Implementing Linux with IBM Disk Storage

Your guide to Linux implementations on IBM eServers

Use Linux efficiently with FAStT and ESS

Explore options relating to Linux in SANs

Bertrand Dufrasne
Ronald Annuss
James Goodwin
Paul McWatt
Arwed Tschoeke

Redbooks

ibm.com/redbooks

International Technical Support Organization

Implementing Linux with IBM Disk Storage

June 2003

SG24-6261-01

Note: Before using this information and the product it supports, read the information in "Notices" on page xiii.

Second Edition (June 2003)

This edition applies to the IBM Enterprise Storage Server, and the IBM FAStT Storage Server, for use with the Linux operating system (Red Hat Enterprise Advanced Server 2.1 and SuSE Linux Enterprise Server Edition 8) on IBM eServers.

Contents

Figures

Notices

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

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

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

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

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

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

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

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

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

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

Trademarks

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

- AIX 5L
- AIX®
- AS/400®
- Balance®
- DB2®
- developerWorks™
- ECKD™
- Enterprise Storage Server™
- ESCON®
- Everyplace™
- FICON™
- FlashCopy®
- IBM.COM™
- IBM®
- iSeries™
- Lotus®
- Notes®
- NUMA-Q®
- OS/2®
- OS/390®
- Parallel Sysplex®
- Perform™
- pSeries™
- Redbooks (logo) ™
- Redbooks™
- RS/6000®
- S/390 Parallel Enterprise Server™
- S/390 Parallel Enterprise Server™
- S/390®
- Seascape®
- ServeRAID™
- ServerProven®
- SP1®
- Tivoli®
- TotalStorage™
- eserver ™
- VM/ESA®
- X-Architecture™
- xSeries™
- z/Architecture™
- z/OS™
- z/VM™
- zSeries™

The following terms are trademarks of other companies:

Intel, Intel Inside (logos), MMX, and Pentium are trademarks of Intel Corporation in the United States, other countries, or both.

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

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

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

SET, SET Secure Electronic Transaction, and the SET Logo are trademarks owned by SET Secure Electronic Transaction LLC.

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

Preface

This IBM® Redbook explains the considerations, requirements, pitfalls, and possibilities when implementing Linux with IBM disk storage products.

This redbook presents the reader with a practical overview of the tasks involved in installing Linux on a variety of IBM @server platforms.

It also introduces the people who are already familiar with Linux to the IBM disk storage products.

Specifically, we worked with the Enterprise Storage Server™ (ESS) and the FAStT Storage Server, covering their installation and configuration with Linux distributions on the IBM eserver xSeries™, zSeries™, and pSeries™ models.

Configurations and practical experiments conducted by a team of professionals at the ITSO, San Jose Center, who implemented these Linux environments are documented in this book. We discuss the implementation of Linux from a storage perspective. It is expected that people who wish to work with Linux and storage subsystems will first become familiar with Linux and its installation documentation and support sources. We provide the basic steps required to get the storage subsystems in a state ready for Linux host attachment, and storage specific steps on preparing the Linux host for attachment to the storage subsystem. We also provide pointers and references to the documents and Web sites that will be of interest at the time you will be doing your implementation.

IT specialists in the field will find this book helpful as a starting point and reference when implementing Linux using the IBM disk storage servers.

Support considerations

Note: Before starting your Linux implementation activities using this redbook, you should check the latest availability status and documentation in respect to Linux, for the products and functions presented in this book.

You can consult your IBM Field Technical Support Specialist for the general support available. You can also find support information at the following Web site:

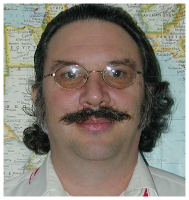
: http://www.storage.ibm.com/hardsoft/products/

The team that wrote this Redbook

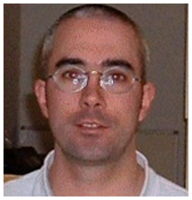
This redbook was produced by a team of specialists from around the world working at the International Technical Support Organization, San Jose Center.

Bertrand Dufrasne is a Certified Consulting I/T Specialist and Project Leader for Disk Storage Systems at the International Technical Support Organization, San Jose Center. He has worked at IBM for 21 years in many IT areas. Before joining the ITSO he worked for IBM Global Services in the US as an IT Architect. He holds a degree in Electrical Engineering.

Ronald Annuss works as an IT Specialist for IBM in Germany. He holds a diploma in Geodesy from the Technical University Berlin. He joined IBM in 1999 and has 8 years of experience with Linux. He is a RHCE and is working on Linux projects in Germany with zSeries and xSeries customers. He also is the co-author of a redbook and a Redpaper on Linux for zSeries.



James Goodwin is a Senior IT Specialist with Americas' Advanced Technical Support for Storage. He regularly develops and presents training materials for IBM Storage products. He has nearly 20 years of experience with open systems. He holds a degree in Mechanical Engineering from the University of New Mexico and has expertise in UNIX-like systems and storage. Prior to joining the ATS group, he worked in technical support for IBM NUMA-Q®.



Paul McWatt is an EMEA xSeries Advanced Technical Support Engineer based in Scotland. He has worked at IBM for 6 years and has 10 years experience of IBM Intel based servers. He is an RHCE and Advisory IT Specialist. His areas of expertise include Linux, storage, and High Availability Clustering. This is Paul's first redbook, but he has written two papers on similar subjects

Arwed Tschoeke is an IBM zSeries Systems Engineer, located in Hamburg, Germany. He worked for four years in the xSeries presales support team specializing in Linux and MSCS. He currently focuses on zOS and cross platform solutions. He holds a degree in Physics from the University of Kaiserslautern, Germany.

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

Bob Haimowitz
Roy Costa
International Technical Support Organization, Poughkeepsie Center

Chuck Grimm
IBM Technical Support Marketing Lead

Richard Heffel
IBM Storage Open Systems Validation Lab

Franck Excoffier
IBM Storage Open Systems Validation Lab

Rainer Wolafka
IBM Storage Open Systems Validation Lab

Timothy Pepper
IBM Storage Development

Mark S Fleming
IBM Storage Systems Level Test Lab

Nick Davis
IBM EMEA Linux Solutions Sales Manager, xSeries

Shawn Andrews
IBM x-Series World Wide Level 3 Support

Christopher M McCann
IBM xSeries World Wide Level 2 Support

Silvio Erdenberger
IBM xSeries Presales Technical Support Germany

Wendy Hung
IBM NOS technology support

Mic Watkins
IBM FAStT Storage Products Development

Marco Ferretti
IBM EMEA xSeries ATS - Education

James Gallagher
IBM pSeries Development Lab

Tan Truong
IBM pSeries Development Lab

Felica A Turner
IBM Raleigh - NOS Technology Support

Become a published author

Join us for a two- to six-week residency program! Help write an IBM Redbook 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'll develop a network of contacts in IBM development labs, and increase your productivity and marketability.

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

ibm.com/redbooks/residencies.html

Comments welcome

Your comments are important to us!

