving images/stage2.img...
```

Figure 7-10 Installation continues

Note: The installation now accesses the files on the FTP server. If there is an error at this point, chances are that there is either a problem with the files on the FTP server or there is an access problem to those files.

Since the .PARM file specified that VNC was to be used to run the install process, the VNCSERVER is now started and you are directed to start a VNC client session to continue with the graphical installation (Figure 7-11).



```
login as: root
Welcome to the anaconda install environment 1.1 for zSeries

Running anaconda, the Red Hat Enterprise Linux Server system installer - please
wait...
Starting VNC...


WARNING!!! VNC server running with NO PASSWORD!
You can use the vncpassword=<password> boot option
if you would like to secure the server.


The VNC server is now running.
Please connect to 9.60.86.52:1 to begin the install...
Press <enter> for a shell
Starting graphical installation...
```

Figure 7-11 Continue installation using VNC

Note: Two popular VNC viewers that we use on a regular basis are listed below. There may be others available, but this is what we have experience with.

- TightVNC

  `http://www.tightvnc.com/download.html`

- RealVNC

  `http://www.realvnc.com/download.html`

Once the VNC connection is made, the first window of the graphical installation process is displayed. If you have not already done so, you may want to review the release notes at this time (Figure 7-12).

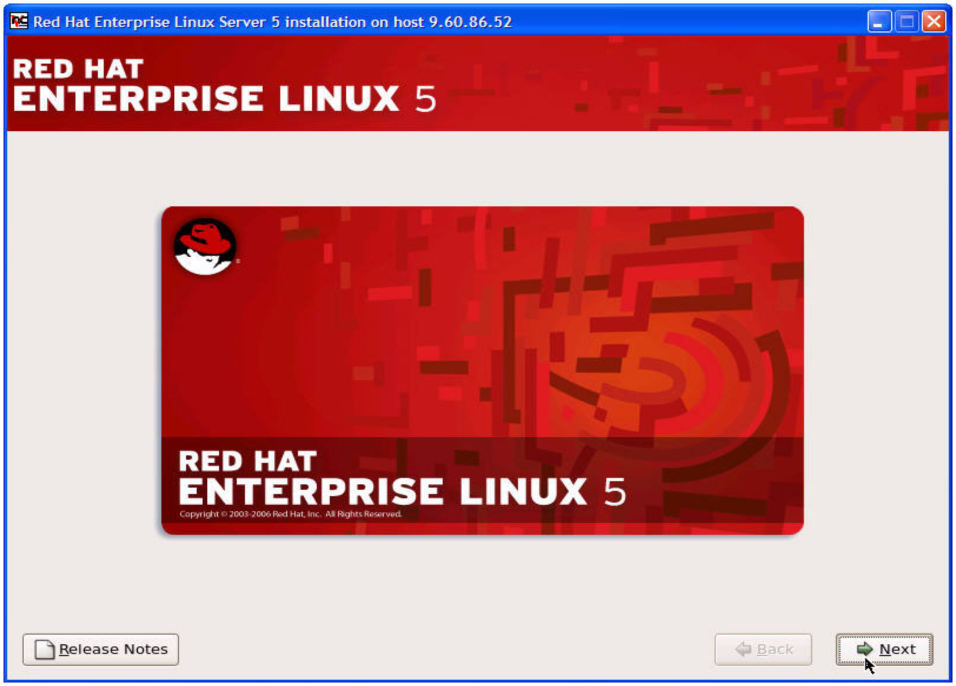
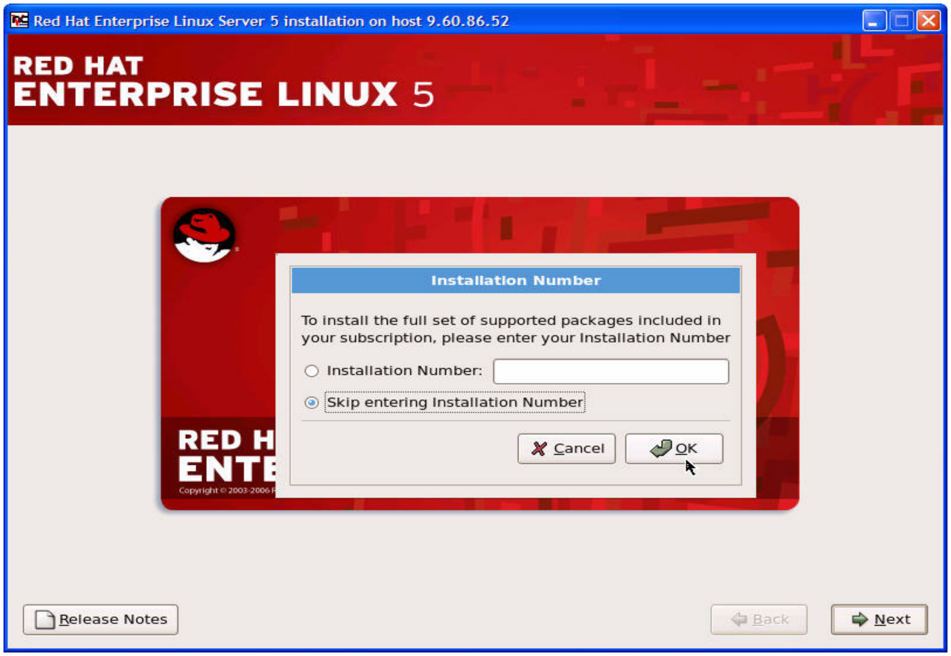

Red Hat Enterprise Linux Server 5 installation on host 9.60.86.52
RED HAT
ENTERPRISE LINUX 5
RED HAT
ENTERPRISE LINUX 5
Copyright © 2003-2006 Red Hat, Inc. All Rights Reserved.
Release Notes
Back
Next

Figure 7-12 Graphical Installation Welcome window

Note: Depending on the system and network configuration, this window may take up to several minutes to load. Be patient.

Enter the installation number if you have one. The installation number is not a license number. This number is used to select a predetermined package selection that will be available to the installer. If you do not have an installation number you can choose to skip entering the installation number. There is an opportunity later in the installation process to select additional packages. In our example the option to skip entering this number was taken (Figure 7-13).


Red Hat Enterprise Linux Server 5 installation on host 9.60.86.52
RED HAT
ENTERPRISE LINUX 5

Installation Number

To install the full set of supported packages included in your subscription, please enter your Installation Number

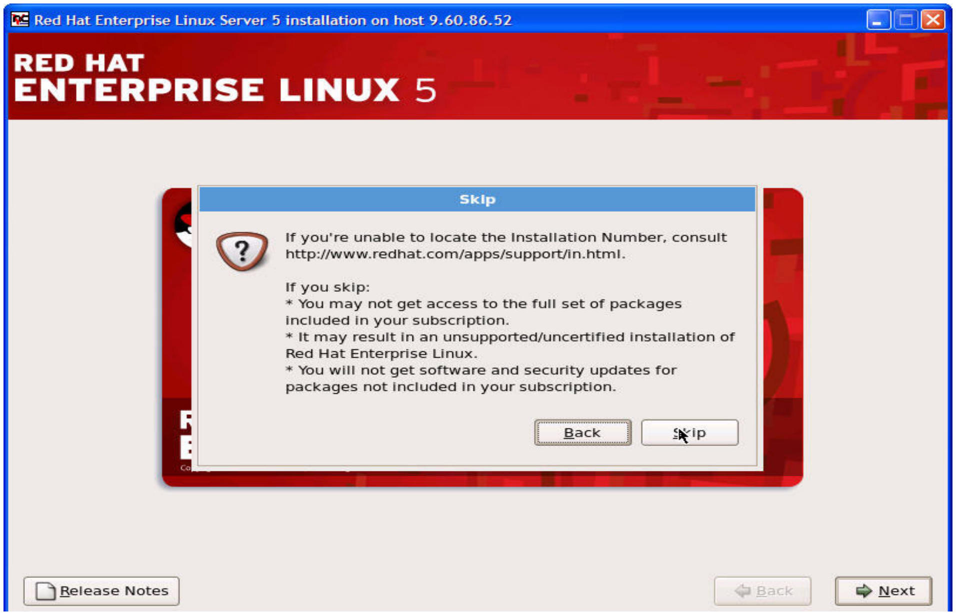
○ Installation Number:

● Skip entering Installation Number

Cancel
OK
Release Notes
Back
Next

Figure 7-13 Installation number

Since no installation number was provided, the installation process wants the installer to confirm the selection (Figure 7-14).



Red Hat Enterprise Linux Server 5 installation on host 9.60.86.52

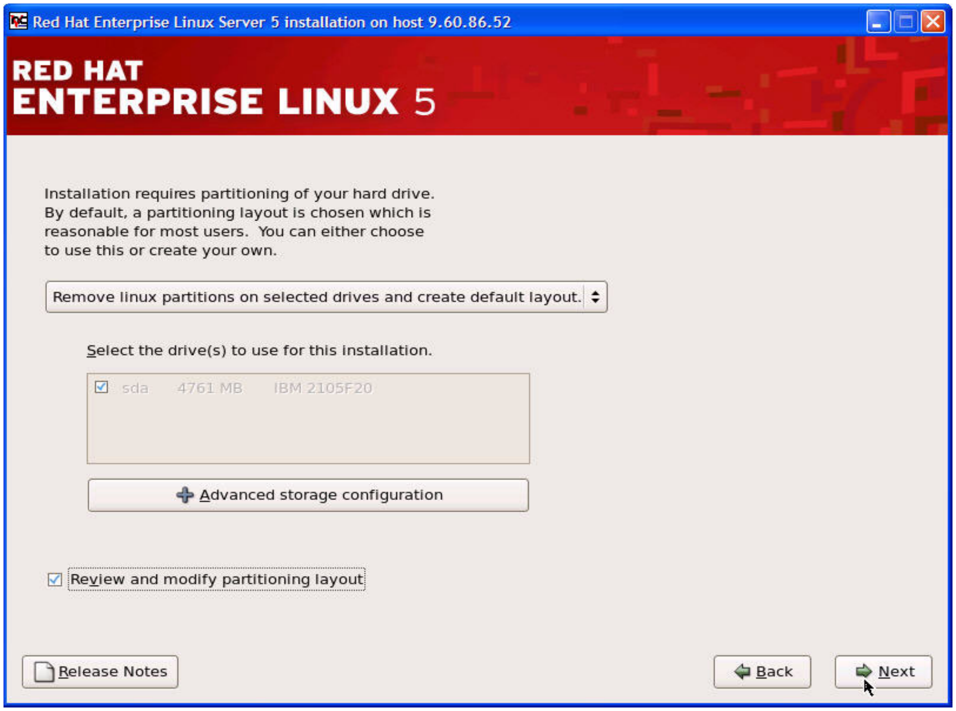
Skip

If you're unable to locate the Installation Number, consult http://www.redhat.com/apps/support/in.html.

If you skip:
* You may not get access to the full set of packages included in your subscription.
* It may result in an unsupported/uncertified installation of Red Hat Enterprise Linux.
* You will not get software and security updates for packages not included in your subscription.

Figure 7-14 Installation number confirmation

This next window shows the physical disk drives that are available for this installation. Disks with a name of dasdx are ECKD DASD. Disks with names of sdx are FCP SCSI disks. So since the parameter FCP1= was defined in the redhat.conf file earlier, there is one FCP SCSI disk available (Figure 7-15). Since this is the first FCP disk available it is named sda. The size of this disk is 4761 MB and it is on an IBM 2105 Shark. If there were more than one FCP device in our redhat.conf file (FCP2=, FCP3=, and so on) we would see those devices listed here also.


Red Hat Enterprise Linux Server 5 installation on host 9.60.86.52
RED HAT
ENTERPRISE LINUX 5

Installation requires partitioning of your hard drive. By default, a partitioning layout is chosen which is reasonable for most users. You can either choose to use this or create your own.

Remove linux partitions on selected drives and create default layout.

Select the drive(s) to use for this installation.

☑ sda 4761 MB IBM 2105F20

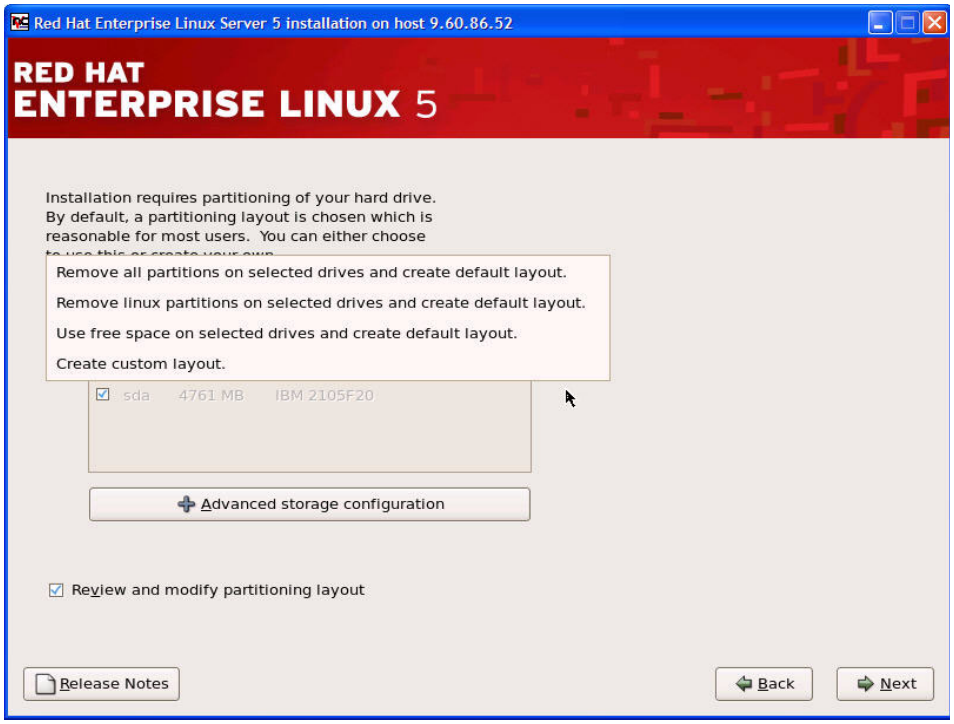
Advanced storage configuration

☑ Review and modify partitioning layout

Figure 7-15 Partitioning available disks

Note: Typically, any devices defined and available to a guest are sensed. However, because of the flexibility of configurations associated with FCP, IBM System z requires that any Fibre Channel Protocol (FCP) device be entered manually (either in the installation program interactively, or specified as unique parameter entries in the CMS conf file).

On the window shown in Figure 7-16, select the pull-down. The first three options allow you to perform an automated installation without having to partition the disks.



Red Hat Enterprise Linux Server 5 installation on host 9.60.86.52

RED HAT ENTERPRISE LINUX 5

Installation requires partitioning of your hard drive. By default, a partitioning layout is chosen which is reasonable for most users. You can either choose to use this or create your own.

- Remove all partitions on selected drives and create default layout.
- Remove linux partitions on selected drives and create default layout.
- Use free space on selected drives and create default layout.
- Create custom layout.

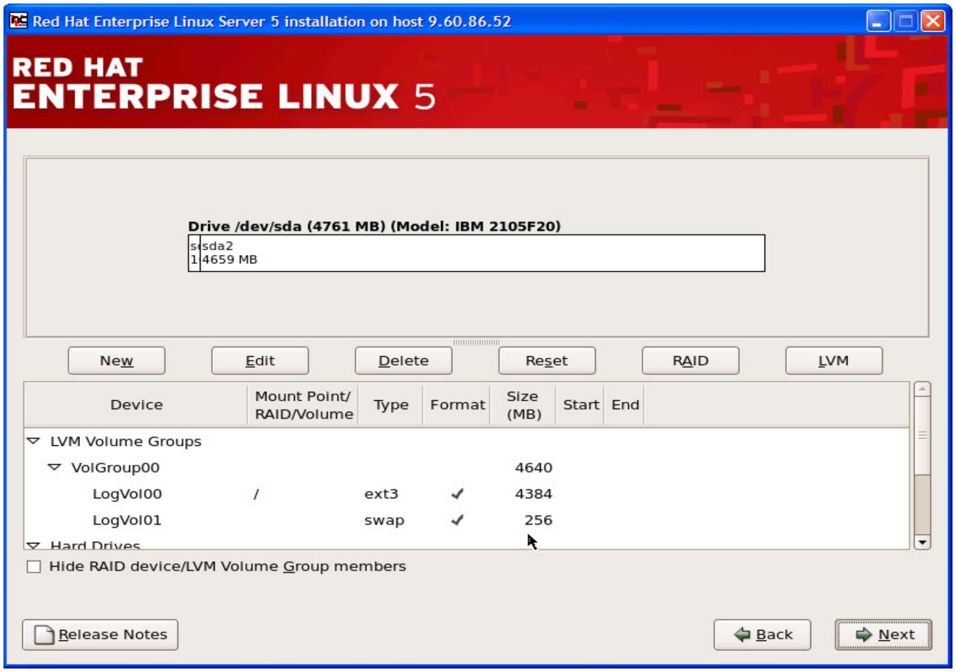
☑ sda 4761 MB IBM 2105F20

☑ Review and modify partitioning layout

Figure 7-16 Disk partitioning, selection

If you do not feel comfortable with partitioning the system, we recommend that you do not choose to create a custom layout, and instead let the installation program partition it for you. For this test installation the option to take the default was taken. However, the option to review or modify the partitioning layout was also checked. This enables a review of the default layout to make sure that the default configuration will in fact meet the needs of the configuration. If not, then there will be an opportunity to alter the configuration. It is much easier to change things during the installation now rather than later.

Disk partitioning enables you to divide your physical storage disks into separate logical partitions and therefore separate mount points. At this point, the partitioning layout should be closely reviewed (see Figure 7-17). Check the sizes of the partitions. Make sure that there is enough space allocated in each partition to accommodate each of the applications. Also consider future growth. If there are resources available now, it is much easier to configure the space at this time during the installation.


Red Hat Enterprise Linux Server 5 installation on host 9.60.86.52
RED HAT
ENTERPRISE LINUX 5

Drive /dev/sda (4761 MB) (Model: IBM 2105F20)

Device	Mount Point/ RAID/Volume	Type	Format	Size (MB)	Start	End
▽ LVM Volume Groups						
▽ VolGroup00				4640		
LogVol00	/	ext3	✓	4384		
LogVol01		swap	✓	256		
▽ Hard Drives						

□ Hide RAID device/LVM Volume Group members

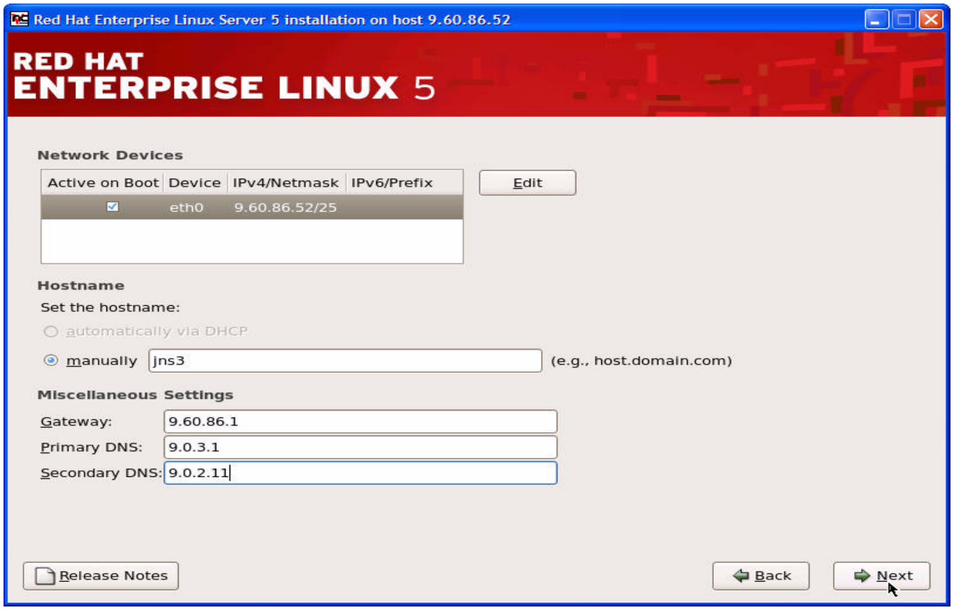
Figure 7-17 Disk partitioning layout

Note: By default the installer tried to configure 1936 M of swap space. This cut the available space of the 4.6 G volume group almost in half. In the example above swap space was reduced to 256 M. Since guests should be configured so that there is little or no swapping, having all this space reserved for swap seems like a waste. Of course, there are exceptions. This should be something that is looked at when the guest is tuned. Although it is not shown in this configuration, we usually recommend that a small swap space of approximately 200–300 M be located on a small vdisk, with some additional space on disk also configured as secondary swap space.

The installation program automatically detects any network devices and displays them in the network devices list.

The eth0 connection shown in Figure 7-18 has already been configured since it is supplied in the redhat.conf file.

Any additional network devices detected can also be configured at this time.



Red Hat Enterprise Linux Server 5 installation on host 9.60.86.52
RED HAT
ENTERPRISE LINUX 5

Network Devices

Active on Boot	Device	IPv4/Netmask	IPv6/Prefix
☑	eth0	9.60.86.52/25	

Hostname

Set the hostname:

○ automatically via DHCP

● manually jns3 (e.g., host.domain.com)

Miscellaneous Settings

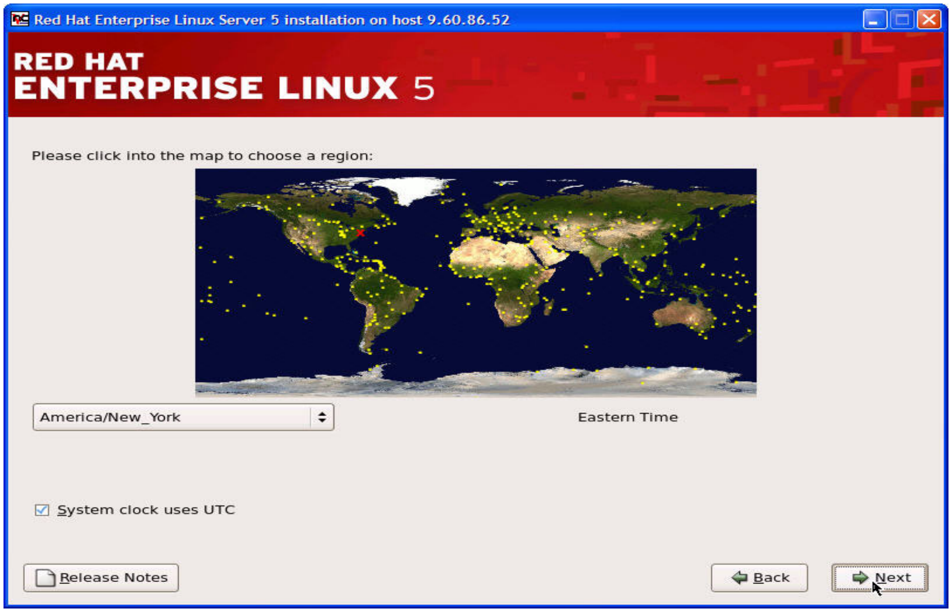
Gateway: 9.60.86.1

Primary DNS: 9.0.3.1

Secondary DNS: 9.0.2.11

Figure 7-18 Network configuration

Set the time zone by selecting the city closest to the computer's physical location (Figure 7-19). Select the pull-down, then scroll through the selections until your time zone is located. You can also click the map to zoom to that area of the map. Then select one of the little yellow dots closet to your location. This changes the time zone to that city and also changes the yellow dot to a red X.



Red Hat Enterprise Linux Server 5 installation on host 9.60.86.52

Please click into the map to choose a region:

America/New_York

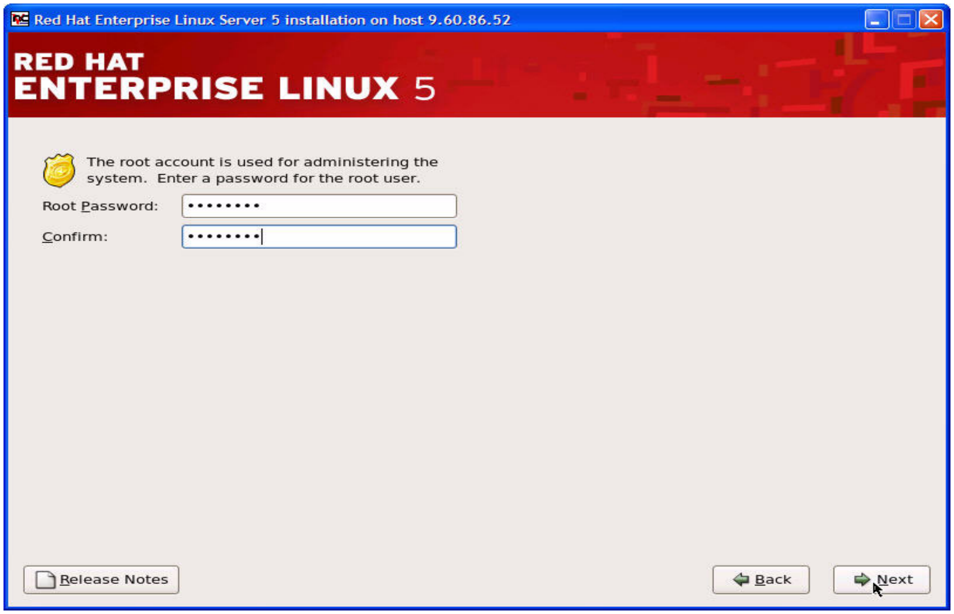
Eastern Time

☑ System clock uses UTC

Figure 7-19 Time zone

Setting up a root account and password is a very important step of the installation (Figure 7-20). The root account is the administrator account for the Linux system.

- The password must be a minimum of six characters in length.
- The password should be easy to remember but difficult for someone to guess. We recommend a combination of upper and lowercase letters and numbers.
- Do not lose this password.


Red Hat Enterprise Linux Server 5 installation on host 9.60.86.52
RED HAT
ENTERPRISE LINUX 5

The root account is used for administering the system. Enter a password for the root user.

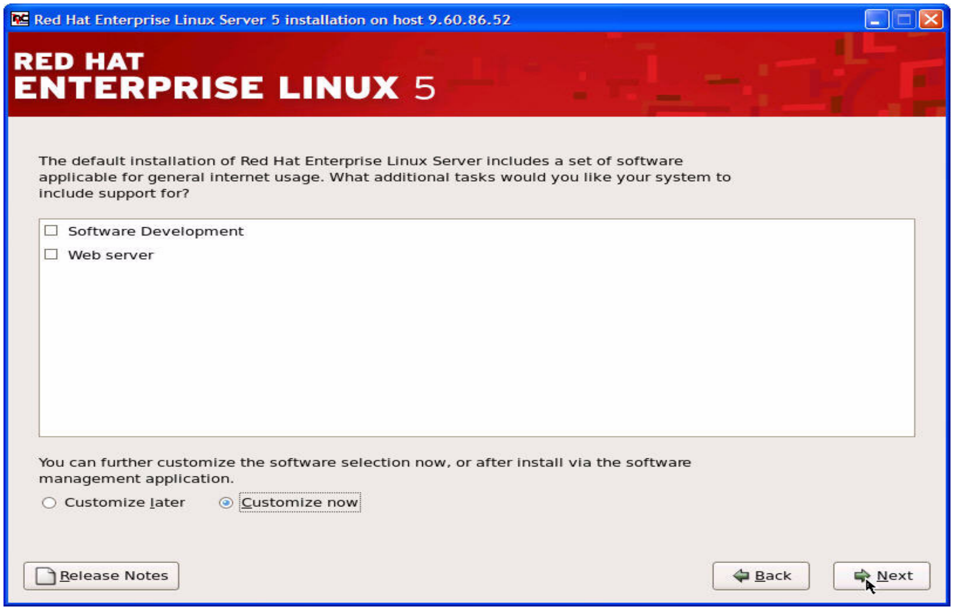
Root Password: ••••••••

Confirm: ••••••••

Figure 7-20 ROOT password

Note: The root account should be used to administer the Linux system. The root user has complete control/access to the entire Linux system. Because of this, we recommend that you log in with a different user and **su** to root only when administrator authority will be needed.

Red Hat includes a default selection of general applications with their installation. At this time any optional package groups can be selected (Figure 7-21).



Red Hat Enterprise Linux Server 5 installation on host 9.60.86.52
RED HAT
ENTERPRISE LINUX 5

The default installation of Red Hat Enterprise Linux Server includes a set of software applicable for general internet usage. What additional tasks would you like your system to include support for?

☐ Software Development

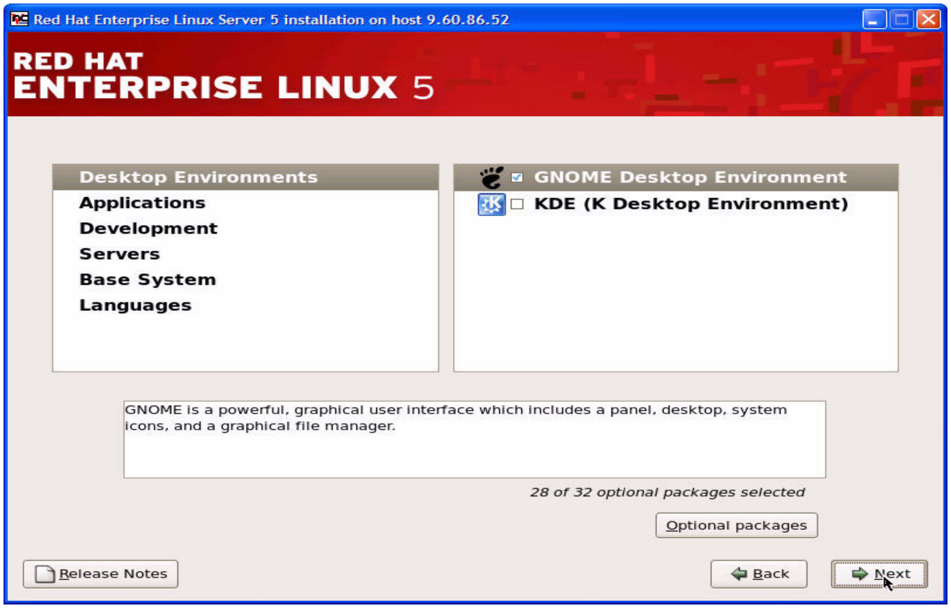
☐ Web server

You can further customize the software selection now, or after install via the software management application.

○ Customize later ● Customize now

Figure 7-21 Package group selection

If the Customize Now button is selected, the installer presents another window where there will be a further opportunity to select and customize additional individual optional packages. In this example the Customize Now button has been selected (Figure 7-22).



☑ GNOME Desktop Environment
□ KDE (K Desktop Environment)

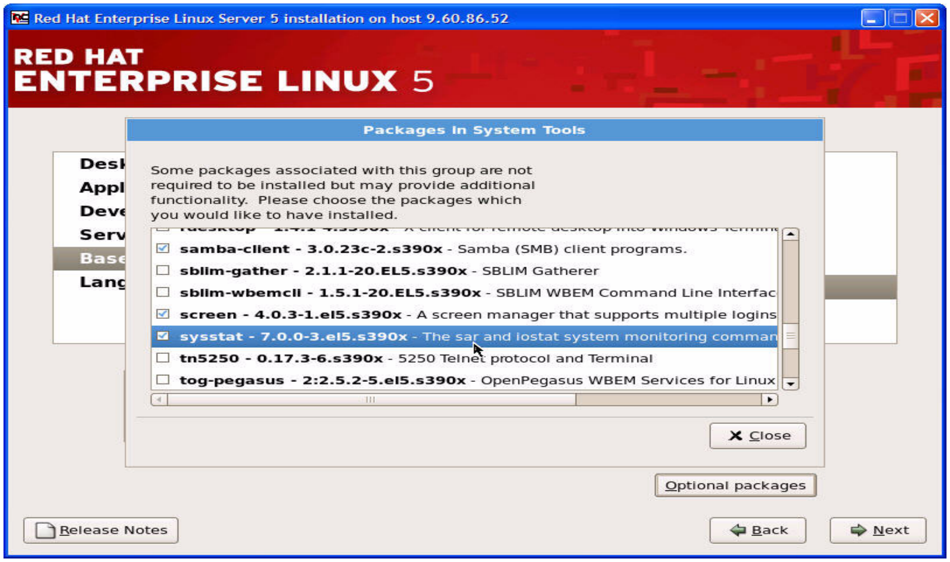
GNOME is a powerful, graphical user interface which includes a panel, desktop, system icons, and a graphical file manager.

28 of 32 optional packages selected

Figure 7-22 Package selection

We recommend that each package group and its optional packages be reviewed (Figure 7-22 on page 131). Add those additional packages that may be needed or wanted. But most of all, we recommend that any unneeded packages be removed. They can always be added later if a need is identified. This reduces some of the storage space used by the guest and also helps to prevent the automatic starting of applications not needed. The idea is to only install what is needed.

Figure 7-23, Figure 7-24 on page 133, and Figure 7-26 on page 134 show the selection of a couple of applications recommended for use within the FCP environment. These are sysstat, lsscsi, and mdadm.



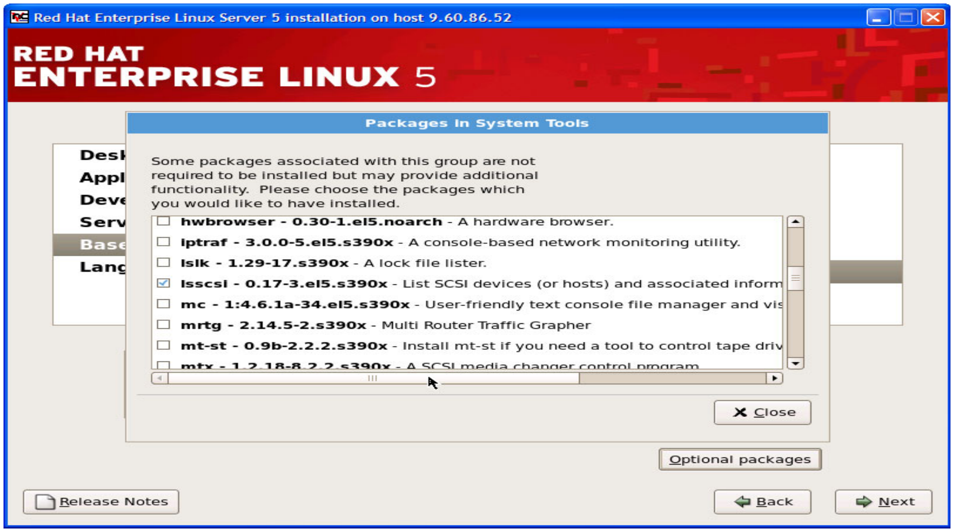
Red Hat Enterprise Linux Server 5 installation on host 9.60.86.52

Packages in System Tools

Some packages associated with this group are not required to be installed but may provide additional functionality. Please choose the packages which you would like to have installed.

- ☑ **samba-client - 3.0.23c-2.s390x** - Samba (SMB) client programs.
- ☐ **sblim-gather - 2.1.1-20.EL5.s390x** - SBLIM Gatherer
- ☐ **sblim-wbemcli - 1.5.1-20.EL5.s390x** - SBLIM WBEM Command Line Interfac
- ☑ **screen - 4.0.3-1.el5.s390x** - A screen manager that supports multiple logins
- ☑ **sysstat - 7.0.0-3.el5.s390x** - The sar and iostat system monitoring comman
- ☐ **tn5250 - 0.17.3-6.s390x** - 5250 Telnet protocol and Terminal
- ☐ **tog-pegasus - 2:2.5.2-5.el5.s390x** - OpenPegasus WBEM Services for Linux

Figure 7-23 sysstat selection



Red Hat Enterprise Linux Server 5 installation on host 9.60.86.52

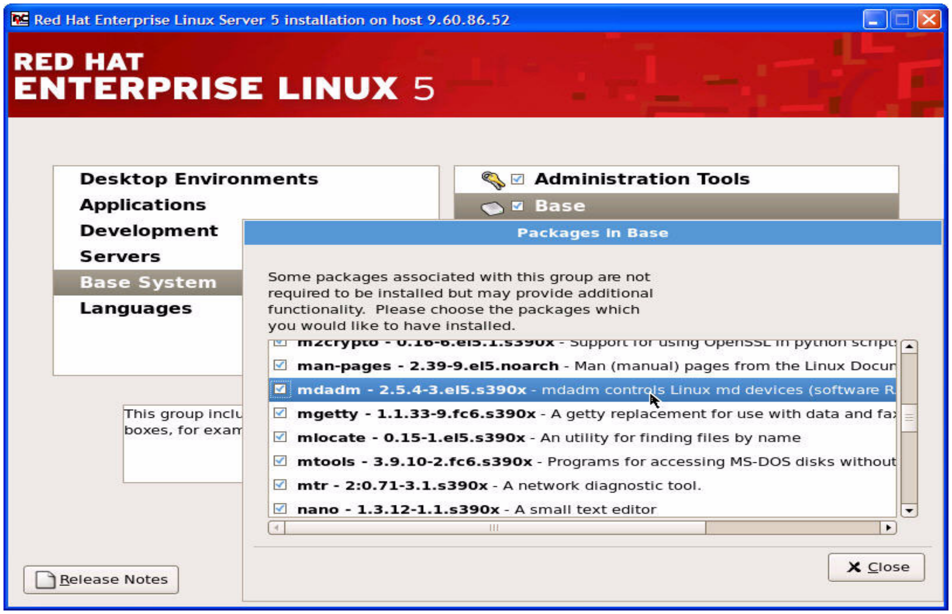
RED HAT ENTERPRISE LINUX 5

Packages in System Tools

Some packages associated with this group are not required to be installed but may provide additional functionality. Please choose the packages which you would like to have installed.

- ☐ **hwbrowser - 0.30-1.el5.noarch** - A hardware browser.
- ☐ **iptraf - 3.0.0-5.el5.s390x** - A console-based network monitoring utility.
- ☐ **lslk - 1.29-17.s390x** - A lock file lister.
- ☑ **lsscsi - 0.17-3.el5.s390x** - List SCSI devices (or hosts) and associated inform
- ☐ **mc - 1:4.6.1a-34.el5.s390x** - User-friendly text console file manager and vis
- ☐ **mrtg - 2.14.5-2.s390x** - Multi Router Traffic Grapher
- ☐ **mt-st - 0.9b-2.2.2.s390x** - Install mt-st if you need a tool to control tape driv
- ☐ **mtx - 1.2.18-8.2.2.s390x** - A SCSI media changer control program

Figure 7-24 lsscsi selection



Red Hat Enterprise Linux Server 5 installation on host 9.60.86.52

RED HAT ENTERPRISE LINUX 5

Packages in Base

Some packages associated with this group are not required to be installed but may provide additional functionality. Please choose the packages which you would like to have installed.

- ☑ **m2crypto - 0.16-6.el5.1.s390x** - Support for using OpenSSL in python scripts
- ☑ **man-pages - 2.39-9.el5.noarch** - Man (manual) pages from the Linux Docur
- ☑ **mdadm - 2.5.4-3.el5.s390x** - mdadm controls Linux md devices (software R
- ☑ **mgetty - 1.1.33-9.fc6.s390x** - A getty replacement for use with data and fa:
- ☑ **mlocate - 0.15-1.el5.s390x** - An utility for finding files by name
- ☑ **mtools - 3.9.10-2.fc6.s390x** - Programs for accessing MS-DOS disks without
- ☑ **mtr - 2:0.71-3.1.s390x** - A network diagnostic tool.
- ☑ **nano - 1.3.12-1.1.s390x** - A small text editor

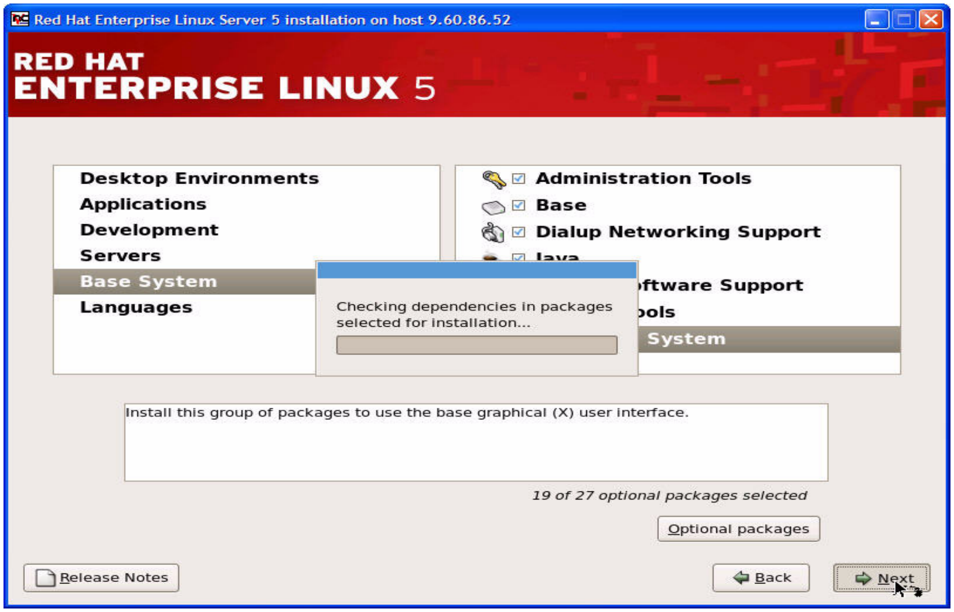
Figure 7-25 mdadm selection

sysstat includes four groups of monitoring tools (sar / sadc / sadf, iostat, mpstat, pidstat) for global system performance analysis.

lsscsi uses information in sysfs to list all scsi devices currently attached to the system.

The mdadm package is used to create, manage, and monitor Linux MD (software RAID) devices. This package will be needed later if the FCP devices will be set up with multipathing.

Once the packages have been selected, clicking **Next** checks for any dependencies. If there are any errors, those dependencies must be taken care of before proceeding (see Figure 7-26).



☑ Administration Tools
☑ Base
☑ Dialup Networking Support
☑ Java
ftware Support
ools
System

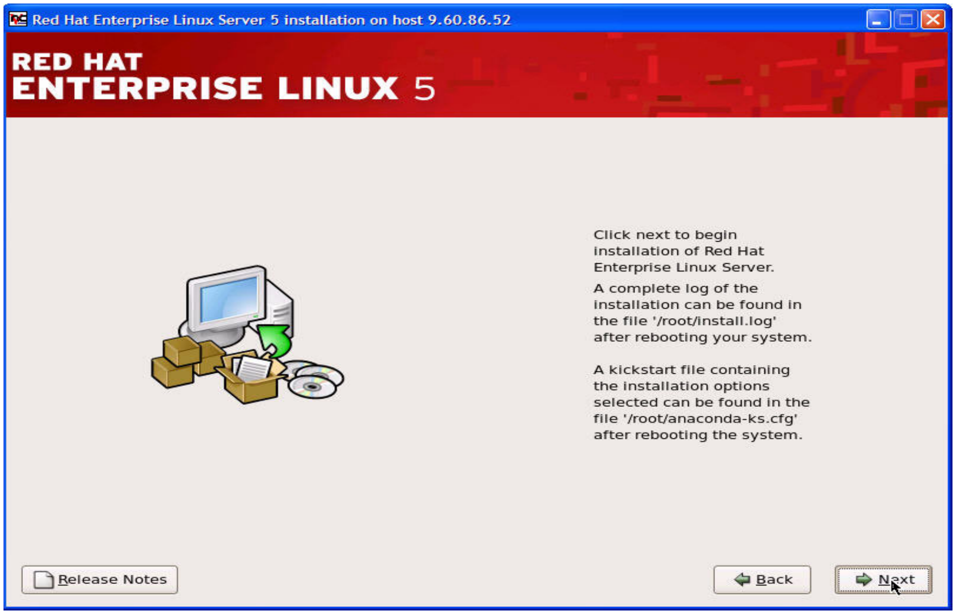
Checking dependencies in packages selected for installation...

Install this group of packages to use the base graphical (X) user interface.

19 of 27 optional packages selected

Figure 7-26 Package dependency check

Figure 7-27, Figure 7-28 on page 136, and Figure 7-29 on page 136 show the windows you should see upon installation start through to completion of installation.

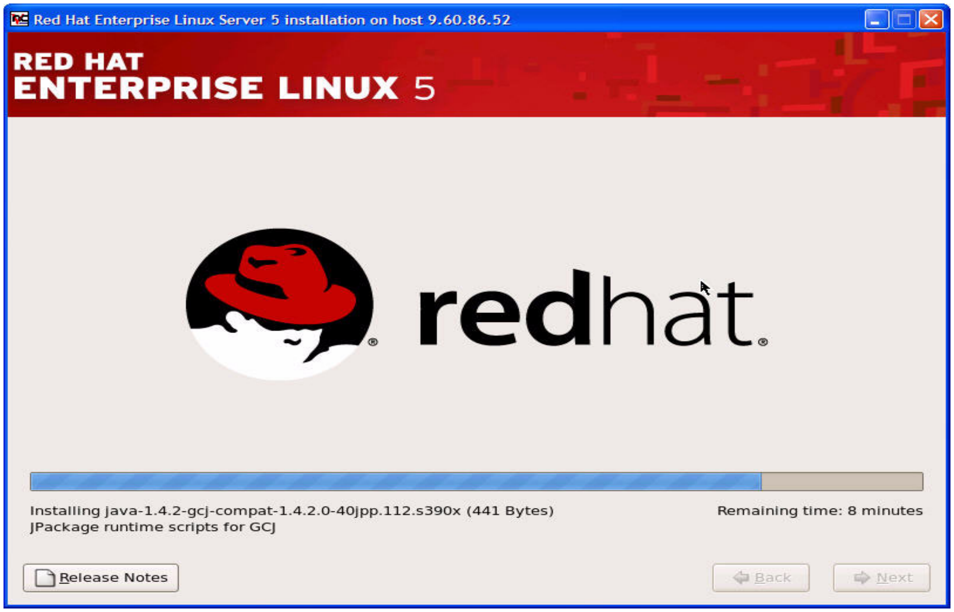

Red Hat Enterprise Linux Server 5 installation on host 9.60.86.52
RED HAT
ENTERPRISE LINUX 5
Release Notes
Back
Next

Click next to begin installation of Red Hat Enterprise Linux Server.

A complete log of the installation can be found in the file '/root/install.log' after rebooting your system.

A kickstart file containing the installation options selected can be found in the file '/root/anaconda-ks.cfg' after rebooting the system.

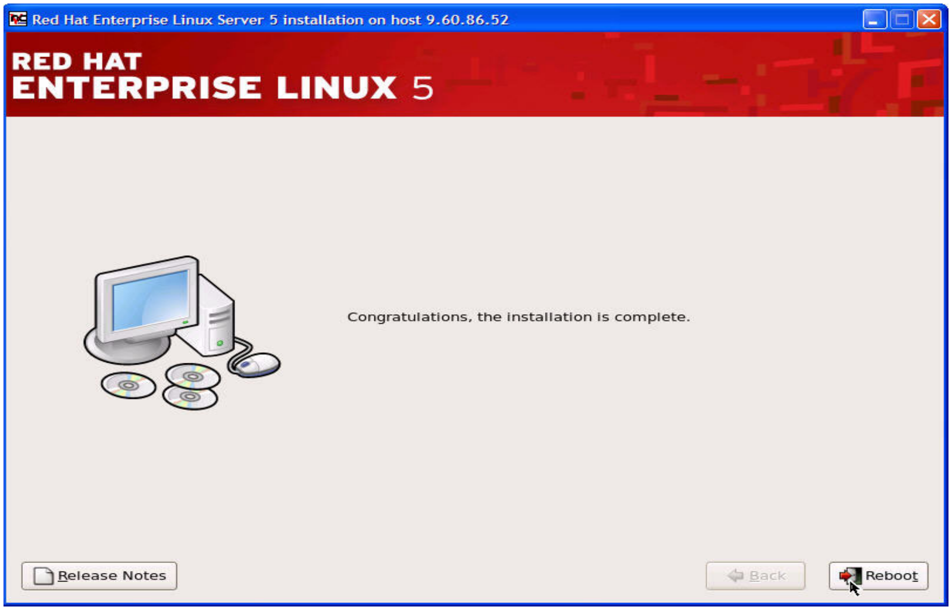
Figure 7-27 Start package installation


Red Hat Enterprise Linux Server 5 installation on host 9.60.86.52
RED HAT
ENTERPRISE LINUX 5
redhat
Release Notes
Back
Next

Installing java-1.4.2-gcj-compat-1.4.2.0-40jpp.112.s390x (441 Bytes)
JPackage runtime scripts for GCJ

Remaining time: 8 minutes

Figure 7-28 Package installation progress


Red Hat Enterprise Linux Server 5 installation on host 9.60.86.52
RED HAT
ENTERPRISE LINUX 5
Release Notes
Back
Reboot

Congratulations, the installation is complete.

Figure 7-29 Installation completion

7.2.3 Completing the install

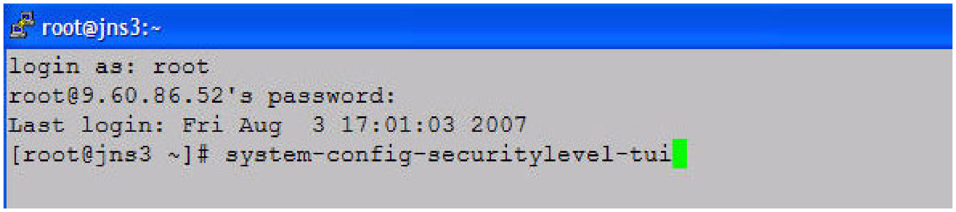
At this time shut down the Linux guest completely and then reboot. When the Linux guest comes back up, it is a good idea to issue the **`df -h`** command and check the mount points and the available space.