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

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

  ibm.com/redbooks

- Send your comments in an Internet note to:

  redbook@us.ibm.com

- Mail your comments to:

  IBM Corporation, International Technical Support Organization
  Dept. QXXE Building 80-E2
  650 Harry Road
  San Jose, California 95120-6099

Summary of changes

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

Summary of changes
for SG24-6261-01
for Implementing Linux with IBM Disk Storage
as created or updated on June 27, 2003.

May 2003, Second Edition

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

New information

- Linux LPAR installation for IBM eserver zSeries and FCP attachment to the ESS
- Linux implementation on IBM eserver r BladeCenter
- Linux implementation on IBM eserver pSeries
- VMware ESX
- BladeCenter SAN utility

Changed information

- Covers Red Hat Enterprise Linux Advanced Server 2.1 and Suse Linux Enterprise Server (SLES) 8.
- FAStT Storage Manager 8.3 and ESS Specialist

1

Introduction to Linux

This chapter provides a brief introduction to Linux:

- Historical perspective
- Open source
- Linux distributions
- IBM’s commitment to Linux
- Summary

1.1 Historical perspective

Years ago when the time came to install an operating system, a good plan was to have some activity to occupy one's time during the tedious phases. With the advent of fast processors, high-bandwidth networks, and high-speed drives for installation media, the luxury of "popping in a tape and going out to lunch" has been reduced to "kickstart the installation and read the splash screens for ten minutes." Readers may choose to save this chapter as a possible diversion during these ever-shorter installation periods.

Nearly everything one reads about Linux nowadays begins with the ritual invocation of "... an operating system created in 1991 as a hobby by Linus Torvalds at the University of Helsinki..." While this is true, it does not do justice to the significance of the work and its broad implications for today's implementers. We know that Linux refers to a UNIX-like operating system, so let us begin with a brief overview of the development of portable operating systems and open source, to see the context in which Linux evolved.

1.1.1 UNIX and the culture of collaboration

In 1969, several AT&T Bell Labs employees[1] began work on an operating system (OS) that would come to be called UNIX. A significant novelty of their development was that the OS was portable across hardware platforms. Prior to this, it was more typical to emulate the (usually) older hardware on a new system, so the OS would run unchanged, or to rewrite the operating system completely for the alternate hardware. Such portability was achievable only through writing most of the OS in a higher level language "above" the hardware. Unlike an OS written in assembly language for a particular architecture, this abstraction[2] and the language ("C") they developed to implement it permitted the study of their OS without much regard to hardware specifics, and by 1976 UNIX was being taught in classes on operating systems at the university[3] level.

At the time, AT&T, for various legal reasons, was permitting free academic access to the source code to UNIX while charging over $20,000 (in 1976 dollars!) for commercial or government access. AT&T halted publication of the source code in university texts as this revealed proprietary Bell Labs code. The era of collaborative programming had arrived.

[1] Ken Thompson, Dennis Ritchie, and J.F. Ossanna

[2] For an alternative view of this principle, see *The Linux Edge,* Linus Torvalds, in Open Sources: Voices from the Open Source Revolution, O'Reilly, 1999 ISBN 1-56592-582-3

[3] Prof. John Lions, University of New South Wales, Australia

1.1.2 GNU and free software

In the US, the academic variant of interest became the Berkeley Systems Distribution (BSD),[4] where virtual memory and networking were added. These advancements permitted large collaborative projects with contributors being scattered throughout the world. Lawsuits eventually ensued among AT&T, the Regents of the University of California, and other parties over access to and distribution of the OS source code. Such constraints on intellectual property rights to code, along with commercialization of academic artificial intelligence projects in the late 1970s and early 1980s, provided strong motivation for one researcher[5] from the Artificial Intelligence Laboratories at the Massachusetts Institute of Technology to write an operating system that was both portable and also would be licensed in a manner that would prevent its eventual constraint by intellectual property claims.

The new OS was to be named GNU, a recursive acronym for “Gnu’s Not UNIX.” This work would be “copylifted” instead of copyrighted; licensed under the GNU General Public License[6] (GPL), which stipulates that all programs run under or derived from GNU must have their source code published freely, relinquishing rights of control while retaining rights of ownership. This was the birth of free (as in freedom) software, in contrast to software in the public domain.

By this time, vendors such as Sun, Hewlett-Packard, and IBM had proprietary commercial offerings derived from licensed AT&T UNIX that were gaining popularity with corporate customers. The nascent GNU development effort began by making tools such as editors, compilers and file utilities available in source form that could be compiled and executed on any platform, standardizing and improving upon those offered by commercial vendors. Around 1990, programmers had contributed a nearly complete operating environment to GNU, with the exception of a kernel. The GNU kernel was to be based on a microkernel architecture for improved portability. This approach required an (arguably better) architecture completely different from that of a monolithic kernel. The GNU kernel project is known as the Hurd. In the words of its principal architect:

> “The Hurd project was begun before Linux was a twinkle in Linus Torvalds' eye, but because it is a more challenging task, and because we were less adept at mobilizing large-scale volunteer excitement, the Linux kernel was developed and deployed much sooner.”[7]

[4] *Twenty Years of Berkeley UNIX: From AT&T-Owned to Freely Redistributable,* Marshall Kirk McKusick, in Open Sources: Voices from the Open Source Revolution, O’reilly, 1999 1-56592-582-3

[5] *The GNU Operating System and the Free Software Movement,* Richard M. Stallman, ibid.

[6] Details at the Free Software Foundation, http://www.fsf.org/licenses/licenses.html

[7] Thomas Bushnell, BSG, in a letter to *Technology Review*, March/April 1999

1.1.3 Linux

With the withdrawal of AT&T source code from the university environment, and the BSD operating system mired in legal challenges by AT&T to its claim to be unencumbered by proprietary code, a small scale, UNIX-like skeleton of an operating system[8] called Minix was published in a text to be used as a teaching tool. It is here that Linus Torvalds enters the story. He decided to write a UNIX-like OS with improved functionality over that of Minix to run on readily available personal computers. His purpose was to teach himself the C language and improve his understanding of operating system design. He and colleague Lars Wirzenius published their source code under the GPL on the Internet for public comment, and Linux was born.

Linux was a kernel without utilities, GNU was an operating environment lacking a finished kernel, and unencumbered non-kernel BSD pieces were available to complete the picture. In short order the components were combined with installation and maintenance tools and made available by distributors; the first of serious note being Slackware[9] in 1993, followed by many others, making the GNU/Linux (or simply Linux, as it has since come to be known) combination readily available. In only a few years, a worldwide Linux community[10] has evolved, comprised of programmers and users attracted by the reliability and flexibility of this “free” operating system.

The term “open source” began to replace the term “free software” as the commercial adoption of GNU/Linux grew. There is in fact a difference, upon which hinges the fate of commercial ventures in this arena.[11]

1.2 Open source

There is a difference between open source and the GNU General Public License as noted on the GNU Web site (http://www.gnu.org). Linux was developed under GNU, which has articulated a philosophy that defines “free code” – the user’s right to use the code – rather than defining what they cannot do, which is the case with proprietary software. This license allows the user to alter, distribute, and even sell the code covered under its license as long as they allow those to whom they distribute the code to do the same.

[8] Dr. Andrew S. Tannenbaum, Vrije Universiteit, Amsterdam, The Netherlands

[9] compiled by Patrick Volkerding

[10] http://counter.li.org estimates 18 million Linux users as of April 2003

[11] e.g., X386 vs. XFree86 - http://www.xfree86.org/pipermail/forum/2003-March/000191.html or the evolution of Cygnus Support

The General Public License promotes free code on the GNU Web page. It also provides protection for the developer and prevents a user from altering the code and then asserting proprietorship over the code. This does not mean the code cannot be sold. According to the GNU Web site,[12] "free software" allows a user to run, copy, distribute, study, change, and improve the software. It must also be available for commercial use.

Standards enable communication among heterogeneous software and systems. Open source can be viewed as a manifestation of this process, and the process itself as a necessity for extending the development of inter-networking. When there is a need for new software or hardware support, or a defect is discovered and reported, the software (creation or correction) can be swiftly done by the user who required the changes (or by the original author), with no need for a design review, cost analysis, or other impositions of a centralized development structure. This is made possible by open source code. IBM has recognized the efficacy of this community and sees the benefit of the rapid and innovative development of robust and stable code to provide the enabling layer for e-business applications.

As a result of the evolutionary development of Linux, pieces of the code are located on various Web sites. Without some integration, it is difficult to install and upgrade the product, keep track of module dependencies, and acquire drivers for the hardware. *Distributions* provide coherent bundles of software in one location that furnish the capabilities needed for a usable server or desktop. Generally, the products of different distributors have a lot of things in common, but there may be particular features or functions that you require that are not available from a given distributor.

1.3 Linux distributions

Linux is available in many configurations from a variety of distributors. Linux advocates have strong preferences for one distribution over the other. Distributions from, for example, Mandrake,[13] Debian,[14] and gentoo[15] are presently available and offer their own advantages in package management, installation, development environment, and ease of use of various tools.

The distributions we used in our investigations were the so-called enterprise distributions, one from Red Hat Software Inc.,[16] and the other from SuSE Holding AG.[17] Both of these distributors also offer workstation distributions,

[12] http://www.gnu.org/philosophy/free-sw.html
[13] http://www.mandrake.com/
[14] http://www.debian.org/
[15] http://www.gentoo.org/
[16] http://ww.redhat.com/

whose relative merits or disadvantages (along with those of many other distributors) are discussed at length on the Internet.

1.3.1 Red Hat Enterprise Linux AS (formerly Advanced Server)

Enterprise Linux AS (formerly Red Hat Linux Advanced Server) is the core operating system and infrastructure enterprise Linux solution from Red Hat, supporting up to 8 CPUs and 16 GB of main memory, and is certified by DISA (US Defense Information Systems Agency) as COE (Common Operating Environment) compliant. It features High Availability Clustering and IP load balancing capabilities, asynchronous I/O support, Linux Standard Base interface conformity, improved SMP integration and reduced I/O-memory copy overhead. Red Hat offers support and maintenance services for their enterprise distributions.

1.3.2 SuSE Linux Enterprise Edition

SuSE Linux Enterprise Server (SLES) is a package distribution of UnitedLinux (see 1.3.3, “UnitedLinux) intended for server application. SuSE distributions tend towards standardized core content as defined by the Linux Standards Base (LSB).[18] SuSE also offers support and maintenance services for their enterprise distributions.

1.3.3 UnitedLinux

UnitedLinux[19] is a partnership of industry-leading Linux companies combining their intellectual property, sales, support, and marketing expertise to produce a uniform distribution of Linux designed for business.

Key elements of the UnitedLinux distribution include POSIX standard asynchronous I/O (AIO), raw I/O enhancements that provide high-bandwidth, low-overhead SCSI disk I/O, and direct I/O that moves data directly between the userspace buffer and the device performing the I/O, avoiding expensive copy operations, and bypassing the operating system's page cache.

Other functionality in focus includes hyper-threading to enable multi-threaded server software applications to execute threads in parallel within each individual server processor; large memory support to take advantage of the Intel Physical Address Extension to support up to 64 GB of physical RAM, and the full 4 GB of virtual addressing space per process; Internet Protocol Version 6 (IPv6), the next generation protocol designed by the IETF to replace the current version Internet

[17] http://www.suse.com/
[18] http://www.opengroup.org/lsb/cert/
[19] http://www.unitedlinux.com

protocol; and LDAPv3, the latest available directory services protocol for better user base management and application integration for mail servers and authentication servers, for instance.

1.4 IBM's commitment to Linux

The history of IBM's involvement with Linux is so large as to be outside the scope of this work. IBM is not a distributor of Linux. IBM is a contributor to Linux, a supplier of servers that run Linux, and a provider of support for customers who choose to use Linux. IBM's commitment to Linux may best be illustrated with some of the many notable highlights over the last several years.

As early as March 1999, IBM had announced key alliances, products, and global support for Linux, including distributors Caldera Systems Inc., (now the SCO Group[20]), Pacific HiTech Inc., (now Turbolinux, Inc.[21]), SuSE Holding AG (all of which later founded UnitedLinux along with distributor Conectiva S.A.[22]), and Red Hat Software Inc.[23] The IBM Thinkpad 600 became the first notebook computer listed as supported hardware by Red Hat.

Linux made its official debut on S/390® servers in May 2000, with commercial deployment by a major ISP for server consolidation following shortly thereafter. In June, SuSE announced Enterprise Linux for the RS6000.

In 2001, IBM won top honors at LinuxWorld for zSeries and iSeries™ systems running Linux, completing the availability of Linux across the entire IBM eServer product line. Complementing this, was the announcement of a broad range of storage options for Linux servers, including the Enterprise Storage Server (ESS) and Fibre Array Storage System (FAStT) external storage arrays. IBM's Linux supercomputer systems were deployed in research and also in oil and gas exploration, and prepackaged Linux clusters for e-business made their debut.

In 2002, continued emphasis on robust computing for small and medium businesses, server consolidation, retail, network infrastructure, advertising, telecom, and life sciences strengthened IBM's position as the most committed Linux player across all market segments. By December 2002 IBM had made the IBM eServer p630 available; the first pSeries system dedicated to Linux support. Internally, IBM began migrating its own business-critical services to Linux, demonstrating its indisputable commitment, and also demonstrating the scalability of Linux e-business solutions.

[20] http://www.sco.com/

[21] http://ww.turbolinux.com/

[22] http://www.conectiva.com/

[23] See press releases at http://www-916.ibm.com/press/prnews.nsf/homepage

Opening in 2003 were the announcement of the new Linux efforts, ranging from Linux supercomputing on demand, to bringing Linux to consumer electronics such as PDAs and handheld computers through the definition of a reference installation and IBM software for Linux (Websphere Micro Environment, Power Manager, DB2® Everyplace™, Tivoli® Device Manager, and IBM's Service Manager Framework).