```
[root@jns3 ~]# df -h
df -h
Filesystem            Size  Used Avail Use% Mounted on
/dev/mapper/VolGroup00-LogVol00
                      4.2G  1.9G  2.2G  46% /
/dev/sda1              99M   13M   82M  14% /boot
tmpfs                 502M     0  502M   0% /dev/shm
```

Figure 7-30 Verifying disk partitioning and space

By default, RHEL 5 turns on the firewall. So at this point the only place root can log on is from the 3270 session, ssh, or from other devices as listed in the /etc/securetty file. If VNC will be used to log into the Linux system, the firewall will need to be customized or turned off. In Figure 7-31, a PUTTY ssh session was started and the user root was logged on.

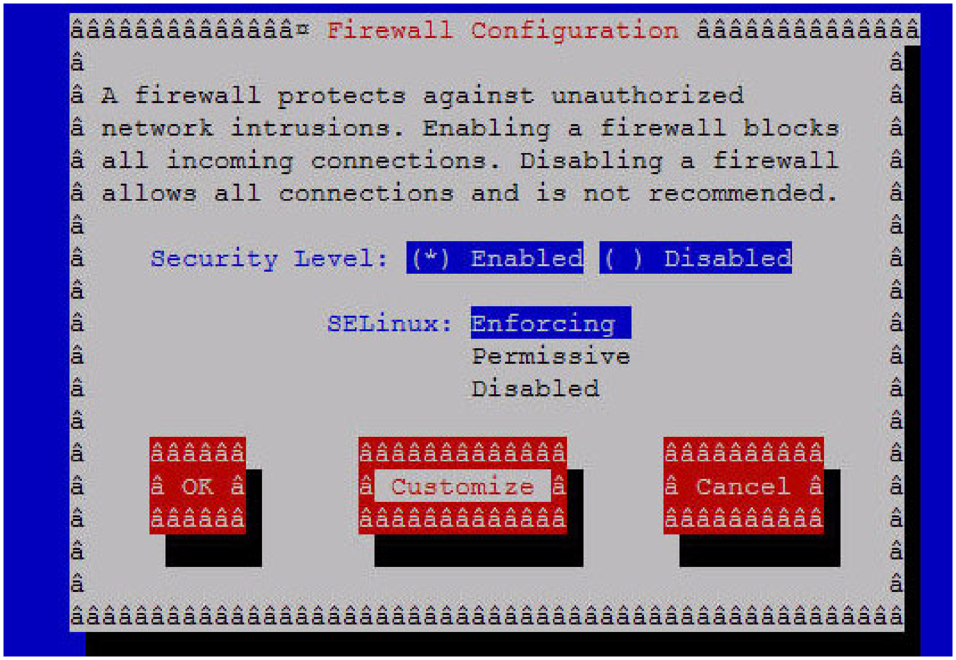
Run *s*ystem-config-securitylevel-tui to bring up the firewall customization panel.



```
login as: root
root@9.60.86.52's password:
Last login: Fri Aug  3 17:01:03 2007
[root@jns3 ~]# system-config-securitylevel-tui
```

Figure 7-31 Starting system-config-secureitylevel-tui

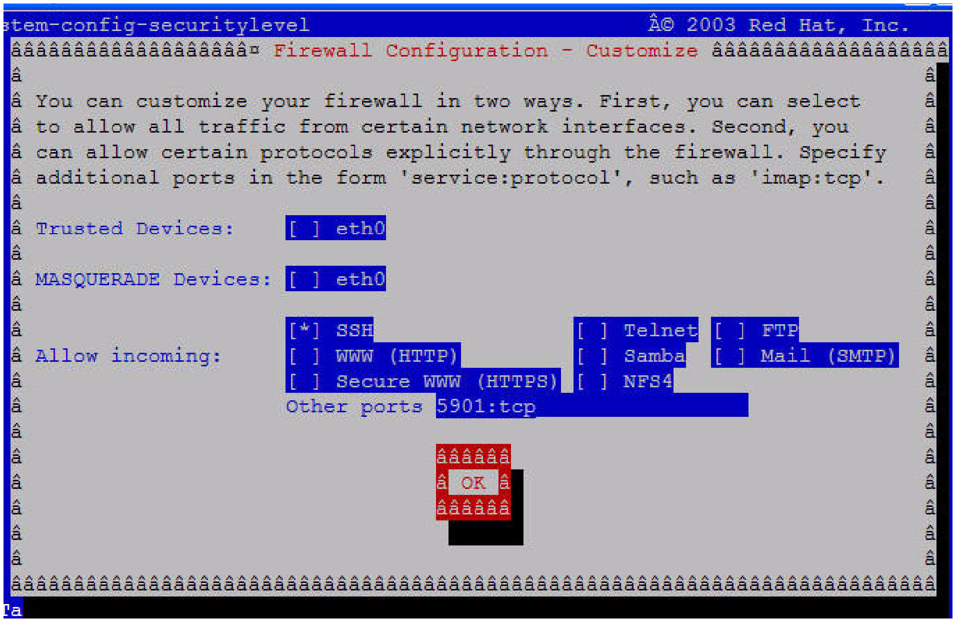
Select **Disable** (Figure 7-32) in order to disable the firewall, or select **Customize**.



```
Firewall Configuration

A firewall protects against unauthorized
network intrusions. Enabling a firewall blocks
all incoming connections. Disabling a firewall
allows all connections and is not recommended.

   Security Level: (*) Enabled ( ) Disabled

                 SELinux: Enforcing
                          Permissive
                          Disabled
```

Figure 7-32 Firewall settings

If Customize is selected, the panel shown in Figure 7-33 is presented.



```
stem-config-securitylevel                                    Â© 2003 Red Hat, Inc.
ââââââââââââââââââââ¤ Firewall Configuration - Customize ââââââââââââââââââââ
â                                                                          â
â You can customize your firewall in two ways. First, you can select       â
â to allow all traffic from certain network interfaces. Second, you        â
â can allow certain protocols explicitly through the firewall. Specify     â
â additional ports in the form 'service:protocol', such as 'imap:tcp'.     â
â                                                                          â
â Trusted Devices:     [ ] eth0                                            â
â                                                                          â
â MASQUERADE Devices:  [ ] eth0                                            â
â                                                                          â
â                      [*] SSH                   [ ] Telnet [ ] FTP        â
â Allow incoming:      [ ] WWW (HTTP)            [ ] Samba  [ ] Mail (SMTP)â
â                      [ ] Secure WWW (HTTPS)    [ ] NFS4                  â
â                      Other ports 5901:tcp                                â
â                                                                          â
â                                  ââââââ                                  â
â                                  â OK â                                  â
â                                  ââââââ                                  â
â                                                                          â
â                                                                          â
ââââââââââââââââââââââââââââââââââââââââââââââââââââââââââââââââââââââââââââ
Ta
```

Figure 7-33 Modifying firewall settings for VNC

To enable the VNC connections, modify the line Other ports to add vnc display:1 which is port 5901. Any other ports that may be needed can also be opened at this time. When done, select **OK** on this window and the next window. You should now be back at your prompt. Start the VNC server (Figure 7-34).

```
[root@jns3 ~]# vncserver :1

New 'jns3:1 (root)' desktop is jns3:1

Starting applications specified in /root/.vnc/xstartup
Log file is /root/.vnc/jns3:1.log
```

Figure 7-34 Starting VNCSERVER on display 1

7.3 Adding SCSI disks

Since this system was installed using the scsi disk, the /etc/zfcp.conf file (Figure 7-35) has been created, and the /etc/modprobe.conf file (Figure 7-36) has been updated to load the alias scsi_hostadapter and the zfcp module.

```
[root@jns3 etc]# cat /etc/zfcp.conf
0.0.6016 0x5005076300c300aa 0x5023000000000000
```

Figure 7-35 /etc/zfcp.conf

```
[root@jns3 etc]# cat /etc/modprobe.conf
alias eth0 qeth
options dasd_mod dasd=299
alias scsi_hostadapter zfcp
```

Figure 7-36 /etc/modprobe.conf

Note: If this is the first FCP disk configured, the file /etc/zfcp.conf needs to be created and the modprobe.conf file needs to be edited to add the line:

```
alias scsi_hostadapter zfcp
```

7.3.1 Adding an additional SCSI FCP LUN

There are just a few steps that need to be done to add another FCP disk to a system that already has FCP scsi configured:

- Make the FCP Paths available. Do one of the following?
  - Update the z/VM User Directory (Figure 7-37 on page 141, see Note).
  - Dynamically add the path via the z/VM CP ATTACH command (Figure 7-38 on page 141).
- Update the file /etc/zfcp.conf (Figure 7-39 on page 142).
- Add the scsi LUN information manually (Figure 7-40 on page 142).

Or

- Run the script /sbin/zfcpconf.sh and have the script do it for you (Figure 7-41 on page 143).

Adding the new dedicated device 6116 to the user directory for this guest (Figure 7-37).

```
USER JNS3 JNS3   512M 1024M G
   MACHINE ESA
   LOADDEV PORT 5005076300c300aa
   LOADDEV LUN 5023000000000000
   CONSOLE 001F 3215 T
   SPOOL 000C 2540 READER *
   SPOOL 000D 2540 PUNCH A
   SPOOL 000E 1403 A
   LINK MAINT 0190 0190 RR
   LINK MAINT 019D 019D RR
   LINK MAINT 019E 019E RR
   LINK TCPIP 0592 0592 RR
   DEDICATE 2708 2708
   DEDICATE 2709 2709
   DEDICATE 270A 270A
   MDISK 191 3390 1878 100 J20W02  MR READ     WRITE    MULTIPLE
   DEDICATE 6016 6016
   DEDICATE 6116 6116
```

Figure 7-37   Adding an additional FCP path to user direct

Note: The user directory is only read during the initial log on of the guest. So in order for any changes to take effect, the guest must be shut down and completely logged off. A cp disconnect will not suffice.

Using the CP ATTACH command (Figure 7-38) adds the device dynamically. The user directory still needs to be updated in order for the device to survive an log off.

```
Ready; T=0.01/0.01 14:33:12
cp att 6116 jns3
FCP 6116 ATTACHED TO JNS3 6116
```

Figure 7-38   Using CP attach command, from z/VM Maint user ID

Note: During the testing of this chapter we uncovered a bug with one of the utility scripts used by the zfcp installer. Because of this error we must add the scsi_id and scsi_lun fields to the zfcp.conf entries. Bugzilla 251719 has been opened and a fix should be available soon:

https://bugzilla.redhat.com/bugzilla/show_bug.cgi?id=251719

Edit the file /etc/zfcp.conf to add the new disk information. The scsi_id and scsi_lun fields must be added as a work around for the above-mentioned bug (Figure 7-39).

```
root@jns3 ~]# vi /etc/zfcp.conf
0.0.6016 0x0 0x5005076300c300aa 0x0 0x5023000000000000
0.0.6116 0x0 0x5005076300cb00aa 0x0 0x5024000000000000
```

Figure 7-39 Add new disk to /etc/zfcp.conf

Note: The fcp disk information must be in the zfcp.conf file in order for the disks to survive a reboot of Linux.

In Figure 7-40, the manual steps for adding another scsi LUN to your existing Linux guest are shown, or skip to Figure 7-41 on page 143 and run the script /sbin/zfcpconf.sh, which uses the information in zfcp.conf to make the configuration changes. We show the steps so that you know what is being done for you.

```
[root@jns3 zfcp]# cd /sys/bus/ccw/drivers/zfcp/0.0.6116
[root@jns3 0.0.6116]# echo 0x5005076300cb00aa > port_add
[root@jns3 0.0.6116]# cd 0x5005076300cb00aa
[root@jns3 0x5005076300cb00aa]# echo 0x5024000000000000 > unit_add
[root@jns3 0x5005076300cb00aa]# cd ..
[root@jns3 0.0.6116]# echo 1 > online
```

Figure 7-40 Adding LUN manually

To save extra typing when adding disks, after updating the /etc/zfcp.conf file skip to this step and run /sbin/zfcpconf.sh pick up the new changes dynamically (Figure 7-41).

```
[root@jns3 ~]# cd /sbin/
[root@jns3 sbin]# ./zfcpconf.sh
```

Figure 7-41 Running zfcpconf.sh

The new disk should now be available to the RHEL5 system. Verify this with the Linux command **lsscsi** to display your FCP disk information (Figure 7-42).

```
[root@jns3 ~]# lsscsi
[0:0:0:0]    disk    IBM      2105F20          .144  /dev/sda
[1:0:0:0]    disk    IBM      2105F20          .144  /dev/sdb
```

Figure 7-42 Use lsscsi to display new disk

7.3.2 Partitioning the new disk using fdisk

Now that the disk is configured it needs to be partitioned. The disk is partitioned by running the Linux **fsdisk** command. In the below example the disk is partitioned with one partition. You can have up to four partitions. Figure 7-43 shows partitioning of the disk using the **fsdisk** command.

```
[root@jns3 /]# fdisk /dev/sdb
Device contains neither a valid DOS partition table, nor Sun, SGI or OSF
disklabel
Building a new DOS disklabel. Changes will remain in memory only,
until you decide to write them. After that, of course, the previous
content won't be recoverable.

Command (m for help): m
Command action
   a   toggle a bootable flag
   b   edit bsd disklabel
   c   toggle the dos compatibility flag
   d   delete a partition
   l   list known partition types
   m   print this menu
   n   add a new partition
   o   create a new empty DOS partition table
   p   print the partition table
   q   quit without saving changes
   s   create a new empty Sun disklabel
   t   change a partition's system id
   u   change display/entry units
   v   verify the partition table
   w   write table to disk and exit
   x   extra functionality (experts only)

Command (m for help): n
Command action
   e   extended
   p   primary partition (1-4)
p
Partition number (1-4): 1
First cylinder (1-1022, default 1): 1
Last cylinder or +size or +sizeM or +sizeK (1-1022, default 1022): 1022

Command (m for help): p

Disk /dev/sdb: 5000 MB, 5000003584 bytes
154 heads, 62 sectors/track, 1022 cylinders
Units = cylinders of 9548 * 512 = 4888576 bytes

   Device Boot      Start         End      Blocks   Id  System
/dev/sdb1               1        1022     4878997   83  Linux

Command (m for help): w
The partition table has been altered!

Calling ioctl() to re-read partition table.
Syncing disks.
```

Figure 7-43 Using fdisk to create the disk partition

7.3.3 Formatting the new partition using mke2fs

In the next example the disk is formatted with the ext3 journaling file system by specifying -j (Example 7-1).

Example 7-1 Creating ext3 file system

```
[root@jns3 /]# mke2fs -j /dev/sdb1
mke2fs 1.39 (29-May-2006)
Filesystem label=
OS type: Linux
Block size=4096 (log=2)
Fragment size=4096 (log=2)
610432 inodes, 1219749 blocks
60987 blocks (5.00%) reserved for the super user
First data block=0
Maximum filesystem blocks=1249902592
38 block groups
32768 blocks per group, 32768 fragments per group
16064 inodes per group
Superblock backups stored on blocks:
        32768, 98304, 163840, 229376, 294912, 819200, 884736

Writing inode tables: done
Creating journal (32768 blocks): done
Writing superblocks and filesystem accounting information: done

This filesystem will be automatically checked every 32 mounts or
180 days, whichever comes first.  Use tune2fs -c or -i to override.
```

7.3.4 Mounting the new disk

At this point the disk can be mounted. In Example 7-2, the disk is mounted to the /mnt mountpoint. Verify the mount by issuing the command **`df -h`**.

Example 7-2 Mounting the new disk

```
[root@jns3 /]# mount -t ext3 /dev/sdb1 /mnt

[root@jns3 /]# df -h
Filesystem            Size  Used Avail Use% Mounted on
/dev/mapper/VolGroup00-LogVol00
                      4.2G  1.9G  2.2G  47% /
/dev/sda1              99M   13M   82M  14% /boot
tmpfs                 502M     0  502M   0% /dev/shm
/dev/sdb1             4.6G  138M  4.3G   4% /mnt
```

7.3.5 Edit fstab

Even though Linux can access the disk and use it, the mount does not survive a reboot. The mount still must be made permanent. Do this by editing the file /etc/fstab. We recommend that a backup of the fstab is done prior to editing. Edit the file using vi to add the mount point (Example 7-3).

Example 7-3 Copying and then editing /etc/fstab

```
[root@jns3 etc]# cp fstab fstab.save
[root@jns3 etc]# vi fstab

/dev/VolGroup00/LogVol00 /                       ext3    defaults        1 1
LABEL=/boot             /boot                   ext3    defaults        1 2
devpts                  /dev/pts                devpts  gid=5,mode=620  0 0
tmpfs                   /dev/shm                tmpfs   defaults        0 0
proc                    /proc                   proc    defaults        0 0
sysfs                   /sys                    sysfs   defaults        0 0
/dev/VolGroup00/LogVol01 swap                    swap    defaults        0 0
/dev/sdb1               /mnt                     ext3    defaults        1 3
```

7.3.6 Running commands mkinitrd and zipl

Example 7-4 and Example 7-5 on page 147 show commands **mkinitrd** and **zipl** running.

Example 7-4 Running mkinitrd

```
[root@jns3 ~]# cd /boot

[root@jns3 boot]# ls -al
total 7011
drwxr-xr-x  4 root root    1024 Aug 10 15:27 .
drwxr-xr-x 24 root root    4096 Aug 10 17:30 ..
-rw-------  1 root root    5632 Aug 10 15:27 bootmap
-rw-r--r--  1 root root   22179 Apr 16 16:00 config-2.6.18-8.1.3.el5
drwxr-xr-x  2 root root    1024 Aug 10 15:09 grub
-rw-------  1 root root 2637048 Aug 10 15:16 initrd-2.6.18-8.1.3.el5.img
drwx------  2 root root   12288 Aug 10 15:03 lost+found
-rw-r--r--  1 root root   48448 Apr 16 16:00 symvers-2.6.18-8.1.3.el5.gz
-rw-r--r--  1 root root  757438 Apr 16 16:00 System.map-2.6.18-8.1.3.el5
-rw-r--r--  1 root root    1464 Jan 23  2007 tape0
-rwxr-xr-x  1 root root 3637832 Apr 16 16:00 vmlinuz-2.6.18-8.1.3.el5

[root@jns3 boot]# mv initrd-2.6.18-8.1.3.el5.img
initrd-2.6.18-8.1.3.el5.img.old
```

```
[root@jns3 boot]# mkinitrd -v --with=scsi_mod --with=zfcp --with sd_mod
initrd-2
.6.18-8.1.3.el5.img 2.6.18-8.1.3.el5
Creating initramfs
Looking for driver for device sda2
Looking for deps of module ccw:t1731m03dt1732dm03: scsi_mod scsi_transport_fc
qd
io zfcp
Looking for deps of module scsi_mod
Looking for deps of module sd_mod: scsi_mod
Looking for deps of module scsi_transport_fc: scsi_mod
Looking for deps of module qdio
Looking for deps of module zfcp: scsi_mod scsi_transport_fc qdio
Looking for deps of module ide-disk
Looking for deps of module dasd_mod
Looking for deps of module dasd_eckd_mod: dasd_mod
Looking for deps of module dasd_fba_mod: dasd_mod
Using modules:     /lib/modules/2.6.18-8.1.3.el5/kernel/fs/jbd/jbd.ko
/lib/modul
es/2.6.18-8.1.3.el5/kernel/fs/ext3/ext3.ko
/lib/modules/2.6.18-8.1.3.el5/kernel/

******** snipped to reduce space *******
******** snipped to reduce space *******

Adding module scsi_mod
Adding module sd_mod
Adding module scsi_transport_fc
Adding module qdio
Adding module zfcp
Adding module dasd_mod with options dasd=299
Adding module dasd_eckd_mod
Adding module dasd_fba_mod
Adding module dm-mod
Adding module dm-mirror
Adding module dm-zero
Adding module dm-snapshot
```

We run the **zipl** command with the -v switch for verbose (Example 7-5). It is important at this time to watch the output from both commands closely to make sure that there are no errors.

Example 7-5 Running zipl

```
[root@jns3 boot]# zipl -V
Using config file '/etc/zipl.conf'
Target device information
  Device..........................: 08:00
  Partition.......................: 08:01
  Device name....................: sda
  Type...........................: disk partition
  Disk layout....................: SCSI disk layout
  Geometry - heads...............: 154
  Geometry - sectors.............: 62
  Geometry - cylinders...........: 1022
  Geometry - start...............: 63
  File system block size.........: 1024
  Physical block size............: 512
  Device size in physical blocks..: 208782
Building bootmap in '/boot/'
Building menu 'rh-automatic-menu'
Adding #1: IPL section 'linux' (default)
  kernel image......: /boot/vmlinuz-2.6.18-8.1.3.el5
  kernel parmline...: 'root=/dev/VolGroup00/LogVol00'
  initial ramdisk...: /boot/initrd-2.6.18-8.1.3.el5.img
  component address:
    kernel image....: 0x00010000-0x003783ff
    parmline........: 0x00001000-0x000011ff
    initial ramdisk.: 0x00800000-0x00a83dff
    internal loader.: 0x0000a000-0x0000a3ff
Preparing boot device: sda.
Detected SCSI PCBIOS disk layout.
Writing SCSI master boot record.
Syncing disks...
Done.
```

The Linux guest can now be safely shut down and then brought back up. The additional FCP scsi disk should be available on /mnt.

8

FCP multipathing on RHEL 5

This chapter provides details for accessing, configuring, and using FCP multipathing on Red Hat Enterprise Linux 5 (RHEL 5).

Topics include:

- Adding the additional LUN devices
- Partitioning the new disks
- Running mkinitrd
- Running zipl
- Creating and editing the /etc/mdadm.conf file
- Creating the multipath devices using mdadm
- Adding a file system to the new disks
- Mount the file systems at startup

8.1 RHEL 5 multipath implementation

Red Hat Enterprise Linux 5 implements multipathing over FCP using the multiple device administration (mdadm) driver. The mdadm implementation does not load balance among the available paths to volumes. Rather, it uses a primary path to a volume, then fails over to a secondary path if there are any problems with the primary path or if there is path saturation. When an active path fails, the md subsystem detects the failure, marks the path failed, then makes a secondary path active. If the failed path comes back, the md subsystem recognizes this and brings it back as a secondary path.

8.2 Configuring multipathing in RHEL 5

In Figure 8-1 the SCSI LUN’s sda and sdb have already been configured. We now need to add the sdc device and set up multipathing to the LUN 0x5024000000000000. Figure 8-1 is an example of our desired configuration. The device 0.0.6216 has already been configured in the user direct entry for this guest and is available to the guest.

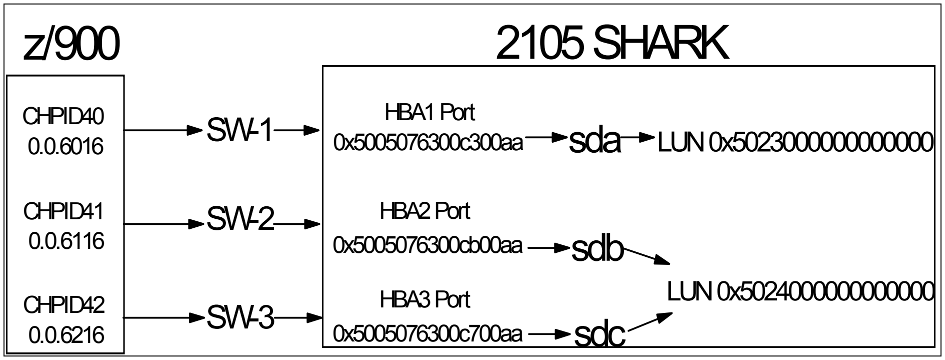
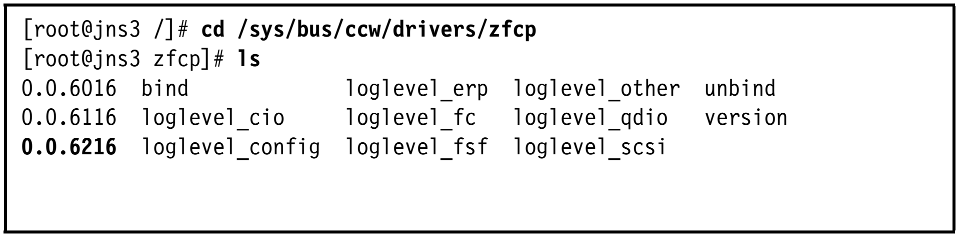


Figure 8-1 Hardware layout

Note: The two devices used will be 6116 and 6216. Each device is on a separate channel, separate switch. Also separate HBA ports and separate storage controllers on the 2105 storage device to avoid a single point of failure. The same LUN 0x5024000000000000 is accessible from both ports on the storage device.

8.2.1 Adding the additional LUN

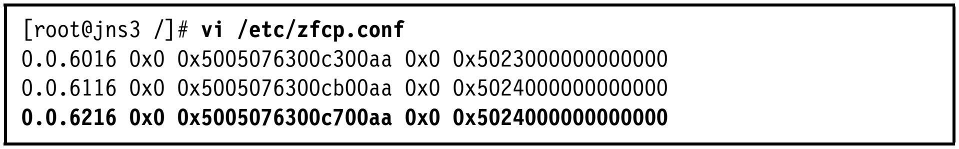
In the next example the device 0.0.6216 has already been added to the z/VM user directory entry for this guest and the guest has been logged off and back on. Linux has been IPLed and the new device has been sensed by RHEL 5. This can verified by doing an ls on the directory /sys/bus/ccw/drivers/zfcp. You can see that there is now a directory for device 0.0.6216 (see Figure 8-2).



```
[root@jns3 /]# cd /sys/bus/ccw/drivers/zfcp
[root@jns3 zfcp]# ls
0.0.6016  bind             loglevel_erp  loglevel_other  unbind
0.0.6116  loglevel_cio     loglevel_fc   loglevel_qdio   version
0.0.6216  loglevel_config  loglevel_fsf  loglevel_scsi
```

Figure 8-2 Verifying the new device has been sensed

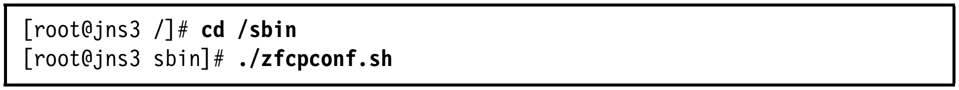
Now that it has been verified that RHEL 5 has sensed the new path, the /etc/zfcp.conf file needs to be edited to add the new port and LUN information (see Figure 8-3).



```
[root@jns3 /]# vi /etc/zfcp.conf
0.0.6016 0x0 0x5005076300c300aa 0x0 0x5023000000000000
0.0.6116 0x0 0x5005076300cb00aa 0x0 0x5024000000000000
0.0.6216 0x0 0x5005076300c700aa 0x0 0x5024000000000000
```

Figure 8-3 Edit the /etc/zfcp.conf file

Once zfcp.conf is updated the script /sbin/zfcpconf.sh is run to configure the new disk. This script reads the /etc/zfcp.conf file and builds the necessary configuration files (see Figure 8-4).