In February 2003 IBM announced its commitment to "work with the Linux community to enter the Common Criteria (CC) certification process for the Linux operating system early this year, and proceed with a progressive plan for certifying Linux at increasing security levels through 2003 and 2004." With its delivery of enterprise middleware including DB2, Websphere, Tivoli and Lotus®, along with robust, secure, and scalable platforms, IBM also demonstrates its commitment to Linux in government, thus completely spanning the marketplace with pervasive support for Linux at every level.

1.5 Summary

Linux is an efficient multi-tasking operating system that can run on a minimal amount of hardware by comparison to today's standards. It can run very well on a 486 processor, for example. It is a multi-user system, and as such, offers various levels of built-in security. It is also a stable system that is capable of sustained up-time. For additional information, some further reading is noted:

The Evolution of the UNIX Time-sharing System, Dennis Ritchie, ca. 1979
(http://cm.bell-labs.com/cm/cs/who/dmr/hist.html)

Overview of the GNU Project, Free Software Foundation, ca. 1998
(http://www.gnu.org/gnu/gnu-history.html)

A Brief History of Hackerdom, Eric S. Raymond, ca. 1998
http://www.tuxedo.org/~esr/writings/hacker-history/

For more details on Linux in general, see:

- Linux Online offers a large amount of information on available applications, distributions, and locations for downloading the code (for free), documentation, education (including online courses), and information regarding hardware as well as a variety of other information: http://www.linux.org/apps/index.html
- The Linux Kernel Archives site is where the kernel sources can be downloaded from (http://www.kernel.org/) but it also contains links to various other related locations.

- The Linux Installation “HOWTO” site is sponsored by the “Linux Documentation Project” (LDP). LDP has free documentation and HOWTO documents that are short guides to do many of the tasks involved with Linux. This particular link is how to install Linux: `http://www.linuxdoc.org/HOWTO/Installation-HOWTO/`
- Linux International offers a nice description of the technical merits of the operating system at: `http://www.li.org`

2

Introduction to IBM TotalStorage

This chapter provides a brief overview and positioning of IBM TotalStorage™ disk products that can be used and managed in a Linux environment:

- IBM TotalStorage Enterprise Storage Server
- IBM TotalStorage FAStT Storage Server

We provide additional details for the models that we used during our experimentations, which are the ESS 800 and the FAStT 700, respectively.

2.1 Enterprise Storage Server

This section introduces the IBM TotalStorage Enterprise Storage Server and discusses some of the benefits that can be achieved when using it. For more detailed information, please refer to the redbook: *IBM TotalStorage Enterprise Storage Server Model 800*, SG24-6424.

2.1.1 ESS Overview

The IBM TotalStorage Enterprise Storage Server (ESS) is IBM's most powerful disk storage server, developed using IBM Seascape® architecture. The ESS provides unmatchable functions for all the server family of e-business servers, and also for the non-IBM (that is, Intel-based and UNIX-based) families of servers. Across all of these environments, the ESS features unique capabilities that allow it to meet the most demanding requirements of performance, capacity, and data availability that the computing business may require.

The Seascape architecture is the key to the development of IBM's storage products. Seascape allows IBM to take the best of the technologies developed by the many IBM laboratories and integrate them, producing flexible and upgradeable storage solutions. This Seascape architecture design has allowed the IBM TotalStorage Enterprise Storage Server to evolve from the initial E models to the succeeding F models, and to the more recent 800 models, each featuring new, more powerful hardware and functional enhancements.

To meet the unique requirements of e-business, where massive swings in the demands placed on systems are common, and continuous operation is imperative, demands very high-performance, intelligent storage technologies and systems that can support any server application. The IBM TotalStorage Enterprise Storage Server has set new standards in function, performance, and scalability in these most challenging environments.

Figure 2-1 on page 13 shows a photograph of an ESS Model 800 with the front covers removed. At the top of the frame are the disk drives, and immediately under them are the processor drawers that hold the cluster SMP processors. Just below the processor drawers are the I/O drawers that hold the SSA device adapters that connect to the SSA loops. Just below the I/O drawers are the host adapter bays that hold the host adapters. At the bottom of the frame are the AC power supplies and batteries.

The ESS in this photo has two cages holding the disk drives (DDMs). If the capacity of this ESS was 64 or fewer disk drives, then the top right side of this ESS would have an empty cage in it. The photo clearly shows the two clusters, one on each side of the frame.

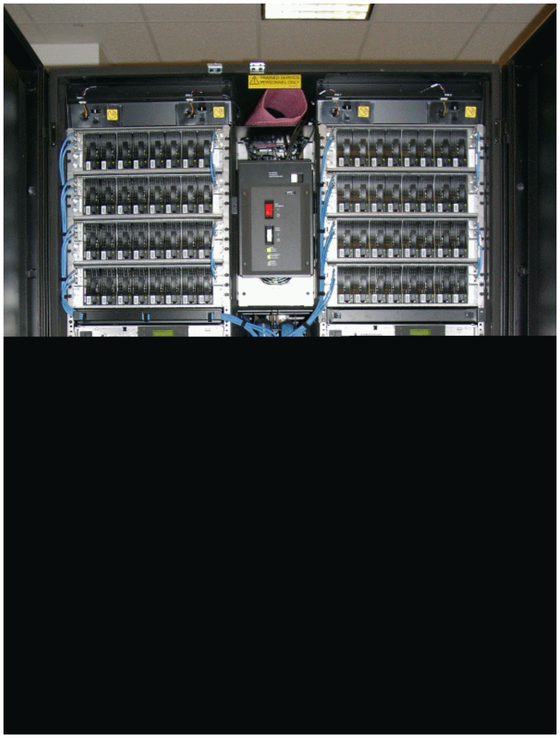


Figure 2-1 Photograph of ESS Model 800

Between the two cages of DDMs is an operator panel that includes an emergency power switch, local power switch, power indicator lights, and message/error indicator lights.

For larger configurations, the ESS base enclosure attaches to an expansion enclosure rack that is the same size as the base ESS, and stands next to the ESS base frame.

2.1.2 ESS features and benefits

The ESS set a new standard for storage servers back in 1999 when it was first available, and since then it has evolved into the F models, and the recently announced third-generation ESS Model 800.

The IBM TotalStorage Enterprise Storage Server Model 800 introduces important changes that dramatically improve the overall value of ESS, and provides a strong base for strategic Storage Area Network (SAN) initiatives.

Storage consolidation

The ESS attachment versatility —and large capacity— enable the data from different platforms to be consolidated onto a single high-performance, high-availability box. Storage consolidation can be the first step towards server consolidation, reducing the number of boxes you have to manage, and allowing you to flexibly add or assign capacity when needed. The IBM TotalStorage Enterprise Storage Server supports all the major operating systems platforms, from the complete set of IBM server series of e-business servers and IBM NUMA-Q, to the non-IBM Intel-based servers, and the different variations of UNIX based servers.