```
[root@jns3 /]# cd /sbin
[root@jns3 sbin]# ./zfcpconf.sh
```

Figure 8-4 Run the script /sbin/zfcpconf.sh

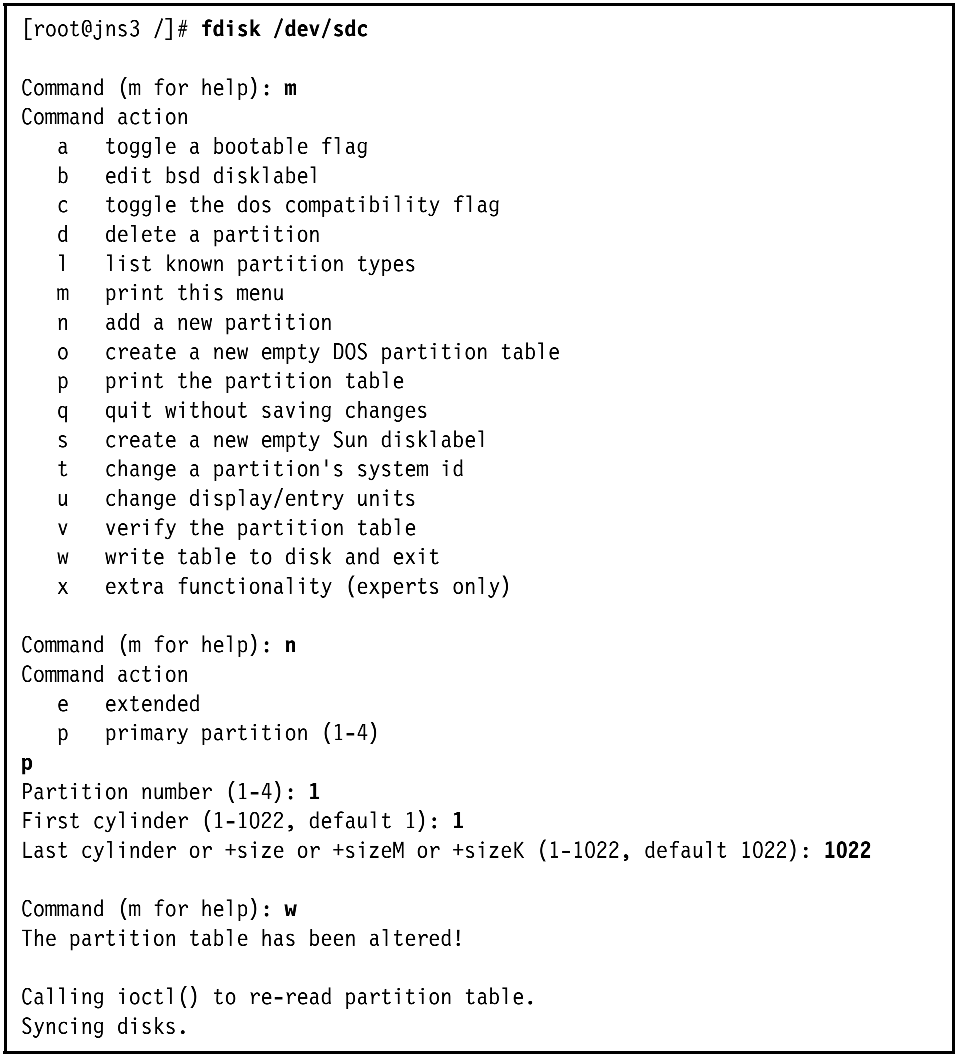
After running the script, the command **lsscsi** should be run to view the FCP disks configured on the guest, making sure that the new disk has been configured (Figure 8-5).

```
[root@jns3 /]# lsscsi
[0:0:0:0]    disk    IBM      2105F20          .144  /dev/sda
[1:0:0:0]    disk    IBM      2105F20          .144  /dev/sdb
[2:0:0:0]    disk    IBM      2105F20          .144  /dev/sdc
```

Figure 8-5   Using lsscsi to view the FCP disks configured on the guest

8.2.2 Running fdisk to partition the new disks

At this point the hardware has been configured and we need to add a partition to the disk using fdisk (Figure 8-6).



```
[root@jns3 /]# fdisk /dev/sdc

Command (m for help): m
Command action
   a   toggle a bootable flag
   b   edit bsd disklabel
   c   toggle the dos compatibility flag
   d   delete a partition
   l   list known partition types
   m   print this menu
   n   add a new partition
   o   create a new empty DOS partition table
   p   print the partition table
   q   quit without saving changes
   s   create a new empty Sun disklabel
   t   change a partition's system id
   u   change display/entry units
   v   verify the partition table
   w   write table to disk and exit
   x   extra functionality (experts only)

Command (m for help): n
Command action
   e   extended
   p   primary partition (1-4)
p
Partition number (1-4): 1
First cylinder (1-1022, default 1): 1
Last cylinder or +size or +sizeM or +sizeK (1-1022, default 1022): 1022

Command (m for help): w
The partition table has been altered!

Calling ioctl() to re-read partition table.
Syncing disks.
```

Figure 8-6 Partitioning the new disk with fdisk

8.2.3 Running mkinitrd and zipl

Up to this point the new disks have been added dynamically. If Linux is shut down or rebooted, the changes made will be gone. We recommend that mkinitrd and zipl are run now. Then reboot the server to make sure that this portion of the procedure went well. Pay close attention to the **mkinitrd** and **zipl** command output, looking for any errors.

If you cannot reboot the server now, you can still proceed (Figure 8-7 on page 155).

Note: In Figure 8-7 the **mv** command and the **mkinitrd** command both span two lines. This is because of the length of the commands and the width of this document. In reality these are one long single command.

```
[root@jn3 sbin]# cd /boot

[root@jn3 boot]# ls -al
total 7012
drwxr-xr-x  4 root root    1024 Aug 15 16:54 .
drwxr-xr-x 24 root root    4096 Aug 16 12:23 ..
-rw-------  1 root root    5632 Aug 15 16:54 bootmap
-rw-r--r--  1 root root   22179 Apr 16 16:00 config-2.6.18-8.1.3.el5
drwxr-xr-x  2 root root    1024 Aug 15 16:31 grub
-rw-------  1 root root 2637942 Aug 15 16:38 initrd-2.6.18-8.1.3.el5.img
drwx------  2 root root   12288 Aug 15 16:25 lost+found
-rw-r--r--  1 root root   48448 Apr 16 16:00 symvers-2.6.18-8.1.3.el5.gz
-rw-r--r--  1 root root  757438 Apr 16 16:00 System.map-2.6.18-8.1.3.el5
-rw-r--r--  1 root root    1464 Jan 23  2007 tape0
-rwxr-xr-x  1 root root 3637832 Apr 16 16:00 vmlinuz-2.6.18-8.1.3.el5

[root@jn3 boot]# mv initrd-2.6.18-8.1.3.el5.img
initrd-2.6.18-8.1.3.el5.img.save

[root@jn3 boot]# mkinitrd -v --with=scsi_mod --with=zfcp --with=sd_mod
initrd-2.6.18-8.1.3.el5.img 2.6.18-8.1.3.el5

Creating initramfs
Looking for deps of module uhci-hcd
Looking for deps of module ext3: jbd
Looking for deps of module jbd
-----
---- snipped to save space ----
-----
Looking for driver for device sda2
Looking for deps of module ccw:t1731m03dt1732dm03: scsi_mod copy from Adding
module dm-snapshot


[root@jn3 boot]# zipl
Using config file '/etc/zipl.conf'
Building bootmap in '/boot/'
Building menu 'rh-automatic-menu'
Adding #1: IPL section 'linux' (default)
Preparing boot device: sda.
Done.
```

Figure 8-7 Running mkinitrd and zipl before rebooting

At this point Linux guest could be safely rebooted and the disks should be available when you come back up. You can verify this by running `lsscsi` again.

8.2.4 Configuring mdadm to manage the multipath device.

Create a file called /etc/mdadm.conf if it does not already exist. This file is used to keep track of which devices are being used by which arrays. Note in Figure 8-8 how the SCSI devices are enumerated in the DEVICE list by enclosing a list in brackets. This file is also used at IPL time to bring the arrays online before their file systems are mounted.

```
[root@jns3 /]# vi /etc/mdadm.conf
DEVICE /dev/sd[bc]1
ARRAY /dev/md0 devices=/dev/sdb1,/dev/sdc1
```

Figure 8-8 Create the /etc/mdadm.conf file

Check the contents of /proc/mdstat, as shown in Figure 8-9, to verify that there are no md devices already defined.

```
[root@jn3 etc]# cat /proc/mdstat
Personalities :
unused devices: <none>
```

Figure 8-9 cat /proc/mdstat

8.2.5 Creating the multipath device using mdadm

Figure 8-10 shows the commands used to create the array.

```
[root@jns3 /]# mdadm -C /dev/md0 --level=multipath --raid-devices=2
/dev/sdb1 /dev/sdc1
mdadm: /dev/sdb1 appears to contain an ext2fs file system
    size=4878996K  mtime=Mon Aug 13 14:18:42 2007
mdadm: /dev/sdc1 appears to contain an ext2fs file system
    size=4878996K  mtime=Mon Aug 13 14:18:42 2007
Continue creating array? yes
mdadm: array /dev/md0 started.
```

Figure 8-10 Create the array and the /etc/mdadm.conf file

We use the command, **`mdadm -C /dev/md0 --level=multipath --raid-devices=2 /dev/sdb1 /dev/sdc1`**, where:

- **`-C`** is the actual create command.
- /dev/md0 is the name of the multipath device.

- --level=multipath means that we are creating a multipath device instead of an array.
- --raid-devices=2 indicates that there will be two devices.
- /dev/sdb1 and /dev/sdc1 are the names of the FCP devices we use.

Verify that **mdadm** has created the multipath device by issuing the command **mdadm --detail /dev/md0**, as shown in Figure 8-11.

```
[root@jns3 /]# mdadm --detail /dev/md0
/dev/md0:
        Version : 00.90.03
  Creation Time : Mon Aug 13 14:50:39 2007
     Raid Level : multipath
     Array Size : 4878912 (4.65 GiB 5.00 GB)
   Raid Devices : 2
  Total Devices : 2
Preferred Minor : 0
    Persistence : Superblock is persistent

    Update Time : Mon Aug 13 14:50:39 2007
          State : clean
 Active Devices : 2
Working Devices : 2
 Failed Devices : 0
  Spare Devices : 0

           UUID : 2ee09eb1:8b342ebb:03296550:bb0461ee
         Events : 0.3

    Number   Major   Minor   RaidDevice State
       0       8       33        0      active sync   /dev/sdc1
       1       8       17        1      active sync   /dev/sdb1
```

Figure 8-11 Verifying the new multipath device

8.2.6 Creating a file system on the devices

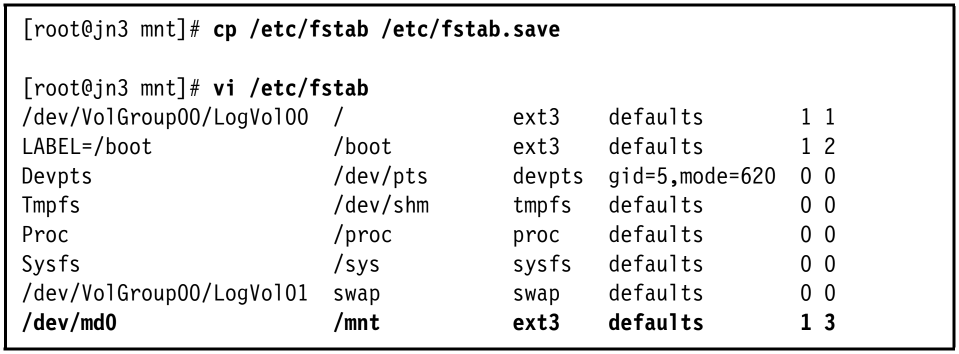
Create a file system on the new device and mount it as shown in Figure 8-12.

```
[root@jns3 /]# mke2fs -j /dev/md0
mke2fs 1.39 (29-May-2006)
Filesystem label=
OS type: Linux
Block size=4096 (log=2)
Fragment size=4096 (log=2)
610432 inodes, 1219728 blocks
60986 blocks (5.00%) reserved for the super user
First data block=0
Maximum filesystem blocks=1249902592
38 block groups
32768 blocks per group, 32768 fragments per group
16064 inodes per group
Superblock backups stored on blocks:
        32768, 98304, 163840, 229376, 294912, 819200, 884736

Writing inode tables: done
Creating journal (32768 blocks): done
Writing superblocks and filesystem accounting information: done

This filesystem will be automatically checked every 29 mounts or
180 days, whichever comes first.  Use tune2fs -c or -i to override.

[root@jns3 boot]# mount /dev/md0 /mnt

[root@jns3 boot]# df -h
Filesystem            Size  Used Avail Use% Mounted on
/dev/mapper/VolGroup00-LogVol00
                      4.2G  1.9G  2.1G  47% /
/dev/sda1              99M   15M   79M  16% /boot
tmpfs                 502M     0  502M   0% /dev/shm
/dev/md0              4.6G  138M  4.3G   4% /mnt
```

Figure 8-12 Creating and mounting a file system

8.2.7 Mount the file systems at startup

Add an entry to /etc/fstab to mount the new RAID device automatically at IPL time. As a precaution and to help in any recovery procedures it is a good idea to back up the current fstab file prior to any editing (Figure 8-13).



```
[root@jn3 mnt]# cp /etc/fstab /etc/fstab.save

[root@jn3 mnt]# vi /etc/fstab
/dev/VolGroup00/LogVol00  /            ext3    defaults        1 1
LABEL=/boot               /boot        ext3    defaults        1 2
Devpts                    /dev/pts     devpts  gid=5,mode=620  0 0
Tmpfs                     /dev/shm     tmpfs   defaults        0 0
Proc                      /proc        proc    defaults        0 0
Sysfs                     /sys         sysfs   defaults        0 0
/dev/VolGroup00/LogVol01  swap         swap    defaults        0 0
/dev/md0                  /mnt         ext3    defaults        1 3
```

Figure 8-13 The new /etc/fstab file

8.2.8 Restart the Linux guest

Shut down and restart the Linux guest. Ensure that the new SCSI devices are available by viewing the startup messages or the /var/log/messages file.

9

FCP security topics

This chapter provides details for securing the FCP network, using LUN masking, zoning, and binding. Topics include:

- LUN masking
- Zoning
- Binding

9.1 LUN masking

LUN masking provides the ability to restrict access to a LUN to a specific group of source WWPNs. For every LUN, we define the WWPNs that are allowed to access it. Definitions are defined to and enforced by the storage controller (Figure 9-1).

> **Important:** For IBM storage products, the default policy is to restrict access to all LUNs. Without explicit access, a WWPN *cannot* access any LUN. (Each WWPN must be granted explicit access to a LUN.) This principle guarantees the maximum possible security.

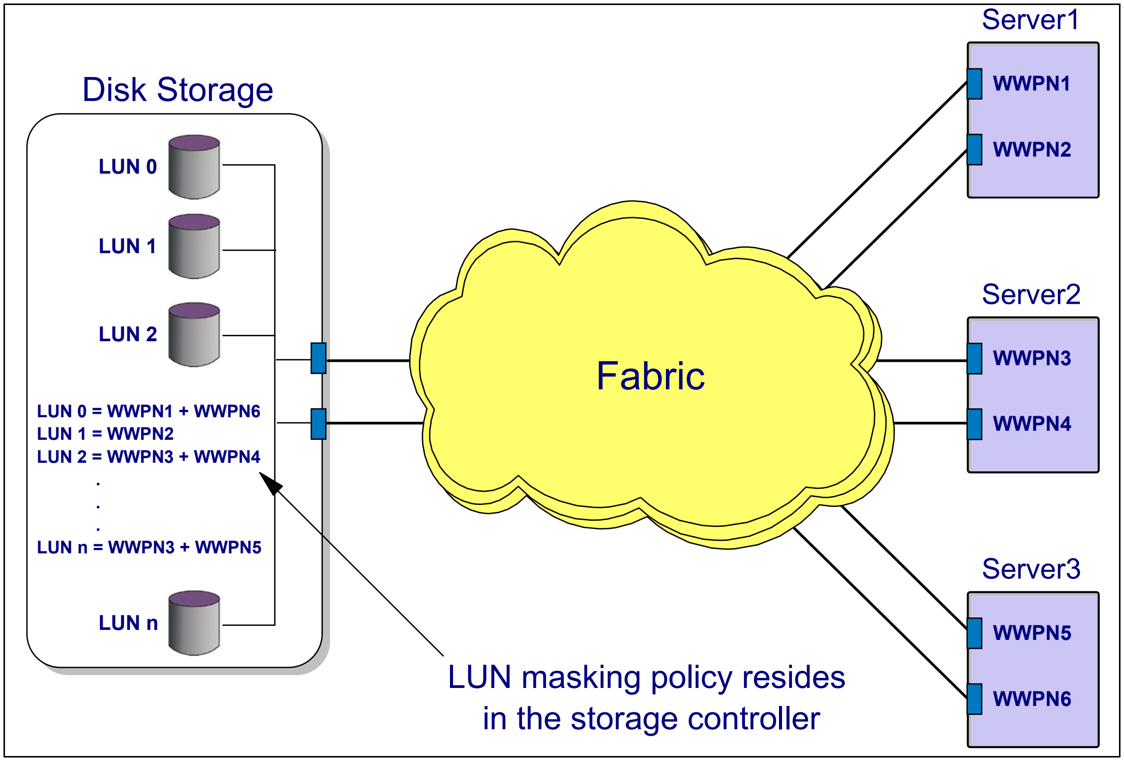


Figure 9-1 LUN masking example

In this example, only WWPN1 and WWPN6 can access LUN 0. WWPN3 may not access LUN 0 (based on the illustrated rules).

Note: If the WWPN of an attached node is changed, the LUN masking definitions must also change to reflect the new WWPN.

9.2 Zoning

Zoning provides the ability to enable or disable communication between nodes in the fabric. Zones are defined in the switch, and are groupings of WWNs (WWPNs and WWNNs). Communication between nodes is permitted or denied based on membership in the corresponding zone. Zoning allows fabric segmentation. Reasons for zoning can include:

- Restricting access to devices with sensitive data (confidential information or system resources, for example).
- Separating FCP and FC traffic within the fabric.
- Controlling the number of paths between servers and devices. This allows management of the traffic within a fabric, and simplifies problem isolation.

Zoning can be established in two ways:

- *Hard zoning* uses the switch identifier and port numbers to define the zones within a fabric.
- *Soft zoning* (or *name server zoning*) defines zones based on the WWPNs of attached devices.

Figure 9-2 illustrates an example of soft zoning.

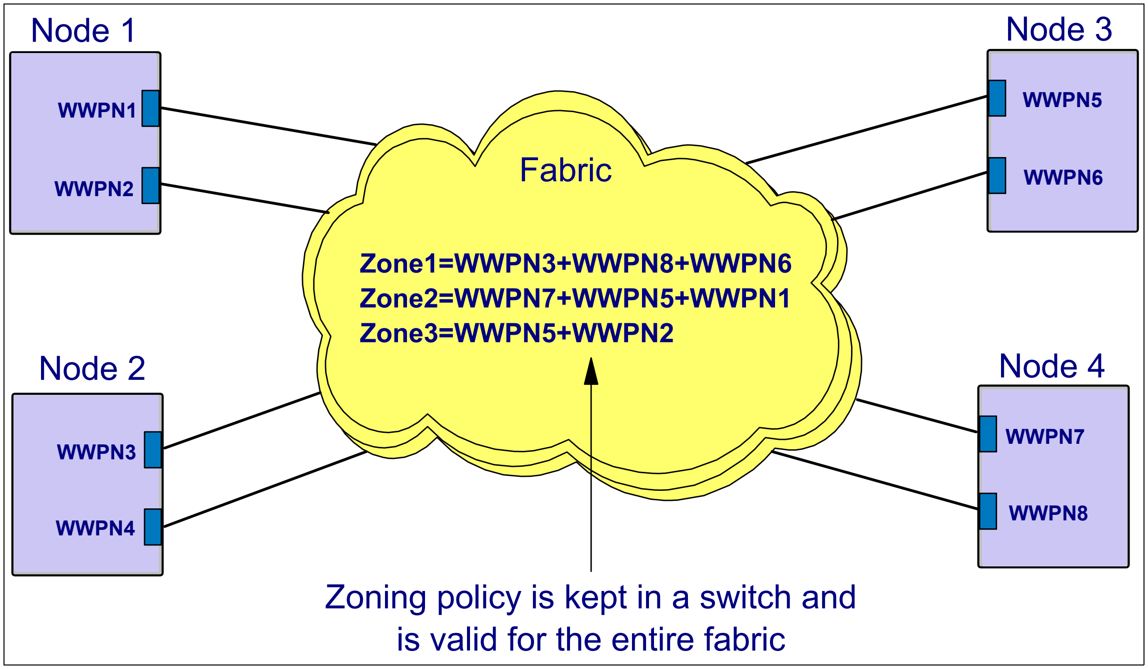


Figure 9-2 Soft zoning example

In Figure 9-2, the members of Zone1 are:

- WWPN3 (on node 2)
- WWPN8 (on node 4)
- WWPN6 (on node 3)

Because they are in the same zone, the WWPNs can communicate with each other. Conversely, WWPN3 *cannot* communicate with WWPN1. (No zone is defined in which both are members.)

Soft zoning offers some advantages over hard zoning:

- With hard zoning, exchanging two cables at the switch could cause the attached devices to move to different zones (and enable access to normally restricted nodes). Therefore, soft zoning can be considered more secure than hard zoning.
- If cables are moved at the switch, isolating problems in fabric can become more difficult.

However, with hard zoning, changing the FCP adapter for any attached device does not require a change to the zoning definition.

Note: Unlike LUN masking, the default policy for zoning enables all attached ports to communicate with each other. It is the administrator’s responsibility to isolate nodes by defining appropriate zones.

9.2.1 Zone definition

Zones are defined on the SAN switch. In Figure 9-3, we show the zone configuration on a Brocade 2109-F32 switch using the **cfgshow** command.

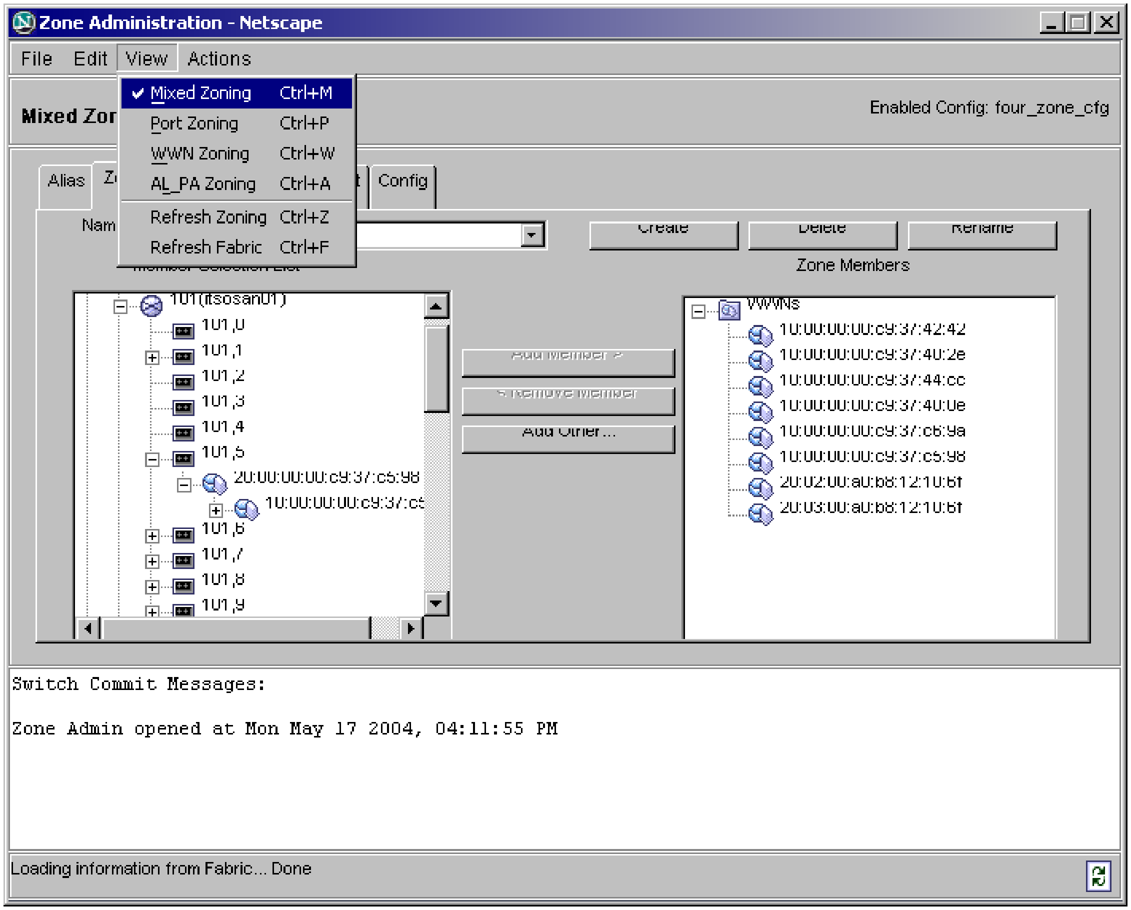
```
itsosan01:admin> cfgshow
Defined configuration:
 cfg:   four_zone_cfg
                gpfs1_zone; zseries_zone; gpfs_trans
 zone:  gpfs1_zone
                .................
 zone:  gpfs_trans
               ..................
 zone:  zseries_zone
                50:05:07:64:01:40:13:e2; 50:05:07:63:00:cd:95:89;
                50:05:07:63:00:c1:95:89; 50:05:07:64:01:00:13:e2;