With a total capacity of more than 27 TB, and a diversified host attachment capability —SCSI, ESCON®, and Fibre Channel/FICON™— the IBM TotalStorage Enterprise Storage Server Model 800 provides outstanding performance while consolidating the storage demands of the heterogeneous set of server platforms that must be dealt with nowadays.

Performance

The IBM TotalStorage Enterprise Storage Server Model 800 integrates a new generation of hardware from top to bottom, allowing it to deliver unprecedented levels of performance and throughput. Key features that characterize the performance enhancements of the ESS Model 800 are:

- New more powerful SSA device adapters
- Double CPI (Common Platform Interconnect) bandwidth
- Larger cache option (64 GB)
- Larger NVS (2 GB non-volatile storage) and bandwidth
- New, more powerful SMP dual active controller processors, with a Turbo feature option
- 2 Gb Fibre Channel/FICON server connectivity, doubling the bandwidth and instantaneous data rate of previous host adapters

Efficient cache management and powerful back-end

The ESS is designed to provide the highest performance for different types of workloads, even when mixing dissimilar workload demands. For example, zSeries servers and open systems put very different workload demands on the storage subsystem. A server like the zSeries typically has an I/O profile that is very cache-friendly, and takes advantage of the cache efficiency. On the other hand, an open system server does an I/O that can be very cache-unfriendly, because most of the hits are solved in the host server buffers. For the zSeries type of workload, the ESS has the option of a large cache (up to 64 GB) and —most important — it has efficient cache algorithms. For the cache unfriendly workloads, the ESS has a powerful back-end, with the SSA high performance disk adapters providing high I/O parallelism and throughput for the ever-evolving high-performance hard disk drives.

Data protection and remote copy functions

Many design characteristics and advanced functions of the IBM TotalStorage Enterprise Storage Server Model 800 contribute to protect the data in an effective manner.

Fault-tolerant design

The IBM TotalStorage Enterprise Storage Server is designed with no single point of failure. It is a fault-tolerant storage subsystem, which can be maintained and upgraded concurrently with user operation.

RAID 5 or RAID 10 data protection

With the IBM TotalStorage Enterprise Storage Server Model 800, there now is the additional option of configuring the disk arrays in a RAID 10 disposition (mirroring plus striping) in addition to the RAID 5 arrangement, which gives more flexibility when selecting the redundancy technique for protecting the users' data.

Peer-to-Peer Remote Copy (PPRC)

The Peer-to-Peer Remote Copy (PPRC) function is a hardware-based solution for mirroring logical volumes from a primary site (the application site) onto the volumes of a secondary site (the recovery site). PPRC is a remote copy solution for the open systems servers and for the zSeries servers.

Two modes of PPRC are available with the IBM TotalStorage Enterprise Storage Server Model 800:

- PPRC synchronous mode, for real-time mirroring between ESSs located up to 103 km apart
- PPRC Extended DIstance (PPRC-XD) mode, for non-synchronous data copy over continental distances

Extended Remote Copy (XRC)

Extended Remote Copy (XRC) is a combined hardware and software remote copy solution for the z/OS™ and OS/390® environments. The asynchronous characteristics of XRC make it suitable for continental distance implementations.

Point-in-Time Copy function

Users still need to take backups to protect data from logical errors and disasters. For all environments, taking backups of user data traditionally takes a considerable amount of time. Usually, backups are taken outside prime shift because of their duration and the consequent impact to normal operations.

Databases must be closed to create consistency and data integrity, and online systems are normally shut down.

With the IBM TotalStorage Enterprise Storage Server Model 800, the backup time has been reduced to a minimal amount of time when using the FlashCopy® function. FlashCopy creates an instant point-in-time copy of data, and makes it possible to access both the source and target copies immediately, thus allowing the applications to resume with minimal disruption.

Storage Area Network (SAN)

The SAN strategy is to connect any server to any storage as shown in Figure 2-2. As SANs migrate to 2 Gb technology, storage subsystems must exploit this more powerful bandwidth. Keeping pace with the evolution of SAN technology, the IBM TotalStorage Enterprise Storage Server Model 800 is introducing new 2 Gb Fibre Channel/FICON host adapters for native server connectivity and SAN integration

These new 2 Gb Fibre Channel/FICON host adapters, which double the bandwidth and instantaneous data rate of the previous adapters available with the F Model, have one port with an LC connector for full-duplex data transfer over long-wave or short-wave fiber links. These adapters support the SCSI-FCP (Fibre Channel Protocol) and the FICON upper-level protocols.

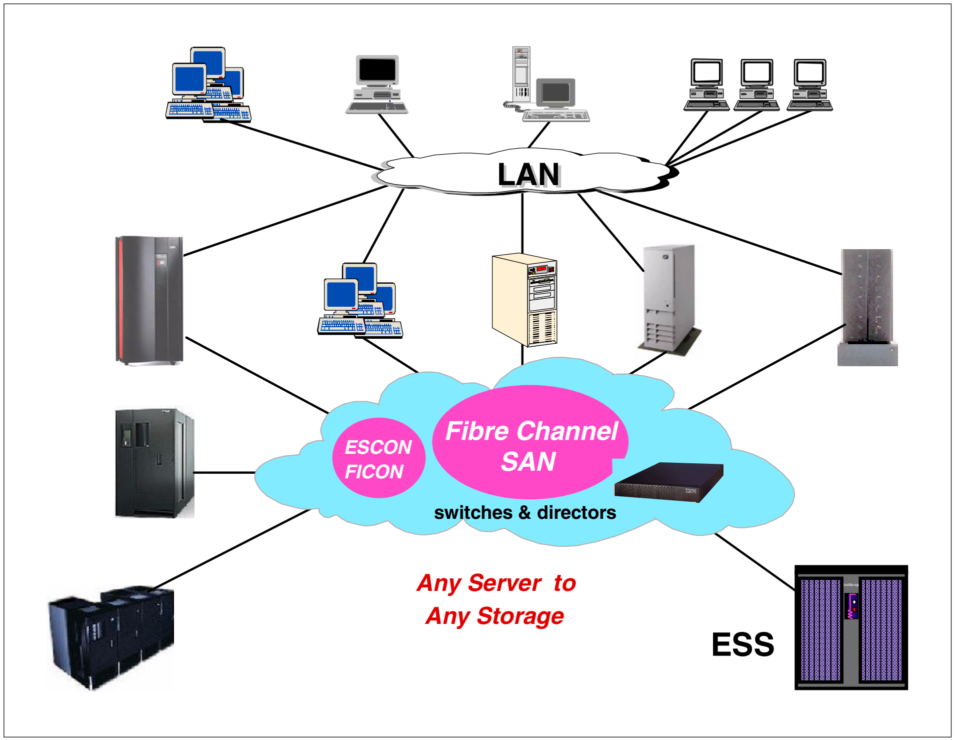


Figure 2-2 Storage area network (SAN)

Fabric support now includes the following equipment:

- IBM TotalStorage SAN switches IBM 3534 Model F08 and IBM 2109 Models F16, F32, S08 and S16
- McDATA “Enterprise to Edge” Directors (IBM 2032 Model 064) for 2 Gb FICON and FCP attachment (up to 64 ports)
- McDATA 16 and 12 port switches for FCP attachment (IBM 2031)
- INRANGE FC/9000 Director for FCP attachment (up to 64 and 128 ports) and FICON attachment (up to 256 ports) — IBM 2042 Models 001 and 128

The ESS supports the Fibre Channel/FICON intermix on the INRANGE FC/9000 Fibre Channel Director and the McDATA ED-6064 Enterprise Fibre Channel Director. With Fibre Channel/FICON intermix, both FCP and FICON upper-level protocols can be supported within the same director on a port-by-port basis. This new operational flexibility can help users to reduce costs with simplified asset management and improved asset utilization.