Effective configuration:
 cfg:   four_zone_cfg
 zone:  gpfs1_zone
                ..................
 zone:  gpfs_trans
                ..................
 zone:  zseries_zone
                50:05:07:64:01:40:13:e2
                50:05:07:63:00:cd:95:89
                50:05:07:63:00:c1:95:89
                50:05:07:64:01:00:13:e2
```

Figure 9-3 Zoning definitions on a Brocade 2109-F32 switch

In this configuration:

- The 50:05:07:64:01:00:13:e2 and 50:05:07:64:01:40:13:e2 WWPNs belong to an IBM System z FCP adapter.
- The 50:05:07:63:00:cd:95:89 and 50:05:07:63:00:c1:95:89 WWPNs belong to an IBM ESS Model 2105-800.

In Figure 9-4, we show a graphic user interface to configure zones on the switch.



Enabled Config: four_zone_cfg

```
Switch Commit Messages:

Zone Admin opened at Mon May 17 2004, 04:11:55 PM
```

Loading information from Fabric... Done

Figure 9-4 Defining zones using a GUI

9.3 Binding

Binding is a method to permit or deny access to ports, nodes, or other switches based on WWPNs or WWNNs. Binding is defined at a switch.

> **Note:** Zoning applies to the entire fabric. Binding applies to a single switch (or port on the switch). Binding definitions are valid for a single switch (not the entire fabric).

Various binding methods are available:

- Port binding

  With port binding, attached nodes are bound to the switch port by their WWPNs. Swapping cables or replacing FCP adapters requires modification of the binding policy.

- Switch binding

  Switch binding is more flexible than port binding. The WWNN is mapped to the switch. This enables cable swapping and card replacement without changing the binding policy.

- Fabric binding

  Fabric binding prevents unauthorized switches from accessing the fabric. The implementation depends on the switch manufacturer.

To summarize:

- LUN masking is performed at the storage controller.

  This defines which LUNs may be accessed from source WWPNs.

- Zoning is performed at the switch.

  This defines access between source and target WWPNs for the entire fabric.

- Binding is also performed at the switch.

  However, binding is specific to a switch. It defines which WWPNs or WWNNs can attach to the switch.

10

The N_Port Virtualization feature

The N_Port ID Virtualization (NPIV) feature in a Fibre Channel Protocol (FCP) fabric is available with IBM System z EC and BC servers. NPIV allows a single FCP port to register multiple Worldwide Port Names (WWPN) with a fabric name server. Each registered WWPN is assigned a unique N_Port ID. With NPIV, a single FCP port can appear as multiple WWPNs in the FCP fabric.

This chapter gives an overview of the NPIV feature and demonstrates how to configure NPIV using a sample configuration as an example.

10.1 Overview of the NPIV feature

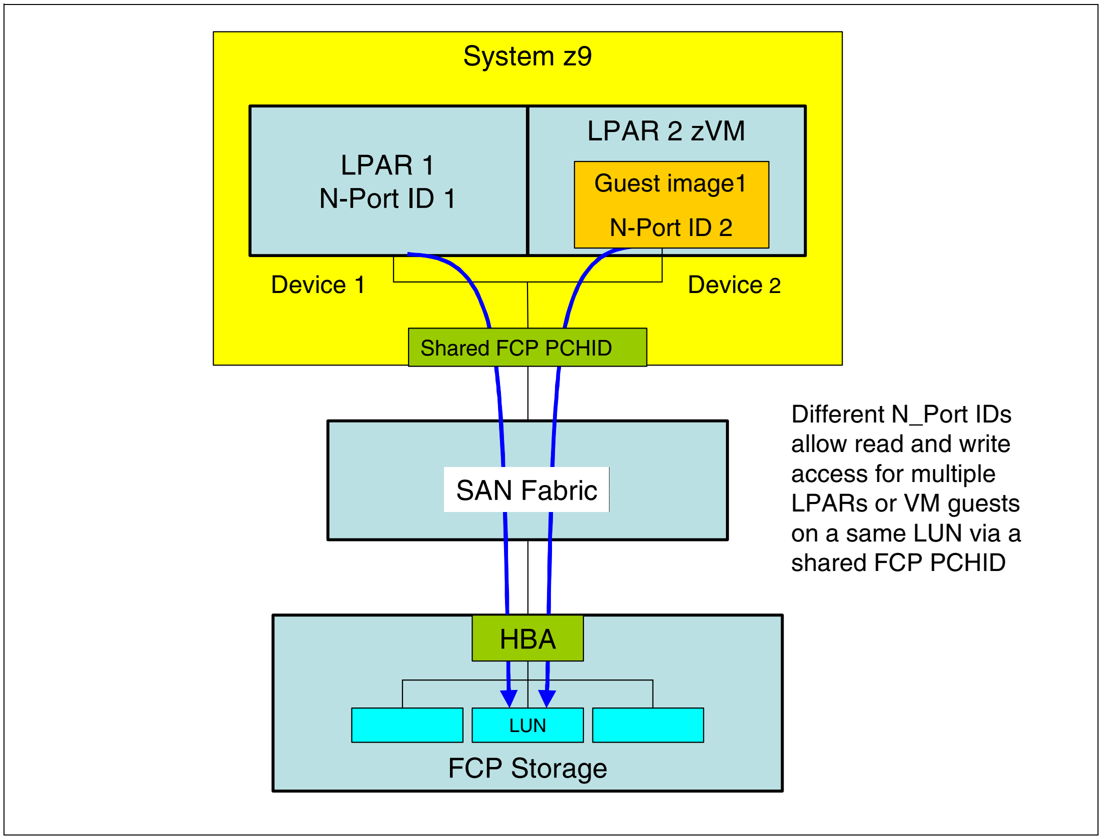
System z FCP channels require a FICON Express adapter. FCP channels can be shared by multiple LPARs. Each port on the adapter is assigned a permanent 64-bit WWPN by the manufacturer. This is used at Fabric Login (FLOGI).

Without the NPIV feature, each operating system image that has an FCP port is identified to the fabric by the permanent WWPN of the port. In this case, all operating system images sharing the port have the same access rights in the fabric. The permanent WWPN of the port determines:

- Zone membership for all images sharing the port
- Logical Unit Number (LUN) access rights for all images sharing the port

With the NPIV feature, the service element (SE) creates new WWPNs for the FCP port at FLOGI. A unique WWPN is then assigned to each FCP subchannel, or device, on the port. The generated NPIV WWPN is registered with the fabric switch and uniquely identifies each FCP subchannel for fabric zoning and LUN masking. Figure 10-1 on page 171 illustrates the NPIV feature.

> **Note:** A FCP port can be shared by LPARs running in either NPIV mode or non-NPIV mode. The IOCDS is same for both modes.



Different N_Port IDs allow read and write access for multiple LPARs or VM guests on a same LUN via a shared FCP PCHID

Figure 10-1 NPIV provides unique WWPNs to servers sharing an FCP port

In Figure 10-1, two LPARs share a single physical FCP port. Each instance registers with the name server. The NPIV WWPN is supported in the Fabric Discovery (FDISC) process.

During Power On Reset (POR) or dynamic I/O activation, each FCP subchannel is assigned a WWPN by the support element regardless of whether the LPAR is

NPIV enabled. If the LPAR is not enabled for NPIV, the microcode does not use the NPIV WWPNs. The SE retains, on its hard drive, the information about the assigned WWPN (to prevent the data from being lost if the system is shut down or the FCP adapter is replaced).

Each FCP subchannel per LPAR receives a different N_Port ID. This allows multiple LPARs and VM guests to read and write to the same Logical Unit Number using the same physical port. Without NPIV, writing to the same LUN over a shared port is not allowed.

10.1.1 Format of the NPIV WWPN

An NPIV WWPN is 64 bits in length. Its format is shown in Figure 10-2.

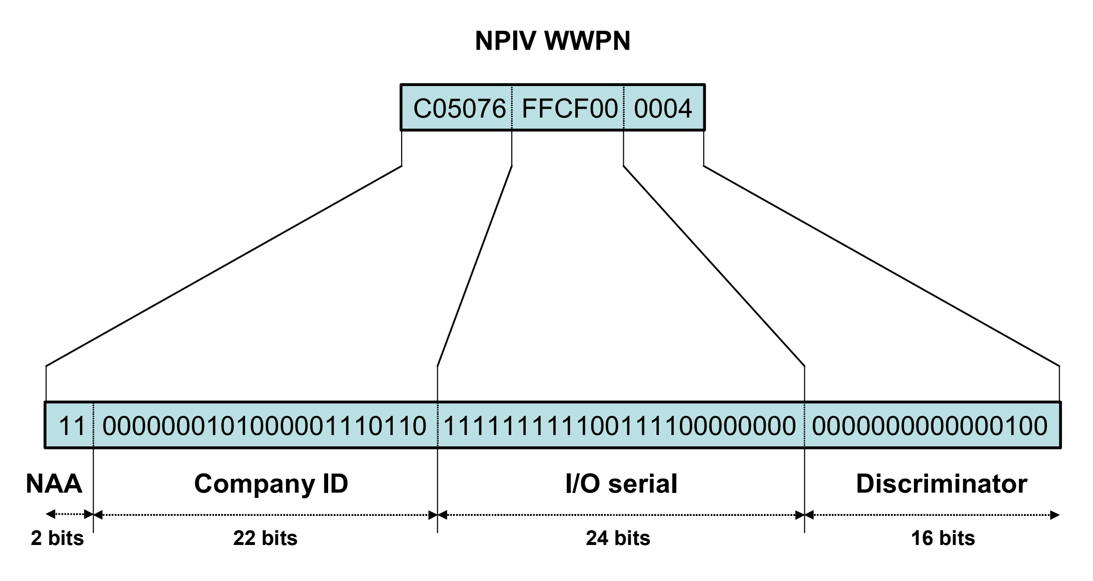


Figure 10-2 Format of an NPIV WWPN

The fields for an NPIV WWPN are:

- Network Address Authority (NAA)

  This two-bit field is always set to binary 11.

- Company ID

  This 22-bit field identifies the server manufacturer and is assigned by IEEE. For NPIV, this value differs from the company ID field in the permanent WWPN of the adapter. The value of the company ID is the same for IBM System z, System p™, and System i™ servers.

- I/O serial

  This 24-bit field uniquely identifies a specific server. A single I/O serial number is assigned to each System p and System i server. System z servers are assigned a range of 128 consecutive I/O serial numbers.

- Discriminator

  The 16-bit NPIV discriminator field is used to generate a unique NPIV WWPN when sharing an FCP port.

A 16-bit discriminator provides up to 64,000 NPIV WWPNs for a shared FCP port. Although this is sufficient for both System p and System i servers, System z systems require a larger address space. To increase the address space, the low order 7 bits of the I/O serial field are combined with the 16-bit discriminator field. The resulting 23-bit effective discriminator field increases the address space to over 8,000,000 unique WWPNs for a shared FCP port.

10.1.2 Configuration considerations

Some general recommendations for using NPIV include:

- Do not use more than 32 subchannels per physical channel in NPIV mode. Also, do not perform more than 128 total target logins (for example, in a configuration with 32 subchannels, limit the number of target logins to no more than an average of 4). Using more subchannels, target logins, or both can create timeouts.
- Zone each NPIV WWPN individually. This can reduce fabric traffic because all participants in the zone are notified when another N_Port joins or leaves the zone.
- Consider using multipathing performance and availability (multipathing is discussed in *Linux on zSeries: Fibre Channel Protocol Implementation Guide*, SG24-6344).
- Enable NPIV on the SAN switch before enabling it on the System z server. If NPIV is not enabled on the switch, the attempt to establish a connection to the fabric fails for all subchannels that are operated in NPIV mode.
- Be aware that each login from a NPIV-mode subchannel into a storage subsystem counts as a separate host login. The IBM TotalStorage® Enterprise Storage Server® (ESS) model 800 supports up to 16 Host Bus Adapters (HBAs). Each HBA supports up to 124 host logins. The ESS itself supports up to 512 logins. Consult the vendor's documentation for limits in your configuration.
- Switches typically limit the number of supported N_Port IDs. Because each NPIV WWPN is assigned an N_Port ID at login, this limit can be exceeded.

- Some switches limit the number of N_Port IDs that can be assigned to a physical port. This limit can be configurable on some switches.

10.2 Configuring the NPIV feature

To demonstrate how to configure NPIV, we used the sample configuration shown in Figure 10-3.

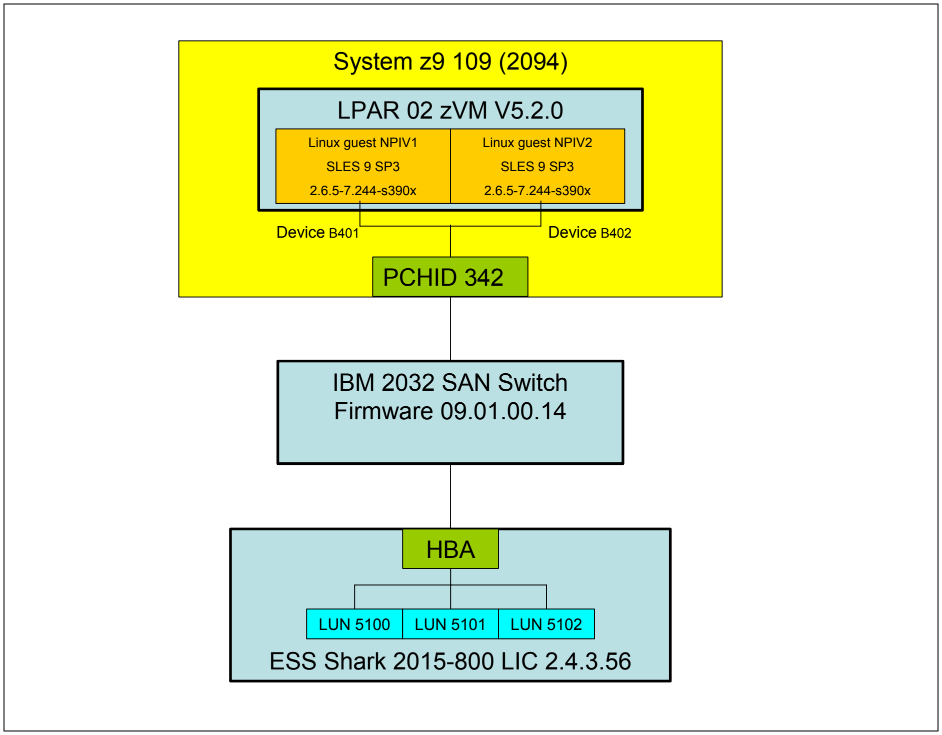


Figure 10-3 Configuration to demonstration NPIV

In this example, two Linux guests access SCSI disks over an NPIV-enabled FCP adapter. The steps we demonstrate include:

- Configuring NPIV on the SAN switch
- Configuring NPIV on the System z server
- Configuring fabric security
- Configuring the Linux server

10.2.1 Configuring NPIV on the SAN switch

Before enabling NPIV on the System z9 server, you must enable NPIV on the switch.

Note: The following examples are specific to our IBM 2032 switch. Other switches that support NPIV work similarly. Follow your switch vendor documentation to enable NPIV on your specific switch.

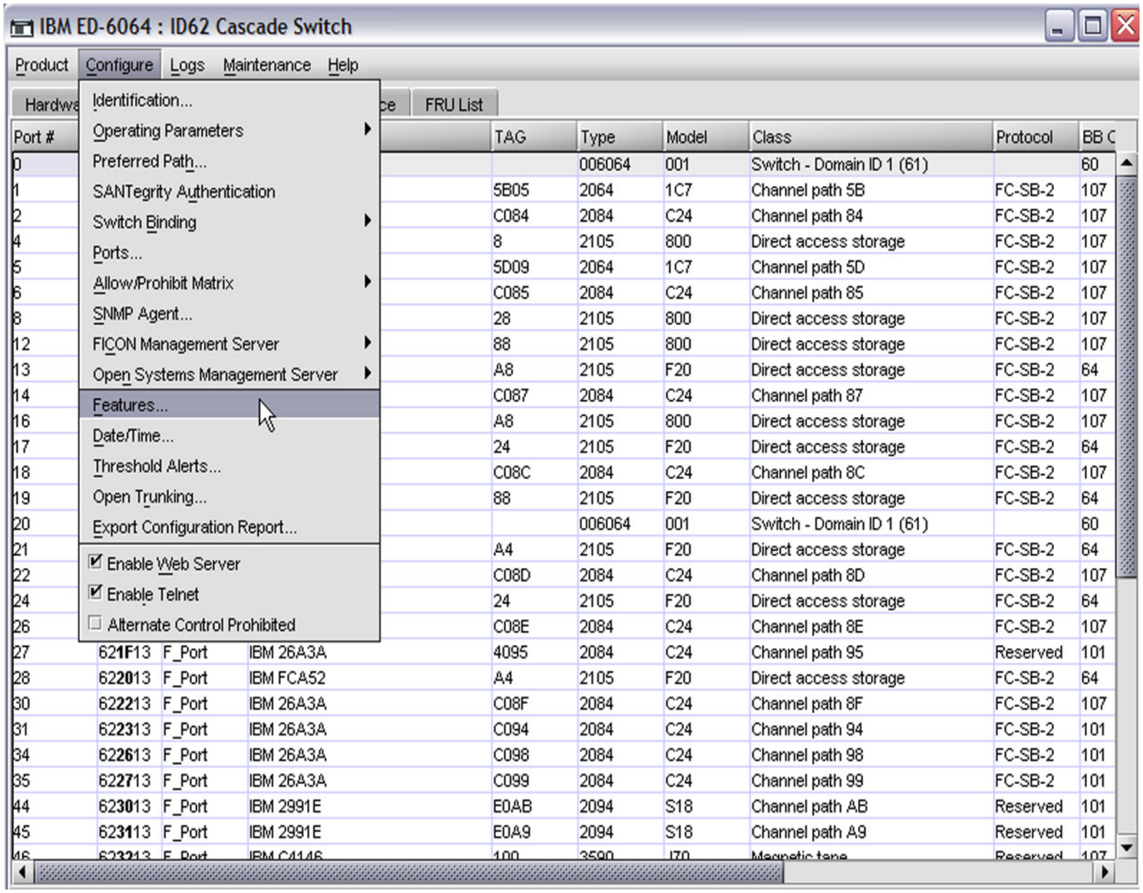
To enable NPIV on the switch:

1. Add the NPIV feature to the switch.
2. Activate the NPIV feature on switch.
3. Configuring NPIV on an individual switch port.

Add the NPIV feature to the switch

To add the NPIV feature to the switch:

1. Start the EFCM switch management application and log on to the switch. Select **Configure** → **Features**, as shown in Figure 10-4.



Port #				TAG	Type	Model	Class	Protocol	BB C
0					006064	001	Switch - Domain ID 1 (61)		60
1				5B05	2064	1C7	Channel path 5B	FC-SB-2	107
2				C084	2084	C24	Channel path 84	FC-SB-2	107
4				8	2105	800	Direct access storage	FC-SB-2	107
5				5D09	2064	1C7	Channel path 5D	FC-SB-2	107
6				C085	2084	C24	Channel path 85	FC-SB-2	107
8				28	2105	800	Direct access storage	FC-SB-2	107
12				88	2105	800	Direct access storage	FC-SB-2	107
13				A8	2105	F20	Direct access storage	FC-SB-2	64
14				C087	2084	C24	Channel path 87	FC-SB-2	107
16				A8	2105	800	Direct access storage	FC-SB-2	107
17				24	2105	F20	Direct access storage	FC-SB-2	64
18				C08C	2084	C24	Channel path 8C	FC-SB-2	107
19				88	2105	F20	Direct access storage	FC-SB-2	64
20					006064	001	Switch - Domain ID 1 (61)		60
21				A4	2105	F20	Direct access storage	FC-SB-2	64
22				C08D	2084	C24	Channel path 8D	FC-SB-2	107
24				24	2105	F20	Direct access storage	FC-SB-2	64
26				C08E	2084	C24	Channel path 8E	FC-SB-2	107
27	621F13	F_Port	IBM 26A3A	4095	2084	C24	Channel path 95	Reserved	101
28	622013	F_Port	IBM FCA52	A4	2105	F20	Direct access storage	FC-SB-2	64
30	622213	F_Port	IBM 26A3A	C08F	2084	C24	Channel path 8F	FC-SB-2	107
31	622313	F_Port	IBM 26A3A	C094	2084	C24	Channel path 94	FC-SB-2	101
34	622613	F_Port	IBM 26A3A	C098	2084	C24	Channel path 98	FC-SB-2	101
35	622713	F_Port	IBM 26A3A	C099	2084	C24	Channel path 99	FC-SB-2	101
44	623013	F_Port	IBM 2991E	E0AB	2094	S18	Channel path AB	Reserved	101
45	623113	F_Port	IBM 2991E	E0A9	2094	S18	Channel path A9	Reserved	101
46	623213	F_Port	IBM C4146	100	3590	J70	Magnetic tape	Reserved	107

Figure 10-4 Navigation to the Features configuration menu

2. In the Features configuration menu, click **New** and enter the NPIV feature key provided by the switch manufacturer. This menu is shown in Figure 10-5.

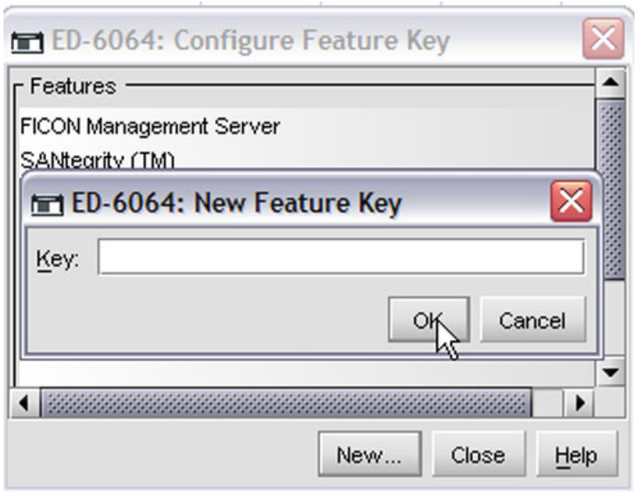
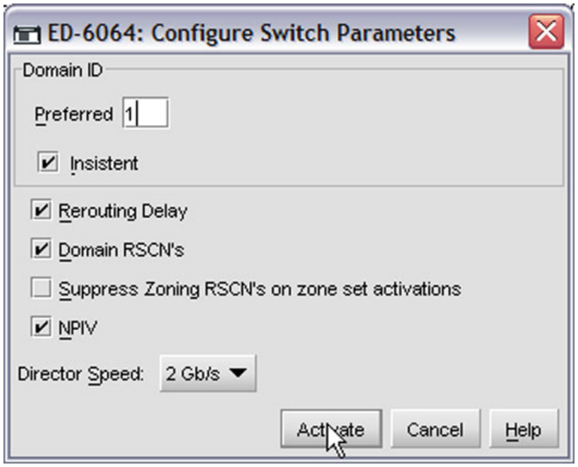


Figure 10-5 Provide the NPIV feature key

Activate the NPIV feature on switch

To activate the NPIV feature on the switch:

1. Select **Configure** → **Operating Parameters** → **Switch Parameters** to open the Configure Switch Parameters menu shown in Figure 10-6.


ED-6064: Configure Switch Parameters

Domain ID

Preferred 1

☑ Insistent

☑ Rerouting Delay

☑ Domain RSCN's

□ Suppress Zoning RSCN's on zone set activations

☑ NPIV

Director Speed: 2 Gb/s

Figure 10-6 Activate the NPIV feature on the switch

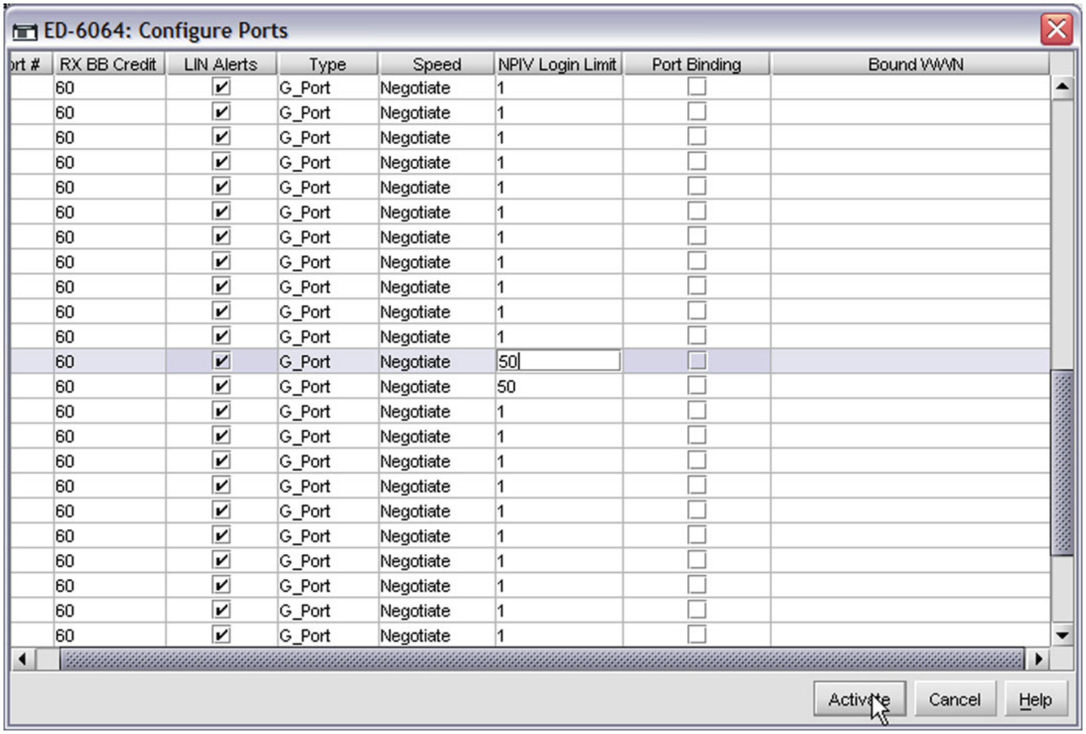
2. Select the NPIV option.

3. Click **Activate**.

Configuring NPIV on an individual switch port

To configure NPIV for an individual port:

1. Select **Configure** → **Ports** to open the port configuration menu shown in Figure 10-7.



rt #	RX BB Credit	LIN Alerts	Type	Speed	NPIV Login Limit	Port Binding	Bound WWN
	60	☑	G_Port	Negotiate	1	☐	
	60	☑	G_Port	Negotiate	1	☐	
	60	☑	G_Port	Negotiate	1	☐	
	60	☑	G_Port	Negotiate	1	☐	
	60	☑	G_Port	Negotiate	1	☐	
	60	☑	G_Port	Negotiate	1	☐	
	60	☑	G_Port	Negotiate	1	☐	
	60	☑	G_Port	Negotiate	1	☐	
	60	☑	G_Port	Negotiate	1	☐	
	60	☑	G_Port	Negotiate	1	☐	
	60	☑	G_Port	Negotiate	1	☐	
	60	☑	G_Port	Negotiate	50	☐	
	60	☑	G_Port	Negotiate	50	☐	
	60	☑	G_Port	Negotiate	1	☐	
	60	☑	G_Port	Negotiate	1	☐	
	60	☑	G_Port	Negotiate	1	☐	
	60	☑	G_Port	Negotiate	1	☐	
	60	☑	G_Port	Negotiate	1	☐	
	60	☑	G_Port	Negotiate	1	☐	
	60	☑	G_Port	Negotiate	1	☐	
	60	☑	G_Port	Negotiate	1	☐	
	60	☑	G_Port	Negotiate	1	☐	
	60	☑	G_Port	Negotiate	1	☐	

Figure 10-7 Configure NPIV for an individual switch port

2. Double-click **NPIV Login Limit** for the port and enter the desired login limit. This limits the total number of WWPN logins (both NPIV logins and the default WWPN login).
3. Click **Activate** to complete configuration.

> **Note:** The switch that we used accepts values between 1 and 256. In the examples that follow, three logins are required (two NPIV logins and the default login). When configuring the port, we arbitrarily assign a login limit of 50. If this value is set too small, an `out of resource in fabric` error message is reported on the Linux host.

10.2.2 Configuring NPIV on the System z server

Once NPIV is configured on the switch, NPIV can be enabled on the System z server. The FCP CHPID must be taken offline to enable the NPIV feature.

Note: NPIV should be enabled on the fabric switch *before* configuring NPIV on the System z server. If NPIV is enabled on the System z server, but not on the switch, the FCP CHPID reverts to non-NPIV mode on fabric login. This could be resolved by turning the corresponding switch port off and on (block and unblock the switch port on our Mcdata switch). Alternatively, but more disruptively, a SE CHPID off/on is required to enable NPIV once the switch is enabled for NPIV.

To enable the NPIV, use the SE or HMC to:

- Enable the NPIV feature on the System z server.
- Find the NPIV WWPNs for the FCP CHPID.
- Find the permanent WWPN for the FCP CHPID.

Enable the NPIV feature on the System z server

The NPIV feature can be enabled from the CHPID Operations menu (Figure 10-8 on page 181) in the SE. From the Hardware Management Console (HMC), select Single Object Operations to navigate to the SE:

1. Select your CPC, right-click it, and select the **Channels** option.
2. Scroll to the CHPID Operations task on the right.

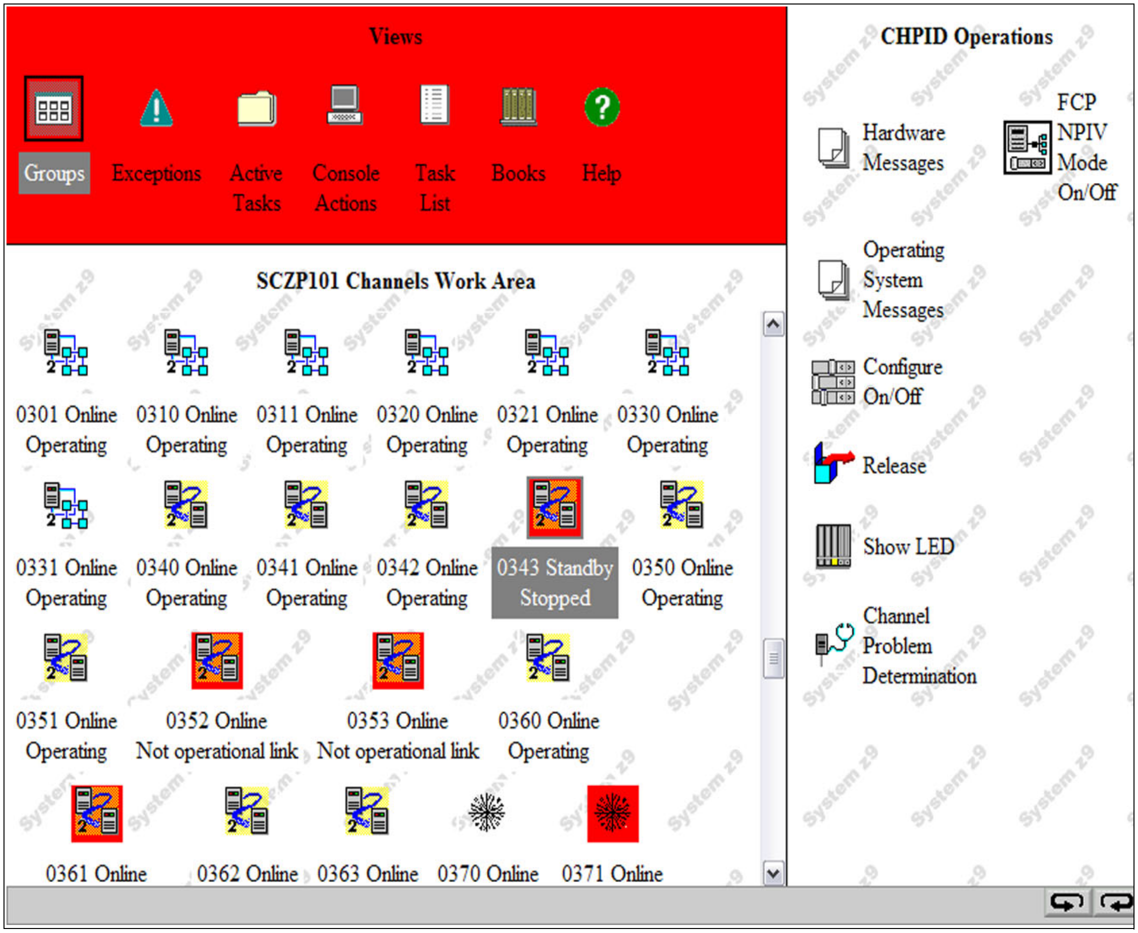
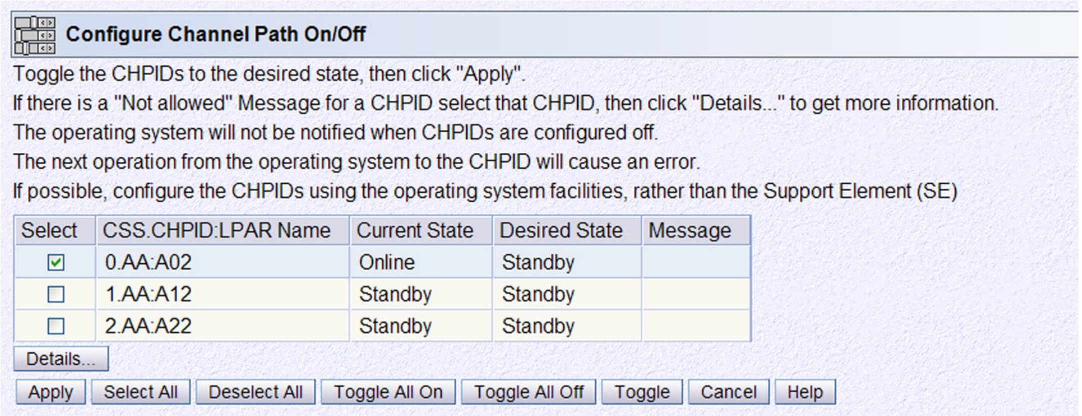


Figure 10-8 The CHPID Operations menu in the SE

The NPIV feature can be selectively enabled for individual LPARs. From the CHPID Operations menu:

1. Set the FCP CHPID to standby as follows:
   a. Double-click **Configure On/Off** to open the menu shown in Figure 10-9.



Configure Channel Path On/Off

Toggle the CHPIDs to the desired state, then click "Apply".

If there is a "Not allowed" Message for a CHPID select that CHPID, then click "Details..." to get more information.

The operating system will not be notified when CHPIDs are configured off.

The next operation from the operating system to the CHPID will cause an error.

If possible, configure the CHPIDs using the operating system facilities, rather than the Support Element (SE)

Select	CSS.CHPID:LPAR Name	Current State	Desired State	Message
☑	0.AA:A02	Online	Standby	
□	1.AA:A12	Standby	Standby	
□	2.AA:A22	Standby	Standby	

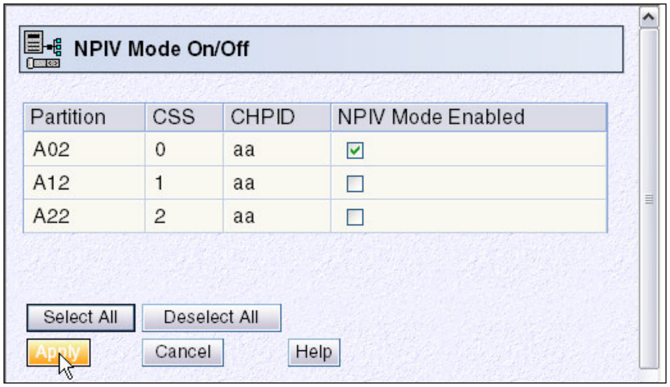
Figure 10-9 Configure Channel Path On/Off menu

   b. Select the appropriate LPARs and click **Toggle** to change the Desired State option to Standby.
   c. Click **Apply** to commit the changes.

2. Enable the NPIV feature as follows:

   a. Select your PCHID number in **Channel Work Area**.

      - From the CHPID Operations menu, double-click **FCP NPIV Mode On/Off** to open the NPIV Mode On/Off menu in Figure 10-10 on page 183.

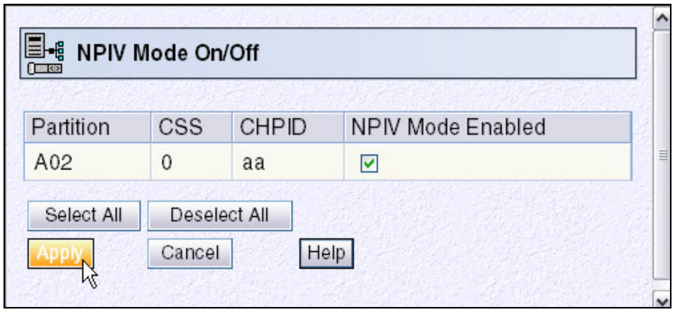


NPIV Mode On/Off

Partition	CSS	CHPID	NPIV Mode Enabled
A02	0	aa	☑
A12	1	aa	□
A22	2	aa	□

Figure 10-10 The NPIV Mode On/Off menu of a PCHID under CPC

- Alternatively, you could select the CHPID number in **Chpids Work Area** under a LPAR image. From the CHPID Operations menu, double-click **FCP NPIV Mode On/Off** to navigate to the NPIV Mode On/Off menu in Figure 10-11.



NPIV Mode On/Off

Partition	CSS	CHPID	NPIV Mode Enabled
A02	0	aa	☑

Figure 10-11 The NPIV Mode On/Off Menu of a CHPID under a LPAR image

   b. Select the NPIV Mode Enable option for each LPAR.

   c. Click **Apply** to commit the changes.

Note: To enable NPIV, the CHPID must be in standby state for the LPAR. If not, the NPIV Mode On/Off option is disabled to prevent any mode changes.

3. Set the FCP CHPID online as follows:

   a. From the CHPID Operations menu, double-click **Configure Channel Path On/Off**.

   b. Select the appropriate LPARs and click **Toggle** to change the Desired State option to **Online**.

   c. Click **Apply** to commit the changes.

Finding the NPIV WWPNs for the FCP CHPID

Once enabled, the NPIV WWPNs that are assigned to an FCP CHPID can be found in the SE. These WWPNs are needed to configure LUN masking on the storage server and zoning in the fabric (switches).

The NPIV WWPNs are accessible from the CPC Configuration menu in the SE (shown in Figure 10-12).

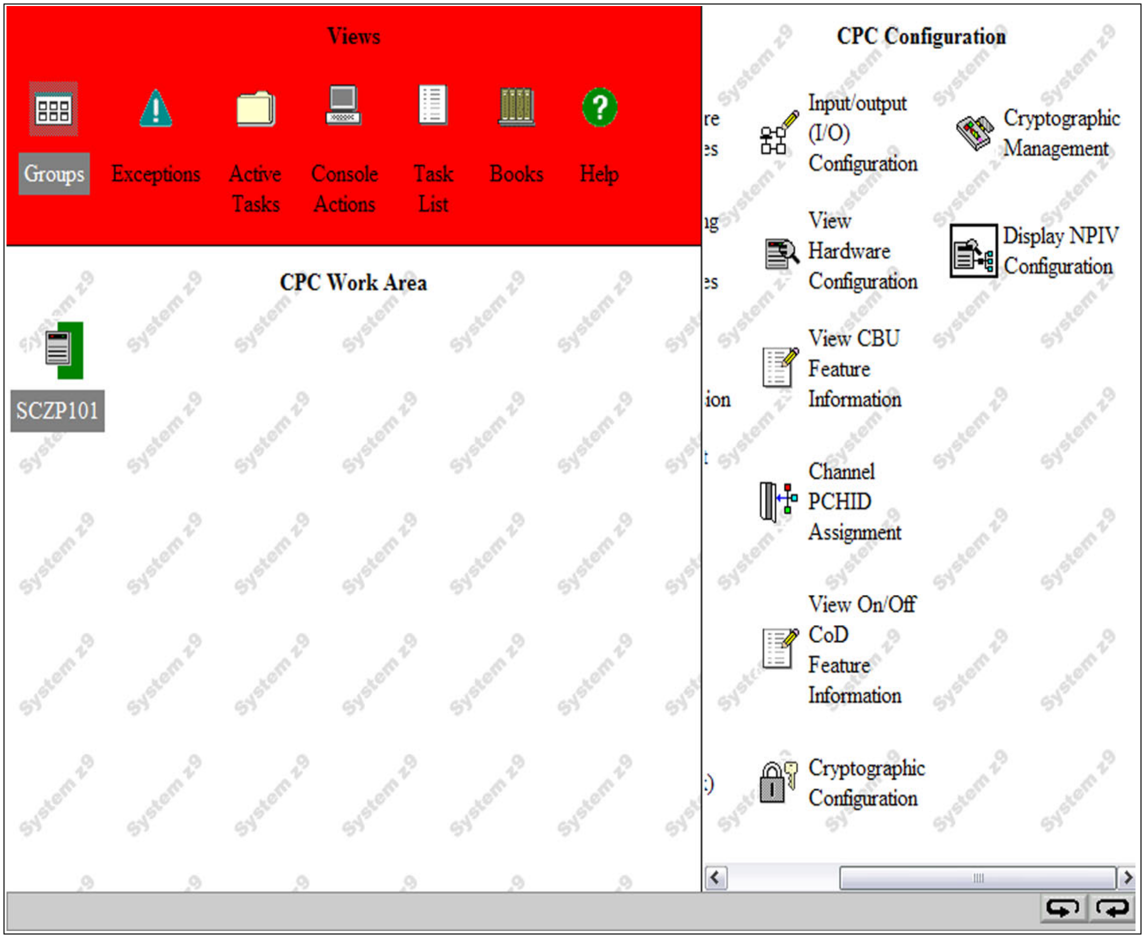
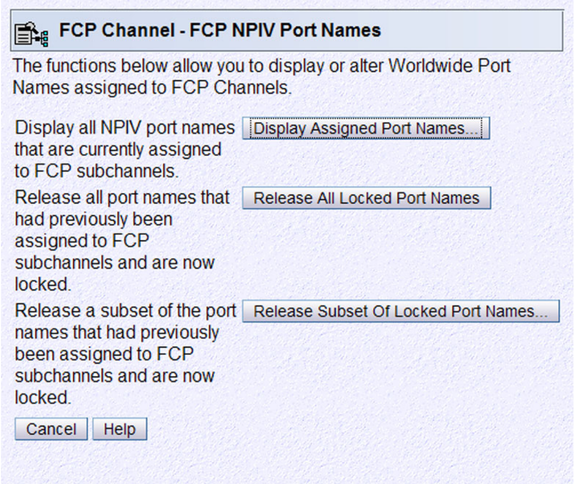

Views
Groups
Exceptions
Active Tasks
Console Actions
Task List
Books
Help
CPC Work Area
SCZP101
CPC Configuration
Input/output (I/O) Configuration
Cryptographic Management
View Hardware Configuration
Display NPIV Configuration
View CBU Feature Information
Channel PCHID Assignment
View On/Off CoD Feature Information
Cryptographic Configuration

Figure 10-12 The CPC configuration menu in the SE

To find the WWPNs:

1. Click **Display NPIV Configuration** to navigate to the FCP Channel - FCP NPIV Port Names menu shown in Figure 10-13.



FCP Channel - FCP NPIV Port Names

The functions below allow you to display or alter Worldwide Port Names assigned to FCP Channels.

Display all NPIV port names that are currently assigned to FCP subchannels.

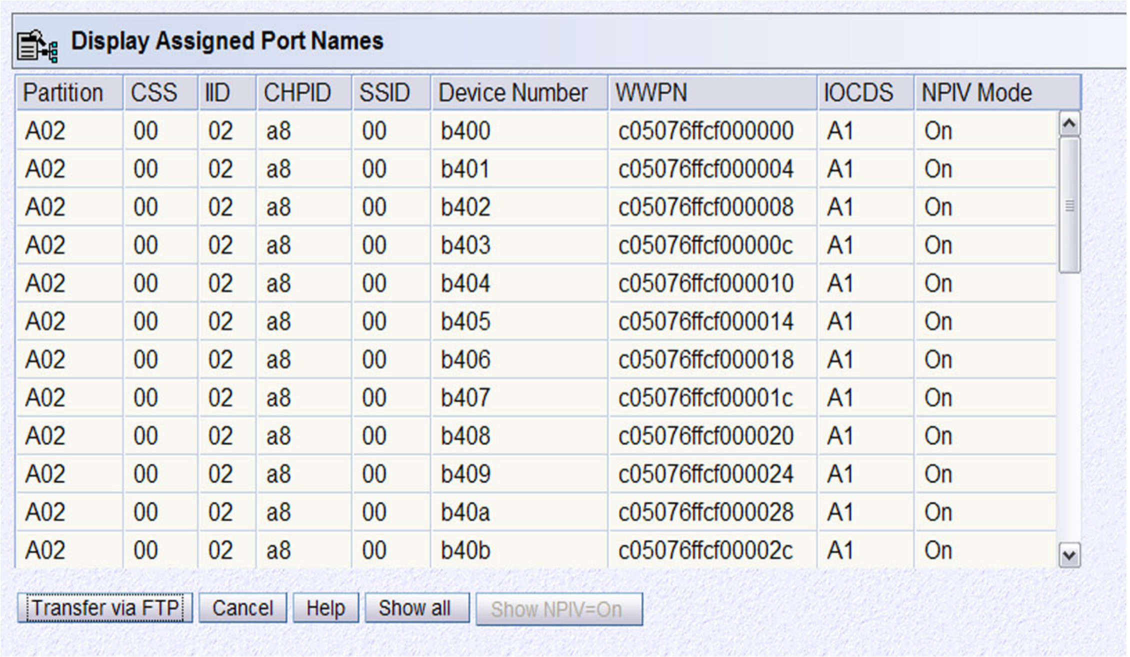
Release all port names that had previously been assigned to FCP subchannels and are now locked.

Release a subset of the port names that had previously been assigned to FCP subchannels and are now locked.

Display Assigned Port Names...
Release All Locked Port Names
Release Subset Of Locked Port Names...
Cancel
Help

Figure 10-13 The FCP Channel - FCP NPIV Port Names menu

2. Click **Display Assigned Port Name** to open the Display Assigned Port Names menu shown in Figure 10-14 on page 187. It is extremely helpful to restrict the number of WWPNs in the display by selecting the (Show NPIV=On) option.



Partition	CSS	IID	CHPID	SSID	Device Number	WWPN	IOCDS	NPIV Mode
A02	00	02	a8	00	b400	c05076ffcf000000	A1	On
A02	00	02	a8	00	b401	c05076ffcf000004	A1	On
A02	00	02	a8	00	b402	c05076ffcf000008	A1	On
A02	00	02	a8	00	b403	c05076ffcf00000c	A1	On
A02	00	02	a8	00	b404	c05076ffcf000010	A1	On
A02	00	02	a8	00	b405	c05076ffcf000014	A1	On
A02	00	02	a8	00	b406	c05076ffcf000018	A1	On
A02	00	02	a8	00	b407	c05076ffcf00001c	A1	On
A02	00	02	a8	00	b408	c05076ffcf000020	A1	On
A02	00	02	a8	00	b409	c05076ffcf000024	A1	On
A02	00	02	a8	00	b40a	c05076ffcf000028	A1	On
A02	00	02	a8	00	b40b	c05076ffcf00002c	A1	On

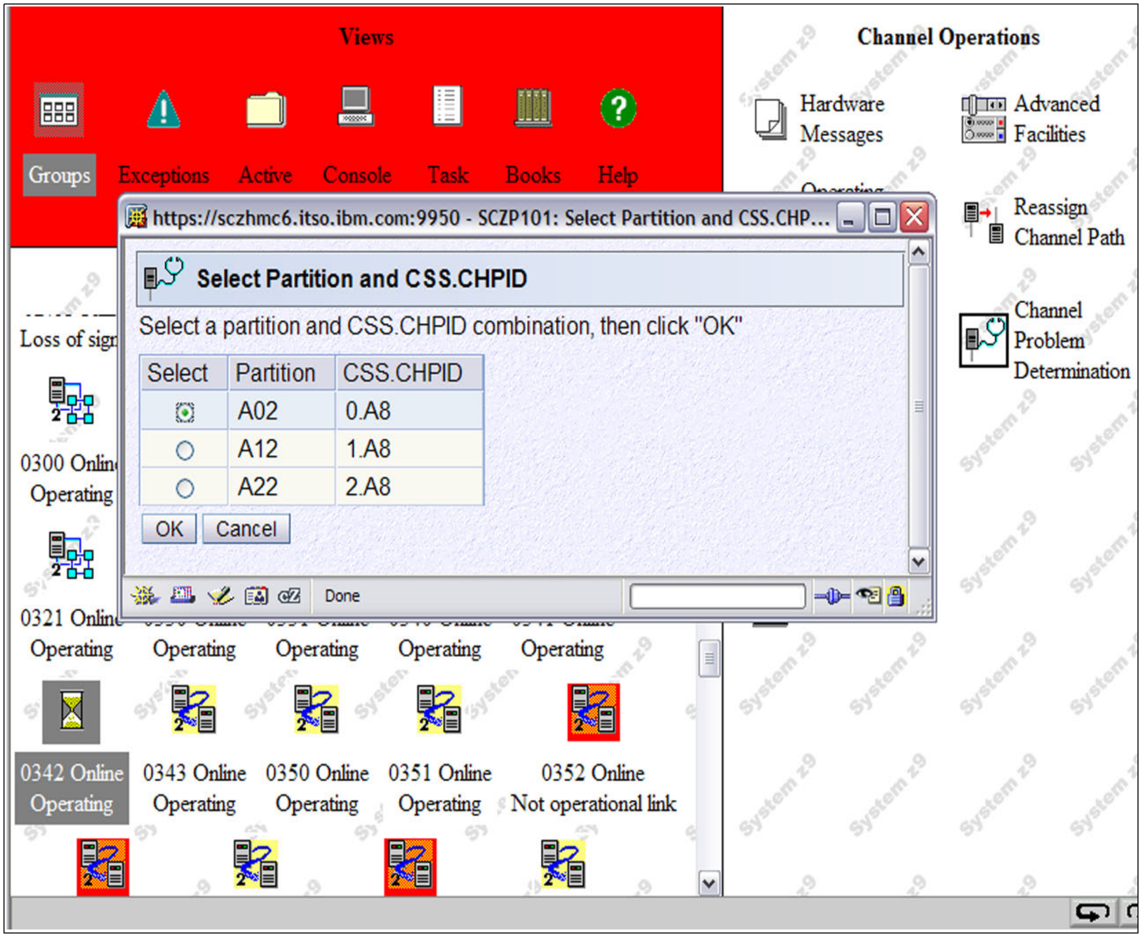
Figure 10-14 The Display Assigned Port Names menu

3. Each FCP subchannel (device) number in an LPAR is assigned a unique NPIV WWPN. Click **Transfer via** FTP to copy a text version of this menu to an FTP server.

Finding the permanent WWPN for the FCP CHPID

The permanent WWPN is also needed for LUN masking and fabric zoning. To find it:

1. From the Channel Operations menu in the SE, click **Channel Problem Determination**, as shown in Figure 10-15 on page 188.



https://sczhmc6.itso.ibm.com:9950 - SCZP101: Select Partition and CSS.CHP...

Select Partition and CSS.CHPID

Select a partition and CSS.CHPID combination, then click "OK"

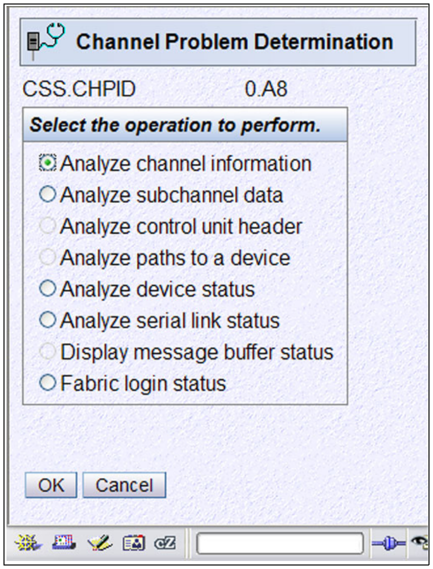
Select	Partition	CSS.CHPID
●	A02	0.A8
○	A12	1.A8
○	A22	2.A8

OK Cancel

Done

Figure 10-15 Select the LPAR in the Channel Problem Determination menu

2. Select the desired LPAR and click **OK** to open the Channel Problem Determination menu shown in Figure 10-16.



Channel Problem Determination

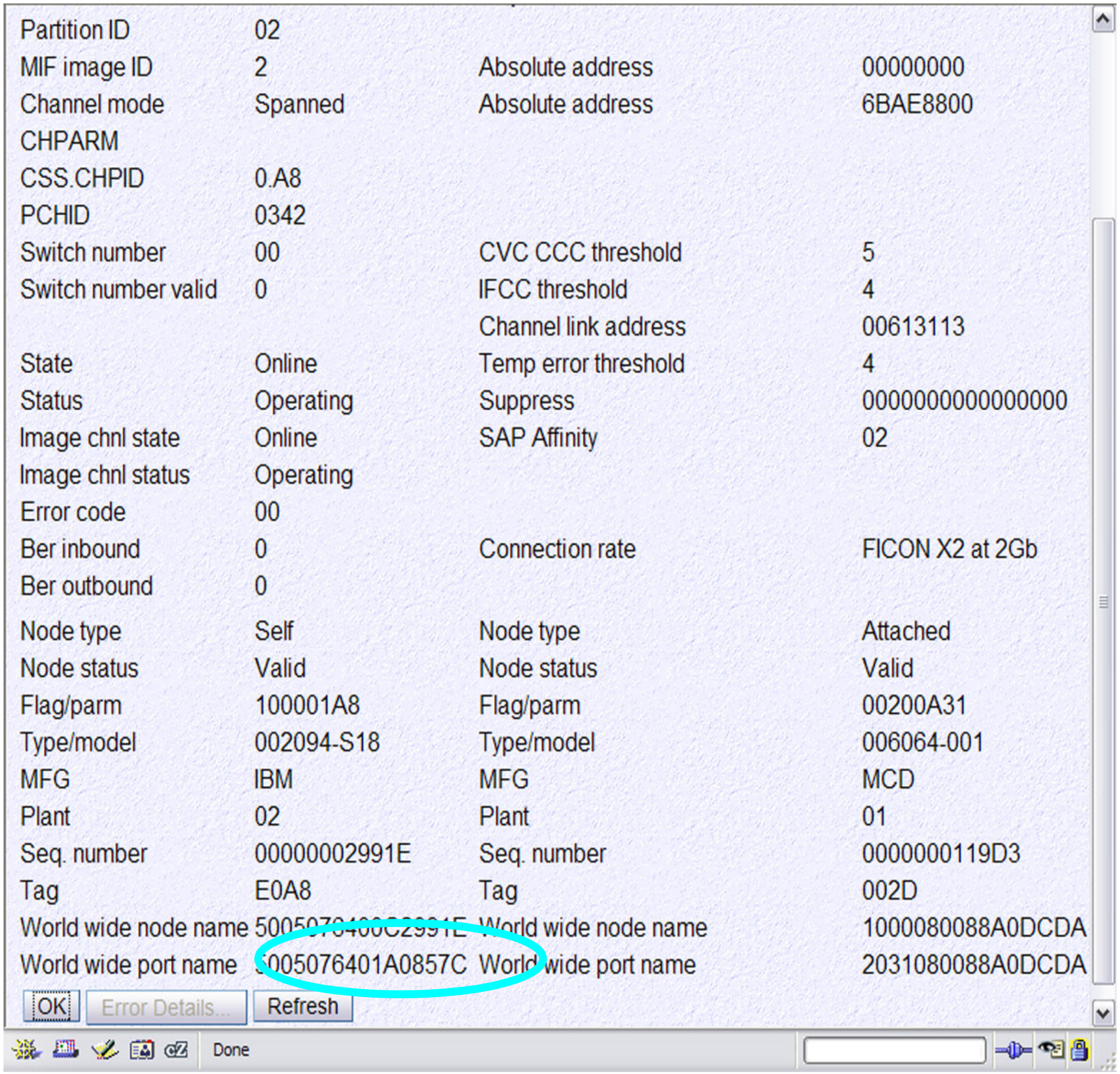
CSS.CHPID 0.A8

Select the operation to perform.

● Analyze channel information
○ Analyze subchannel data
○ Analyze control unit header
○ Analyze paths to a device
○ Analyze device status
○ Analyze serial link status
○ Display message buffer status
○ Fabric login status

Figure 10-16 Channel Problem Determination menu

3. Select the Analyze channel information option and click **OK** to open the menu shown in Figure 10-17.



Partition ID	02		
MIF image ID	2	Absolute address	00000000
Channel mode	Spanned	Absolute address	6BAE8800
CHPARM			
CSS.CHPID	0.A8		
PCHID	0342		
Switch number	00	CVC CCC threshold	5
Switch number valid	0	IFCC threshold	4
		Channel link address	00613113
State	Online	Temp error threshold	4
Status	Operating	Suppress	0000000000000000
Image chnl state	Online	SAP Affinity	02
Image chnl status	Operating		
Error code	00		
Ber inbound	0	Connection rate	FICON X2 at 2Gb
Ber outbound	0		
Node type	Self	Node type	Attached
Node status	Valid	Node status	Valid
Flag/parm	100001A8	Flag/parm	00200A31
Type/model	002094-S18	Type/model	006064-001
MFG	IBM	MFG	MCD
Plant	02	Plant	01
Seq. number	00000002991E	Seq. number	0000000119D3
Tag	E0A8	Tag	002D
World wide node name	5005076400C2991E	World wide node name	1000080088A0DCDA
World wide port name	5005076401A0857C	World wide port name	2031080088A0DCDA

Figure 10-17 The Analyze Channel Information menu

4. Record the default WWPN for the CHPID (highlighted in Figure 10-17).

10.2.3 Configuring fabric security

The NPIV and permanent (default) WWPNs must be defined in the fabric (switch) zoning and LUN masking on the storage server.

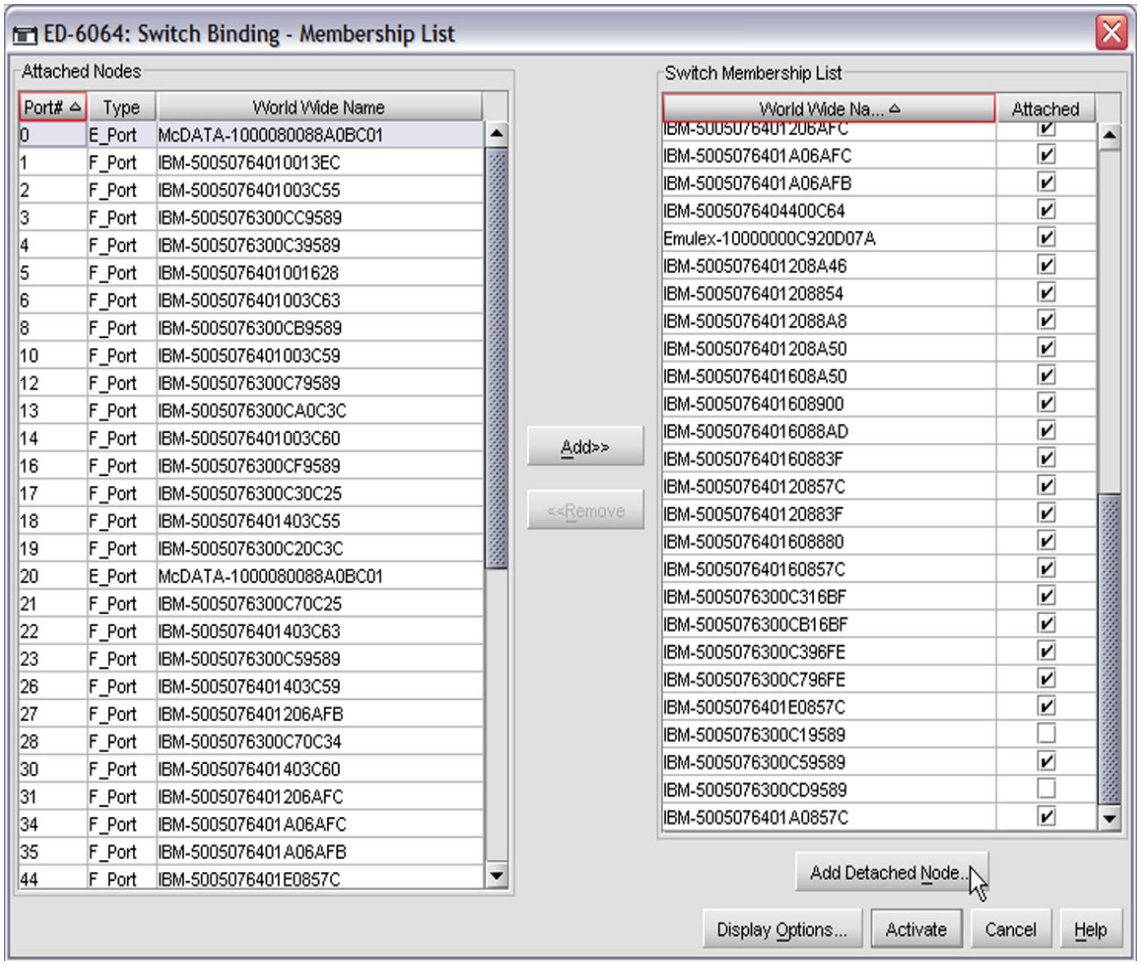
Note: Zoning and LUN masking configuration are specific to the switch and storage server that are used in the fabric. Consult your vendor documentation for details.

In the examples that follow, the authors used a McData SAN switch and an ESS 800.

Configuring fabric zoning on the switch

To do this:

1. From the EFCM switch management application, select **Configure** → **Switch Binding** → **Edit Membership List** to open the Switch Binding - Membership List menu in Figure 10-18.



ED-6064: Switch Binding - Membership List

Attached Nodes

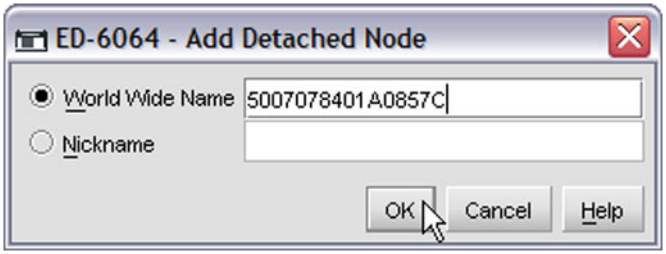
Port#	Type	World Wide Name
0	E_Port	McDATA-1000080088A0BC01
1	F_Port	IBM-50050764010013EC
2	F_Port	IBM-5005076401003C55
3	F_Port	IBM-5005076300CC9589
4	F_Port	IBM-5005076300C39589
5	F_Port	IBM-5005076401001628
6	F_Port	IBM-5005076401003C63
8	F_Port	IBM-5005076300CB9589
10	F_Port	IBM-5005076401003C59
12	F_Port	IBM-5005076300C79589
13	F_Port	IBM-5005076300CA0C3C
14	F_Port	IBM-5005076401003C60
16	F_Port	IBM-5005076300CF9589
17	F_Port	IBM-5005076300C30C25
18	F_Port	IBM-5005076401403C55
19	F_Port	IBM-5005076300C20C3C
20	E_Port	McDATA-1000080088A0BC01
21	F_Port	IBM-5005076300C70C25
22	F_Port	IBM-5005076401403C63
23	F_Port	IBM-5005076300C59589
26	F_Port	IBM-5005076401403C59
27	F_Port	IBM-5005076401206AFB
28	F_Port	IBM-5005076300C70C34
30	F_Port	IBM-5005076401403C60
31	F_Port	IBM-5005076401206AFC
34	F_Port	IBM-5005076401A06AFC
35	F_Port	IBM-5005076401A06AFB
44	F_Port	IBM-5005076401E0857C

Switch Membership List

World Wide Na...	Attached
IBM-5005076401206AFC	☑
IBM-5005076401A06AFC	☑
IBM-5005076401A06AFB	☑
IBM-5005076404400C64	☑
Emulex-10000000C920D07A	☑
IBM-5005076401208A46	☑
IBM-5005076401208854	☑
IBM-50050764012088A8	☑
IBM-5005076401208A50	☑
IBM-5005076401608A50	☑
IBM-5005076401608900	☑
IBM-50050764016088AD	☑
IBM-500507640160883F	☑
IBM-500507640120857C	☑
IBM-500507640120883F	☑
IBM-5005076401608880	☑
IBM-500507640160857C	☑
IBM-5005076300C316BF	☑
IBM-5005076300CB16BF	☑
IBM-5005076300C396FE	☑
IBM-5005076300C796FE	☑
IBM-5005076401E0857C	☑
IBM-5005076300C19589	□
IBM-5005076300C59589	☑
IBM-5005076300CD9589	□
IBM-5005076401A0857C	☑

Figure 10-18 Switch Binding - Membership List menu

2. Click **Add Detached Node** to open the menu shown in Figure 10-19 on page 193.



● World Wide Name 5007078401A0857C
○ Nickname

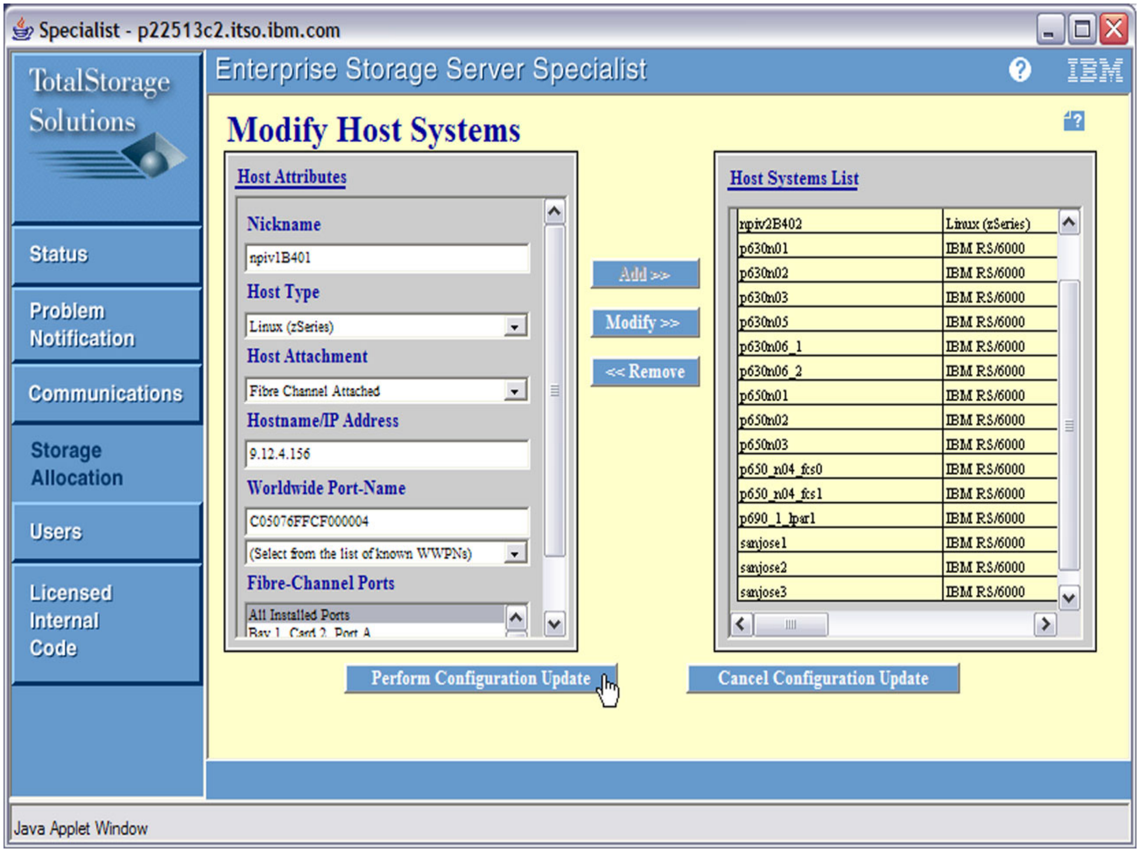
Figure 10-19 Add Detached Node menu

3. Add the FCP subchannel WWPN to the named zone and click **OK**. In this example, the authors added the permanent WWPN to the default zone for the switch.
4. Click **Activate** on the Switch Binding - Membership List menu to commit the changes.

Configuring LUN masking on the storage server

LUN masking is configured on the storage server. Access to a specific LUN is granted to a WWPN or group of WWPNs with LUN masking. To configure LUN masking on the ESS, log on to the ESS Specialist application and start by defining the Linux server to the storage server as follows:

1. Click **Storage Allocation** → **Open System Storage™** → **Modify Host Systems** to open the Modify Host System menu shown in Figure 10-20 on page 194.



Modify Host Systems

Host Attributes

Nickname: npiv1B401

Host Type: Linux (zSeries)

Host Attachment: Fibre Channel Attached

Hostname/IP Address: 9.12.4.156

Worldwide Port-Name: C05076FFCF000004

(Select from the list of known WWPNs)

Fibre-Channel Ports: All Installed Ports; Bay 1, Card 2, Port A

Host Systems List

npiv2B402	Linux (zSeries)
p630n01	IBM RS/6000
p630n02	IBM RS/6000
p630n03	IBM RS/6000
p630n05	IBM RS/6000
p630n06_1	IBM RS/6000
p630n06_2	IBM RS/6000
p650n01	IBM RS/6000
p650n02	IBM RS/6000
p650n03	IBM RS/6000
p650_n04_fcs0	IBM RS/6000
p650_n04_fcs1	IBM RS/6000
p690_1_lpar1	IBM RS/6000
sanjose1	IBM RS/6000
sanjose2	IBM RS/6000
sanjose3	IBM RS/6000

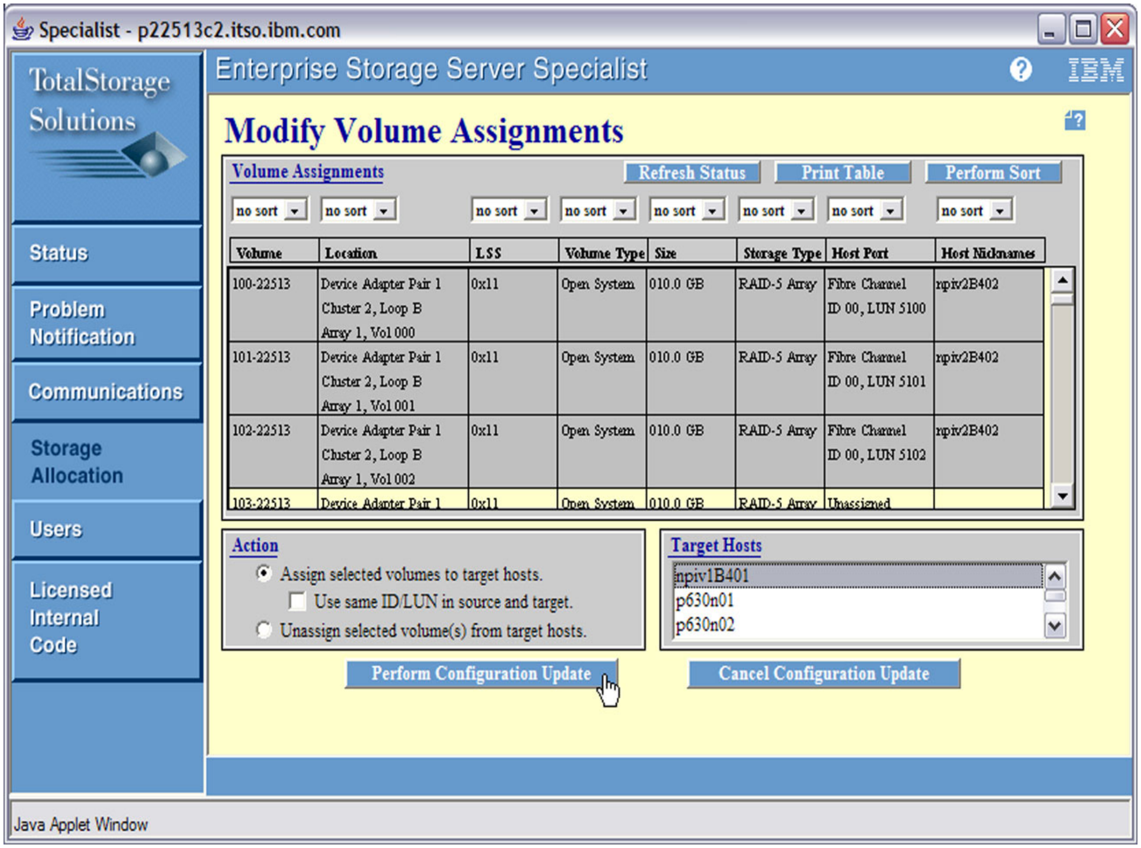
Figure 10-20 The Modify Host System menu

2. Complete the form in the menu and click **Add** to define a host type of Linux.
3. Click **Perform Configuration Update** to commit the change.
4.

> **Note:** The authors defined the Linux server using the nickname npiv1B401. For the WWPN, we supplied the NPIV WWPN for the B401 device (the device that is used by the NPIV1 Linux guest).

Next, you assign some existing storage volumes to the Linux server. From the Modify Volume Assignments menu for Linux server npiv2B402 (shown in Figure 10-21):

1. Select three LUNs (5100-5102) from the Volume Assignments list.
2. Click **Assign selected volumes to target hosts**.
3. Select **npiv1B401** from the Target Hosts list.



Modify Volume Assignments

Volume Assignments

Volume	Location	LSS	Volume Type	Size	Storage Type	Host Port	Host Nicknames
100-22513	Device Adapter Pair 1 Cluster 2, Loop B Array 1, Vol 000	0x11	Open System	010.0 GB	RAID-5 Array	Fibre Channel ID 00, LUN 5100	npiv2B402
101-22513	Device Adapter Pair 1 Cluster 2, Loop B Array 1, Vol 001	0x11	Open System	010.0 GB	RAID-5 Array	Fibre Channel ID 00, LUN 5101	npiv2B402
102-22513	Device Adapter Pair 1 Cluster 2, Loop B Array 1, Vol 002	0x11	Open System	010.0 GB	RAID-5 Array	Fibre Channel ID 00, LUN 5102	npiv2B402
103-22513	Device Adapter Pair 1	0x11	Open System	010.0 GB	RAID-5 Array	Unassigned	

Action

● Assign selected volumes to target hosts.

□ Use same ID/LUN in source and target.

○ Unassign selected volume(s) from target hosts.

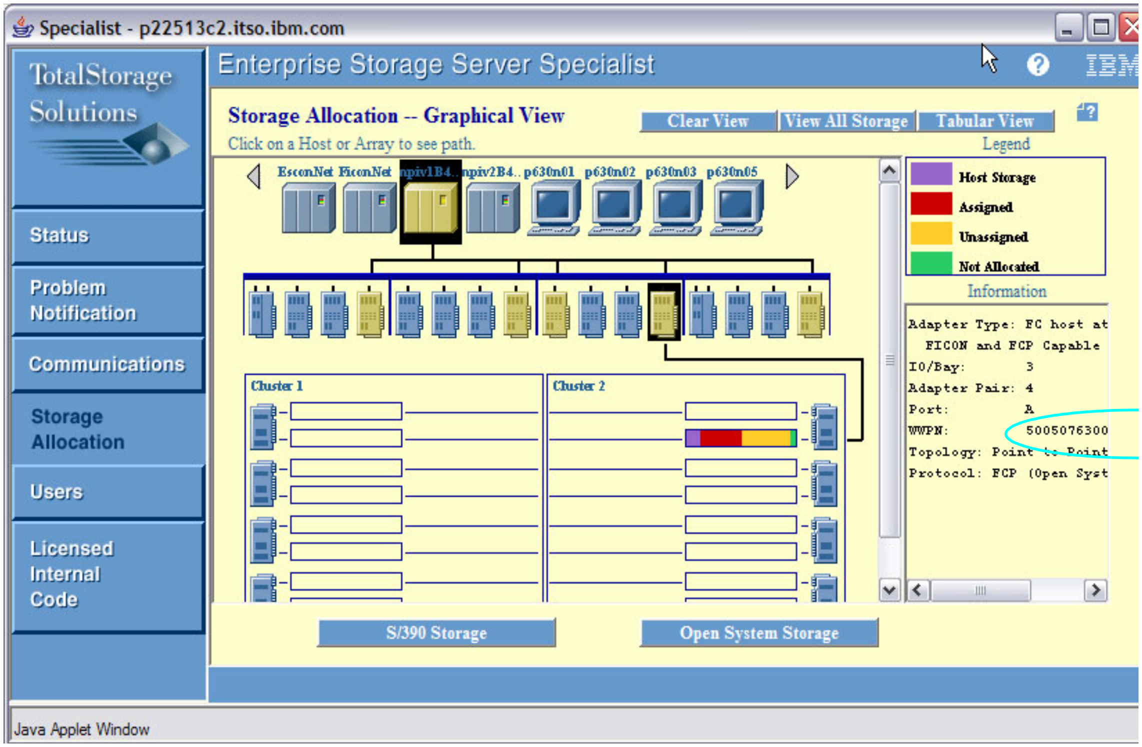
Target Hosts

npiv1B401
p630n01
p630n02

Figure 10-21 The Modify Volume Assignments menu

4. Click **Perform Configuration Update** to commit the changes.

Finally, you find the WWPN of the HBA on the ESS that is used to access the LUNs in the Storage Allocation - Graphic View menu shown in Figure 10-22.



- 🟪 Host Storage
- 🟥 Assigned
- 🟨 Unassigned
- 🟩 Not Allocated

Information