The extensive connectivity capabilities make the ESS the unquestionable choice when planning the SAN solution. For the complete list of the ESS fabric support, please refer to:

`http://www.storage.ibm.com/hardsoft/products/ess/supserver.htm`

For a description of the IBM TotalStorage SAN products, please refer to:

`http://www.storage.ibm.com/ibmsan/products/sanfabric.html`

2.2 IBM Fibre Array Storage Technology (FAStT)

IBM Fibre Array Storage Technology (FAStT) solutions are designed to support the large and growing data storage requirements of business-critical applications.

The FAStT Storage Server is a RAID controller device that contains Fibre Channel (FC) interfaces to connect the host systems and the disk drive enclosures. The Storage Server provides high system availability through the use of hot-swappable and redundant components. The Storage Server features two RAID controller units, redundant power supplies, and fans. All these components are hot-swappable, which assures excellent system availability. A fan or power supply failure will not cause downtime, and such faults can be fixed while the system remains operational. The same is true for a disk failure if fault-tolerant RAID levels are used. With two RAID controller units and proper cabling, a RAID controller or path failure will not cause loss of access to data. The disk enclosures can be connected in a fully redundant manner, which provides a very high level of availability. On the host side FC connections, you can use up to four mini-hubs.

The FAStT Storage Sever can support high-end configurations with massive storage capacities (up to 33 Tb per FAStT controller) and a large number of heterogeneous host systems. It offers a high level of availability, performance, and expandability.

2.2.1 FAStT models and expansion enclosure

At the time of writing this book, the FAStT900 and FAStT600 represented the newest addition to the FAStT family, complementing and joining the range of the earlier FAStT200, FAStT500, and FAStT700.

- The FAStT200 and FAStT600 are ideally placed for low-end mid-range setups such as workgroup and small department consolidations.

- The FAStT500 and FAStT700 are considered mid-range and are aimed at department storage consolidations, server clusters, and small to mid-range SANs.
- The FAStT900 is considered an entry-level enterprise product positioned for enterprise storage consolidation, mission critical databases, large SAN's, and high performance I/O.

The following tables show some of the differences between the FAStT storage server range.

Table 2-1 Machine matrix

	FAStT200	**FAStT500**	**FAStT600**	**FAStT700**	**FAStT900**
Machine Type	3542	3552	1722-60U/X	1742	1742-90U/X
Max Logical Drives	512	512	1024	2048	2048
Host Port Link Rate (Gb/sec)	1	1	2	1,2	1,2
Drive Port Link Rate (Gb/sec)	1	1	2	1,2	1,2
Max Physical Drives	66	100/220	42	224	224

Table 2-2 Performance comparisons

	FAStT200	**FAStT500**	**FAStT600**	**FAStT700**	**FAStT900**
Burst I/O rate cache reads (512 bytes)	11,800 IOPS	60,000	55,000	110,000	148,000
Sustained I/O rate disk reads (4k)	4,610 IOPS	20,000	21,500	31,300 1 Gb 38,000 2 Gb	53,200
Sustained I/O rate disk writes (4k)	1,290 IOPS	5,200	6,010	5,980 1 Gb 8,500 2 Gb	10,900
Sustained throughput cache read (512k)	190MB/s	400	--	424	800

	FAStT200	FAStT500	FAStT600	FAStT700	FAStT900
Sustained throughput disk read (512k)	170MB/s	370	380	380	772
Sustained throughput disk write (512k)	105MB/s	240	320	240	530

Table 2-3 FAStT700 verses FAStT900

	FAStT700	FAStT900
Processor	Celeron 566MHz	Pentium III 850MHz
Processor Memory	--	128MB
NVSRAM	32KB	128KB
Primary PCI Bus	32bit @ 33MHz	64bit @ 66MHz
Buffer Bus	0.5GB/s	1.6GB/s

2.2.2 FAStT700 features and benefits

For illustration and test purposes, during the writing of this book we used a FAStT700 and EXP700.

For more detailed information about these models, see the following redbook: *IBM TotalStorage FAStT700 and Copy Services*, SG24-6808.

The FAStT700 Storage Server has controllers which use the 2 Gbps Fibre Channel standard on the host side, as well as on the drive side. It connects via mini-hubs to the newer FAStT FC2-133 Host Bus Adapters (HBA) and the 2109 F16 Fibre Channel switch to give a full 2 Gbps fabric.

The FAStT700 attaches up to 220 FC disks via 22 EXP500 expansion units or up to 224 FC disks via 16 EXP700 expansion units to provide scalability for easy growth. To avoid single points of failure, it also supports high availability features such as hot-swappable RAID controllers, two dual redundant FC disk loops, write cache mirroring, redundant hot-swappable power supplies, fans, and dual AC line cords.

Figure 2-3 shows a rear view of the FAStT700, illustrating some of its components.

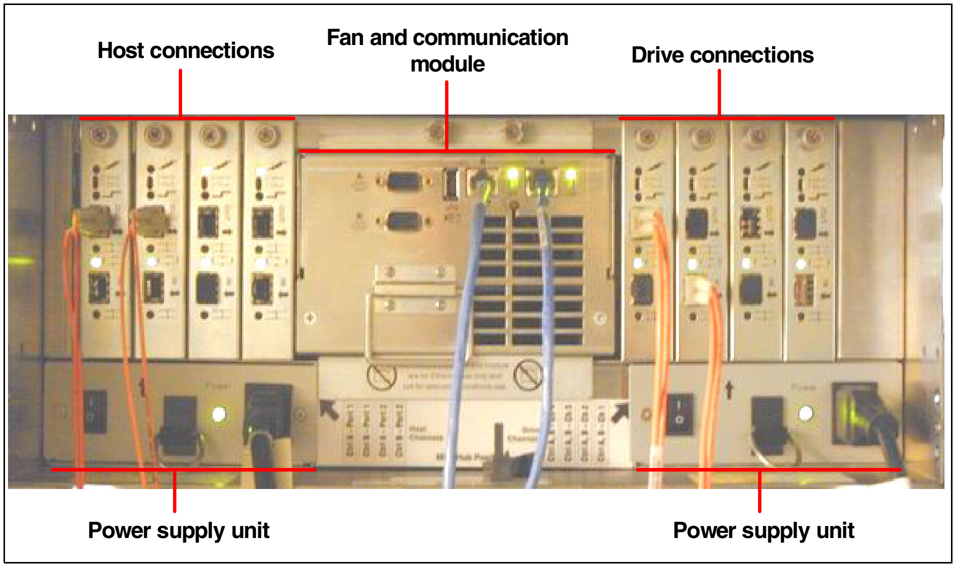


Figure 2-3 FAStT700 - Rear view

Using the latest Storage Manager software, the FAStT700 supports FlashCopy, Dynamic Volume Expansion, and remote mirroring with controller based support for up to 64 storage partitions. RAID levels 0,1,3,5, and 10 are supported, and for performance, it includes a 2 GB battery backed cache (1 GB per controller).