```
Adapter Type: FC host at
  FICON and FCP Capable
IO/Bay:         3
Adapter Pair: 4
Port:           A
WWPN:           5005076300
Topology: Point to Point
Protocol: FCP (Open Syst
```

Figure 10-22 The Storage Allocation - Graphic View menu

Of the 16 ports in the storage server, five are configured for FCP. Any one of these five can be used to access the LUNs. The HBA in bay 3, port 4 connects to the switch. Click this HBA to display the Information panel on the right (where the WWPN can be found).

10.2.4 Configuring the Linux server

At this point, the Linux server can be configured to use the NPIV WWPN. With NPIV, each device on the FCP CHPID is assigned a unique WWPN. The WWPN used by Linux (LPAR or VM guest) or z/VM (for a EDEVICE) is determined by the FCP subchannel (device) of the server.

> **Note:** To avoid data corruption, never mount a disk partition in read/write mode to more than one server without proper software support.

You can use the following steps to bring up the disk volume to your Linux images:

1. Attach FCP devices to Linux images from z/VM.
2. Use YaST to set up the Disk Volumes.
3. Verify the NPIV WWPN login.

Attaching the FCP devices to the Linux guest

When running under z/VM, the FCP device must be attached to the virtual machine of the Linux guest. To attach the device dynamically, use the CP ATTACH command. From the MAINT user, the authors attached two FCP devices to two Linux guests (B401 to Linux guest NPIV1 and B402 to Linux guest NPIV2). The commands are illustrated in Figure 10-23.