Storage Manager

To manage the FAStT700 storage server, use the IBM FAStT Storage Manager software. Refer to Chapter 9., “FAStT Storage Manager” on page 187 for details. At a glance, this software allows you to:

- Configure arrays and logical drives
- Assign your logical drives into storage partitions
- Replace and rebuild failed disk drives
- Expand the size of arrays and logical volumes
- Convert from one RAID level to another
- Perform troubleshooting and management tasks, like checking the status of FAStT Storage Server components, update the firmware or RAID controllers, and similar actions
- Configure and manage FlashCopy Volumes and Remote Volume Mirroring (FlashCopy and RVM are premium features that must be purchased)

It is also possible to use the serial interface and a terminal emulation utility. However, this is only meant for advanced troubleshooting and management. It should only be used when other management methods fail, and must be done under the supervision of IBM level 2 support.

The FAStT Storage Manager software supports two premium features that can be enabled by purchasing a premium feature key, as well as several new standard features. Some of these features include:

- **FlashCopy:** A premium feature that supports the creation and management of FlashCopy logical drives. A FlashCopy logical drive is a logical point-in-time image of another logical drive, called a *base logical drive*, in the storage subsystem. A FlashCopy is the logical equivalent of a complete physical copy, but you create it much more quickly, and it requires less disk space. Because FlashCopy is a host addressable logical drive, you can perform backups using FlashCopy while the base drive remains online and user accessible. In addition, you can write to the FlashCopy logical drive to perform application testing or scenario development and analysis. The maximum FlashCopy logical drives allowed is one half of the total logical drives supported by your controller model.
- **Remote Volume Mirroring:** A premium feature that is used for online, real-time replication of data between storage subsystems over a remote distance. In the event of a disaster or unrecoverable error on one storage subsystem, the Remote Volume Mirror option enables you to promote a second storage subsystem to take responsibility for normal input/output (I/O) operations. When RVM is enabled, the maximum number of logical drives per storage subsystem is reduced.
- **Dynamic logical drive expansion:** Enables you to increase the capacity of an existing logical drive
- **2048 logical drive support:** Enables you to increase the number of defined logical drives up to 2048 for each storage subsystem
- **Storage partitioning:** Supports up to 64 storage partitions

Storage Partitioning

Storage Partitioning allows you to connect multiple host systems to the same storage server. It is a way of assigning logical drives to specific host systems or groups of hosts. This is known as *LUN masking*.

Logical drives in a storage partition are only visible and accessible by their assigned group or individual hosts. Heterogeneous host support means that the host systems can run different operating systems. However, be aware that all the host systems within a particular storage partition must run the same operating system because they will have unlimited access to all logical drives in this partition. Therefore, the file systems on these logical drives must be compatible with the host systems.

3

SuSE Linux on zSeries

This chapter describes the implementation of SuSE Linux Enterprise Server for IBM zSeries Version 8 (SLES 8) and reviews disk storage attachment options supported under Linux. Special focus is on the newly available Fibre Channel Protocol (FCP) support for Linux on zSeries and the attachment of the IBM TotalStorage Enterprise Storage Server (ESS) using Fibre Channel.

This chapter discusses the following:

- Introduction to zSeries, Linux and storage options
- Hardware and software requirements
- The test environment
- Linux installation in a logical partition and as a z/VM™ guest
- Attachment of ESS through Fibre Channel

3.1 Introduction to zSeries and Linux

The IBM eserver zSeries is the successor of the S/390 series and IBM's premier server for e-business transaction workloads and has a long history in running commercial applications with unmatched reliability and availability. In 1999 Linux was ported to S/390 and became available in early 2000. Since then, many new features and device support have been added and more, and more customers are using Linux for S/390 and zSeries either for running new workloads or consolidating servers.

While Linux for zSeries supports the Extended Count Key Data (ECKD™) format on Direct Access Storage Devices (DASD) from the beginning, the support of Fibre Channel attached devices was added recently in February 2003. The aim of this book is to show the usage of IBM disk storage with Linux among the different IBM eServer platforms; in this chapter we will focus on the new zSeries Fibre Channel Protocol (FCP) support.

For more detailed information on zSeries and Linux in general, the reader should refer to the following redbooks: *Linux for S/390*, SG24-4987, and *Linux for IBM zSeries and S/390: Distributions*, SG24-6264, as well as to the SuSE installation manual.

Linux on zSeries can run in three different modes:

- Native using the whole machine
  - A native installation is rarely used since only one Linux system can run at any one time on the machine.
- In a logical partition (LPAR)
  - The zSeries processors support the partitioning of the native hardware into logical partitions. Logical partitions are an allocation of the available processor resource (either shared or dedicated), devices that are dedicated but can be serially switched between partitions and an allocation of memory. In LPAR mode, up to 15 operating system instances can concurrently run on one machine without interfering with each other.
- As a guest system under the hypervisor z/VM
  - When using z/VM as hypervisor to run operating systems as guests, only the hardware resources limit the amount of concurrently running instances.

The last two options are depicted in Figure 3-1

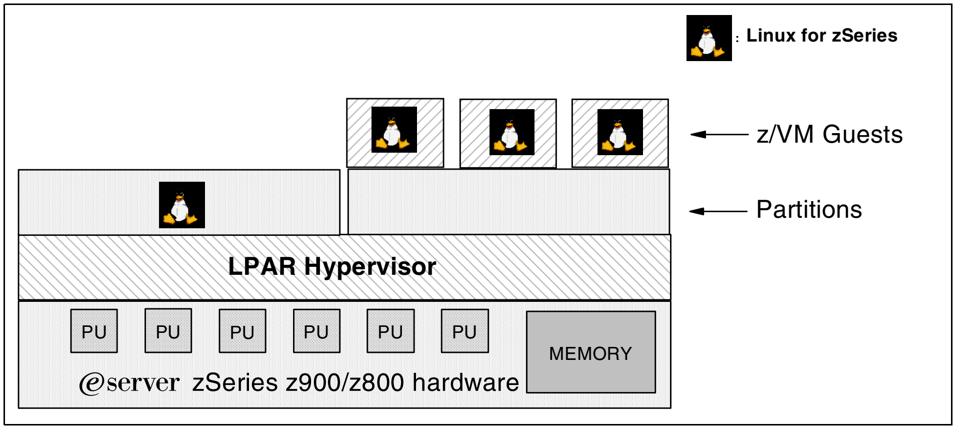


Figure 3-1 Running Linux on zSeries in an LPAR or as z/VM guest

3.1.1 zSeries specific storage attachments

The I/O component of z/Architecture™, inherited from its predecessors such as the ESA/390 architecture, is based on channels, control units, and devices. Channels provide a well-defined interface between a server and the attached control units. Originally implemented on parallel copper media, today's channels such as ESCON® (Enterprise System Connection) and FICON™ (Fibre Connection) use optical serial cables. Also, the parallel channel was a multi-drop interface, while both ESCON and FICON are switched point-to-point connections, extended to complex connection infrastructures via ESCON or FICON switches or directors, as depicted in Figure 3-2.