```
Q CHPID A8
Path A8 online to devices B400 B401 B402 B403 B404 B405 B406 B407
Path A8 online to devices B408 B409 B40A B40B B40C B40D B40E B40F
Path A8 online to devices B410 B411 B412 B413 B414 B415 B416 B417
Path A8 online to devices B418 B419 B41A B41B B41C B41D B41E B41F
Path A8 online to devices B4FC B4FD
Ready; T=0.01/0.01 14:16:27
ATTACH B401 TO NPIV1
FCP B401 ATTACHED TO NPIV1 B401
Ready; T=0.01/0.06 14:16:46
ATTACH B402 TO NPIV2
FCP B402 ATTACHED TO NPIV2 B402
Ready; T=0.01/0.06 14:16:53
Q FCP
FCP  B401 ATTACHED TO NPIV1    B401 CHPID A8
FCP  B402 ATTACHED TO NPIV2    B402 CHPID A8
Ready; T=0.01/0.01 14:16:57
```

Figure 10-23 Attaching FCP devices to Linux guests

Both Linux guests are running when the FCP devices are attached. Figure 10-24 on page 198 shows the messages that appear in the console for guest NPIV1.

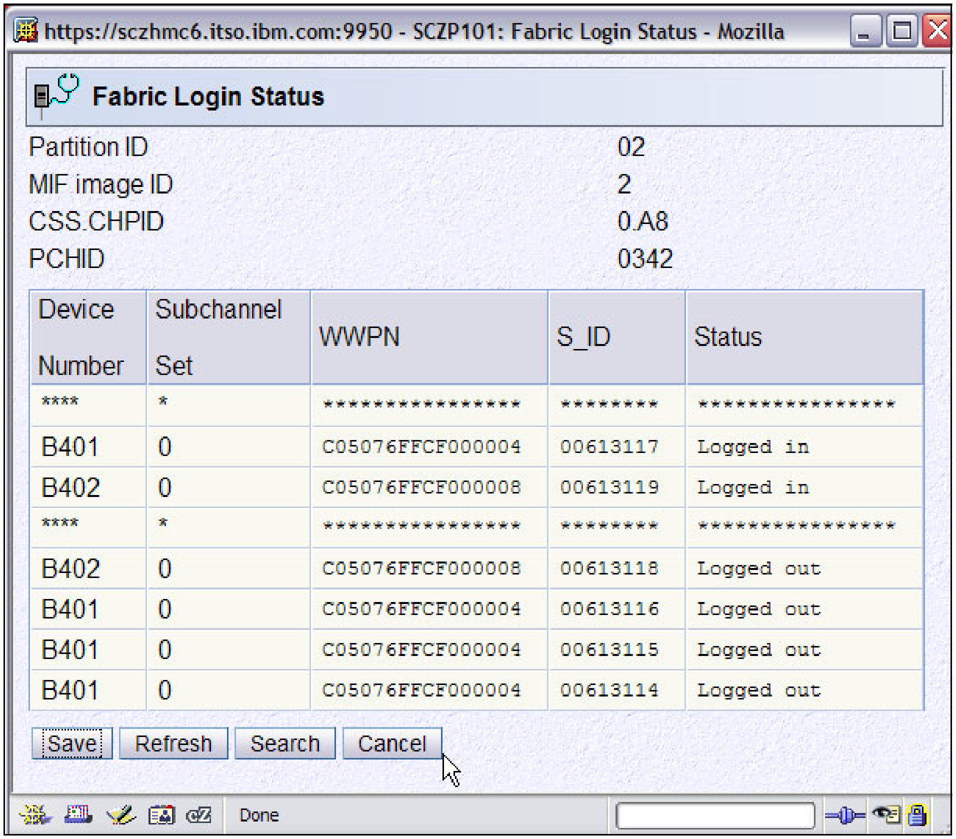
```
FCP B401 ATTACHED TO NPIV1 B401
zfcp: adapter 0.0.b401: operational again
Feb 22 14:22:51 npiv1 kernel: crw_info : CRW reports slct=0, oflw=0, chn=0, rsc=
3, anc=1, erc=4, rsid=12
Feb 22 14:22:51 npiv1 kernel: zfcp: adapter 0.0.b401: operational again
zfcp: The adapter 0.0.b401 reported the following characteristics:
WWNN 0x5005076400c2991e, WWPN 0xc05076ffcf000004, S_ID 0x00613128,
adapter version 0x3, LIC version 0x600, FC link speed 2 Gb/s
zfcp: Switched fabric fibrechannel network detected at adapter 0.0.b401.
Feb 22 14:22:52 npiv1 kernel: zfcp: The adapter 0.0.b401 reported the following
characteristics:
Feb 22 14:22:52 npiv1 kernel: WWNN 0x5005076400c2991e, WWPN 0xc05076ffcf000004,
S_ID 0x00613128,
Feb 22 14:22:52 npiv1 kernel: adapter version 0x3, LIC version 0x600, FC link sp
eed 2 Gb/s
Feb 22 14:22:52 npiv1 kernel: zfcp: Switched fabric fibrechannel network detecte
d at adapter 0.0.b401.
```

Figure 10-24 Console messages from a Linux guest (NPIV1)

As expected, the WWPN for the B401 device uses the NPIV WWPN that is assigned by the SE (shown in Figure 10-14 on page 187).

Note: If the NPIV WWPN is unable to successfully log in to the fabric, its WWPN is assigned the value 0x0000000000000000. In this case, check zoning at the switch (see "Configuring fabric zoning on the switch" on page 192). If an `out of resource in fabric` message is shown, the reason might be that the NPIV login limit was set too small on the switch port (see "Configuring NPIV on an individual switch port" on page 179).

Login status can be checked from the Channel Problem Determination menu in the SE (see Figure 10-16 on page 189). Select **Fabric Login Status** to display the menu shown in Figure 10-25 on page 199.



https://sczhmc6.itso.ibm.com:9950 - SCZP101: Fabric Login Status - Mozilla
Fabric Login Status
Save
Refresh
Search
Cancel
Done

Partition ID	02
MIF image ID	2
CSS.CHPID	0.A8
PCHID	0342

Device Number	Subchannel Set	WWPN	S_ID	Status
****	*	****************	********	****************
B401	0	C05076FFCF000004	00613117	Logged in
B402	0	C05076FFCF000008	00613119	Logged in
****	*	****************	********	****************
B402	0	C05076FFCF000008	00613118	Logged out
B401	0	C05076FFCF000004	00613116	Logged out
B401	0	C05076FFCF000004	00613115	Logged out
B401	0	C05076FFCF000004	00613114	Logged out

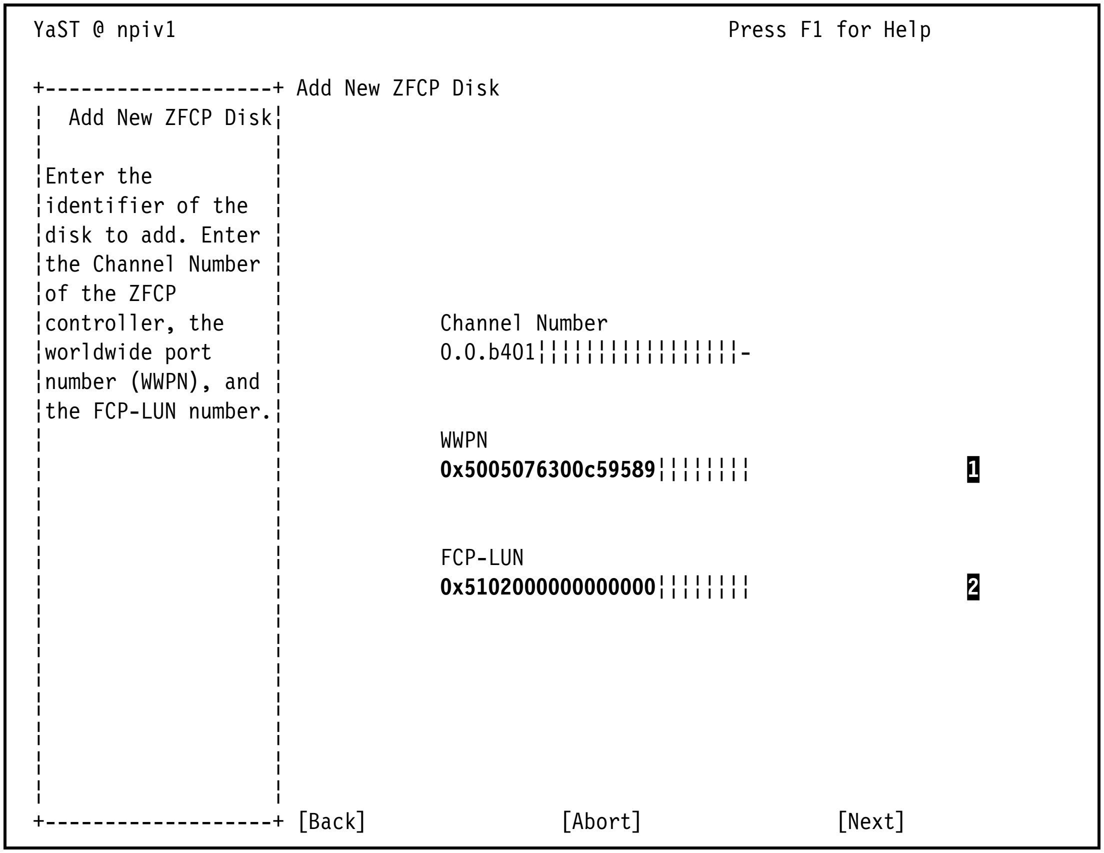
Figure 10-25 Fabric Login Status menu

If the WWPN is in Logged in status but the LUN cannot be accessed, be sure the NPIV WWPN is authorized to access the LUN at the storage server (see “Configuring LUN masking on the storage server” on page 194).

Using YaST to set up the disk volumes

Once the device is attached, the FCP disk can be added to a Linux server as follows:

1. From YaST, select **Hardware** → **zFCP** and click **Add** to open the menu shown in Figure 10-26.



```
YaST @ npiv1                                               Press F1 for Help

+--------------------+ Add New ZFCP Disk
¦  Add New ZFCP Disk¦
¦                    ¦
¦Enter the           ¦
¦identifier of the   ¦
¦disk to add. Enter  ¦
¦the Channel Number  ¦
¦of the ZFCP         ¦
¦controller, the     ¦            Channel Number
¦worldwide port      ¦            0.0.b401¦¦¦¦¦¦¦¦¦¦¦¦¦¦¦¦¦-
¦number (WWPN), and  ¦
¦the FCP-LUN number.¦
¦                    ¦            WWPN
¦                    ¦            0x5005076300c59589¦¦¦¦¦¦¦¦                  1
¦                    ¦
¦                    ¦
¦                    ¦            FCP-LUN
¦                    ¦            0x5102000000000000¦¦¦¦¦¦¦¦                  2
¦                    ¦
¦                    ¦
¦                    ¦
¦                    ¦
¦                    ¦
¦                    ¦
¦                    ¦
+--------------------+ [Back]                [Abort]                [Next]
```

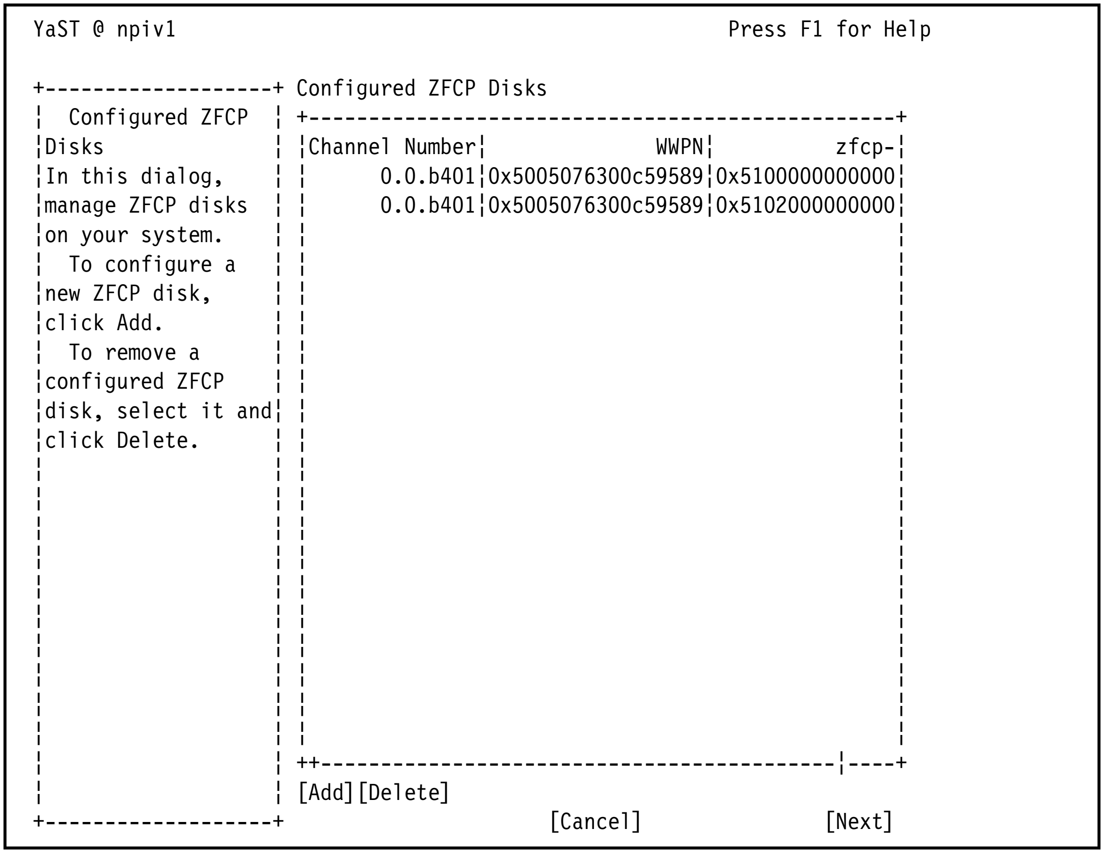
Figure 10-26 YaST menu to configure FCP disk

2. In this menu, the FCP device (0.0.b401) is automatically detected by the Linux kernel. We can replace it if it is not our intended device. We provided:

   – The WWPN of the storage (0x5005076300c59589)
   – The LUN of the SCSI disk (0x5102000000000000)

3. Click **Next** when done to continue.

In Figure 10-27, YaST displays two LUNs that have been added to the Linux guest (0x5100000000000000 and 0x5102000000000000).



```
YaST @ npiv1                                              Press F1 for Help

+-------------------+ Configured ZFCP Disks
¦  Configured ZFCP  ¦ +-------------------------------------------------+
¦Disks              ¦ ¦Channel Number¦              WWPN¦          zfcp-¦
¦In this dialog,    ¦ ¦      0.0.b401¦0x5005076300c59589¦0x5100000000000¦
¦manage ZFCP disks  ¦ ¦      0.0.b401¦0x5005076300c59589¦0x5102000000000¦
¦on your system.    ¦ ¦                                                 ¦
¦  To configure a   ¦ ¦                                                 ¦
¦new ZFCP disk,     ¦ ¦                                                 ¦
¦click Add.         ¦ ¦                                                 ¦
¦  To remove a      ¦ ¦                                                 ¦
¦configured ZFCP    ¦ ¦                                                 ¦
¦disk, select it and¦ ¦                                                 ¦
¦click Delete.      ¦ ¦                                                 ¦
¦                   ¦ ¦                                                 ¦
¦                   ¦ ¦                                                 ¦
¦                   ¦ ¦                                                 ¦
¦                   ¦ ¦                                                 ¦
¦                   ¦ ¦                                                 ¦
¦                   ¦ ¦                                                 ¦
¦                   ¦ ¦                                                 ¦
¦                   ¦ ¦                                                 ¦
¦                   ¦ ¦                                                 ¦
¦                   ¦ ¦                                                 ¦
¦                   ¦ ++-------------------------------------------¦----+
¦                   ¦ [Add][Delete]
+-------------------+                      [Cancel]               [Next]
```

Figure 10-27 YaST menu for configured FCP devices

We used the **dmesg** command to see that the two LUNs were added as the /dev/sda and /dev/sdb partitions, as shown in Figure 10-28 on page 202.

```
zfcp: The adapter 0.0.b401 reported the following characteristics:
WWNN 0x5005076400c2991e, WWPN 0xc05076ffcf000004, S_ID 0x00613127,
adapter version 0x3, LIC version 0x600, FC link speed 2 Gb/s
zfcp: Switched fabric fibrechannel network detected at adapter 0.0.b401.
  Vendor: IBM      Model: 2105800          Rev: 3.56
  Type:   Direct-Access                    ANSI SCSI revision: 03
SCSI device sda: 19531264 512-byte hdwr sectors (10000 MB)
SCSI device sda: drive cache: write back
 sda: sda1 sda2
Attached scsi disk sda at scsi0, channel 0, id 1, lun 0
  Vendor: IBM      Model: 2105800          Rev: 3.56
  Type:   Direct-Access                    ANSI SCSI revision: 03
SCSI device sdb: 19531264 512-byte hdwr sectors (10000 MB)
SCSI device sdb: drive cache: write back
 sdb: sdb1
Attached scsi disk sdb at scsi0, channel 0, id 1, lun 1
Attached scsi generic sg0 at scsi0, channel 0, id 1, lun 0,  type 0
Attached scsi generic sg1 at scsi0, channel 0, id 1, lun 1,  type 0
```

Figure 10-28 Linux kernel messages when adding SCSI devices

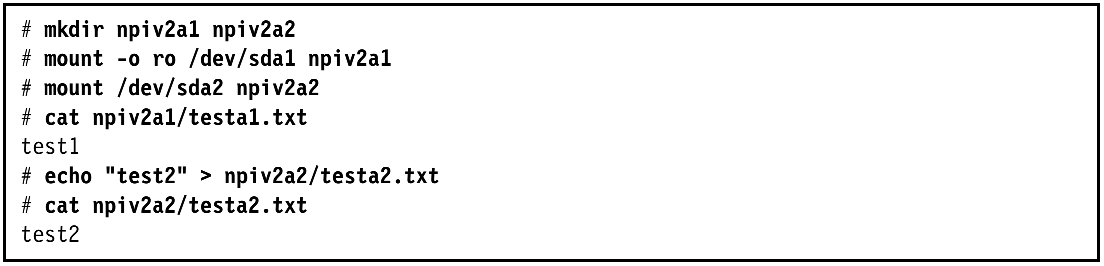
Each has a capacity of 10,000 MB. The kernel recognizes two existing partitions (/dev/sda1 and /dev/sda2) on the /dev/sda device. These are formatted as ext3 partitions.

On the NPIV1 Linux guest, we mounted the /dev/sda1 partition in read/write mode and the /dev/sda2 partition in read-only mode. Next, we created a test text file on the /dev/sda1 partition in Figure 10-29.

```
# mkdir npiv1a1 npiv1a2
# mount /dev/sda1 npiv1a1
# mount -o ro /dev/sda2 npiv1a2
# echo "test1" > npiv1a1/testa1.txt
# cat npiv1a1/testa1.txt
test1
```

Figure 10-29 Mounting the SCSI partitions on the NPIV1 Linux guest

On the NPIV2 Linux guest, we mounted the /dev/sda1 partition in read-only mode, and the /dev/sda2 partition in read/write mode. This allows only one host to have write access to a shared partition. In Figure 10-30 on page 203, we read the test file that was created in the NPIV1 guest and created a test text file in the /dev/sda2 partition.



```
# mkdir npiv2a1 npiv2a2
# mount -o ro /dev/sda1 npiv2a1
# mount /dev/sda2 npiv2a2
# cat npiv2a1/testa1.txt
test1
# echo "test2" > npiv2a2/testa2.txt
# cat npiv2a2/testa2.txt
test2
```

Figure 10-30 Mounting the SCSI partitions on the NPIV2 Linux guest

> **Note:** If a shared disk partition is changed by a Linux image with read/write access, the other Linux images that have read-only access to that partition will not be updated until the disk partition is remounted.

Related publications

The publications listed in this section are considered particularly suitable for a more detailed discussion of the topics covered in this book.

IBM Redbooks

For information about ordering these publications, see “How to get IBM Redbooks” on page 206. Note that some of the documents referenced here may be available in softcopy only.

- *Linux on System z, Device Drivers, Features, and Commands February, 2007*, SC33-8289
- *Linux on zSeries: Fibre Channel Protocol Implementation Guide*, SG24-6344

Other publications

These publications are also relevant as further information sources:

- "SCSI initial program loading for zSeries" by G. Banzhaf, et al., *IBM Journal of Research and Development,* Vol 48, No. 3/4, 2004

Online resources

These Web sites and URLs are also relevant as further information sources:

- T10 Technical Committee

  http://www.t10.org
  http://www.t11.org
- I/O Connectivity

  http://www.ibm.com/servers/eserver/zseries/connectivity/
- Enterprise Volume Management System

  http://evms.sourceforge.net/
- Multipathing with SCSI disks

  http://www.ibm.com/developerworks/linux/linux390/perf/tuning_res_dasd_multipath.html

- LVM How To

  http://www.tldp.org/HOWTO/LVM-HOWTO/
- PuTTY download page

  http://www.chiark.greenend.org.uk/~sgtatham/putty/download.html
- VNC viewers
  - TightVNC

    http://www.tightvnc.com/download.html
  - RealVNC

    http://www.realvnc.com/download.html

How to get IBM Redbooks

You can search for, view, or download IBM Redbooks, Redpapers, Hints and Tips, draft publications and Additional materials, as well as order hardcopy Redbooks or CD-ROMs, at this Web site:

ibm.com/redbooks

Help from IBM

IBM Support and downloads

ibm.com/support

IBM Global Services

ibm.com/services

Index

IBM Redbooks

Fibre Channel Protocol for Linux and z/VM on IBM System z

Fibre Channel Protocol for Linux and z/VM on IBM System z

FCP concepts

FCP multipathing

N_Port virtualization

This IBM Redbooks publication builds upon the existing Linux for zSeries Fibre Channel Protocol Implementation Guide, SG24-6344-00, to produce a System z book that includes (among other things):

- A general introduction to Fibre Channel Protocol (FCP) concepts
- Operating system support and usage
- FCP naming and addressing
- FCP devices in the 2.6 Linux kernel
- N-Port ID Virtualization
- FCP Security topics

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

SG24-7266-00 ISBN 0738486620