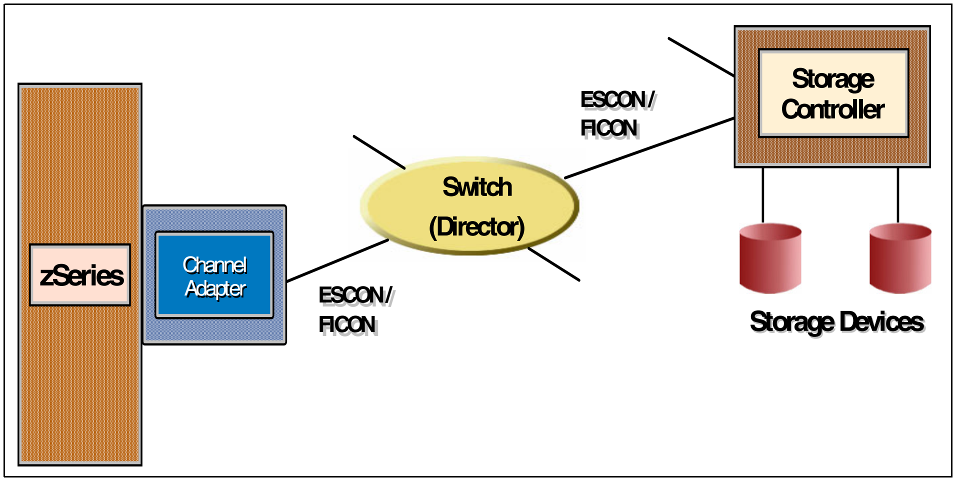


Figure 3-2 Typical zSeries storage attachment concept

ESCON channels use a unique physical interface and transmission protocol. FICON, on the other hand, is based on industry-standard Fibre Channel lower-level protocols.

In both cases, however, the higher-level I/O protocol used by software, based on channel programs consisting of channel command words (CCWs), is unique to mainframes adhering to z/Architecture or its predecessors. For access to disk storage, a specific set of CCWs is used, defined by the ECKD protocol. This protocol describes not only the commands and responses exchanged across the channel, but also the format of the data as it is recorded on the disk storage medium. For tape, similar CCW-based command protocols exist, but there are no associated access protocols for media such as DVDs or scanners, because the necessary software device drivers and control units have never been provided.

zSeries and storage device addressing

With z/Architecture, software addresses a storage device using a 16-bit device number, which uniquely identifies the device. Such a device number is mapped to one or more physical paths to the device. This path is typically described by an address quadruple consisting of:

1. **Channel path identifier**: (CHPID)
   Identifies the channel that provides a path to the device
2. **Physical link address**:
   Identifies a route through a switch or director
3. **Control unit address**:
   Specifies the control unit
4. **Unit address** (UA):
   Identifies the device

Please refer to the left side of Figure 3-6 on page 32.

In the case of a disk, this is often a logical device partitioned out of the disk space provided by an array of physical disks. In order to provide redundancy and/or increased I/O bandwidth, or to allow load balancing, there is typically more than one physical path to a device. While these paths are specified using different address quadruples, software still uses a single device number to address the device, independent of the path that is chosen for any particular I/O request.

Another characteristic of the z/Architecture I/O scheme and the ECKD architecture for disk, is the sophisticated support for sharing channels, control units, and devices among multiple operating systems, which may run on the same or different zSeries systems.

3.1.2 Distributed storage attachments

The attachment of storage controllers in the distributed world is predominantly based on the SCSI standard, using a control-block-based command and status protocol. Today, there is a clear distinction between the physical level and the command/status level of the SCSI protocol.

FCP advantages

Based on this higher-level SCSI command/status protocol, a new standard has been defined called FCP, which stands for Fibre Channel Protocol for SCSI.

FCP employs the SCSI command/status protocol on top of an underlying Fibre Channel transmission protocol. Due to the superiority of the optical Fibre Channel connection regarding speed, distance, and reliability, storage attachments via FCP have the highest growth rate in today's distributed storage world.

Traditionally, distributed storage controllers have been attached via parallel SCSI cabling, which allowed only very limited distances between the server and the controller. And although the SCSI architecture has provisions for physical controller and device sharing, the length constraints of parallel SCSI cables impose natural limits on this sharing capability. Also, there has been little need for a capability to share controllers and devices among multiple operating systems running concurrently on the same server, because such server virtualization techniques are just starting to develop in the distributed world.

FCP storage addressing

Traditional SCSI devices have a simple addressing scheme: originally, the device address, known as the target address was simply a number in the range 0-7 (extended to 0-15 in newer SCSI architecture). As shown in Figure 3-3 on page 30, SCSI also supports the definition of Logical Unit Numbers (LUNs); each LUN allows for a sub-addressing of the physical device.

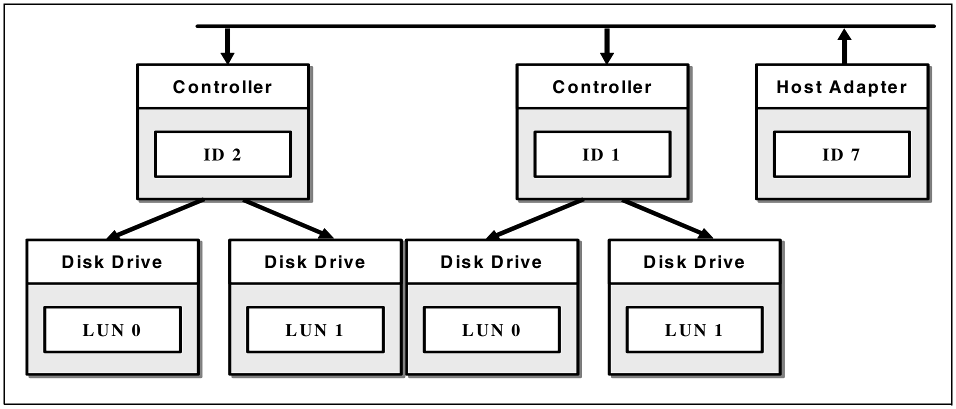


Figure 3-3 SCSI addressing

For FCP, a new addressing scheme, based on World Wide Names (WWN) was developed. Each node and each port (as an FCP port) on a node is assigned a WWN, respectively known as World Wide Node Name (WWMN) or World Wide Port Name (WWPN). This is illustrated in Figure 3-4.

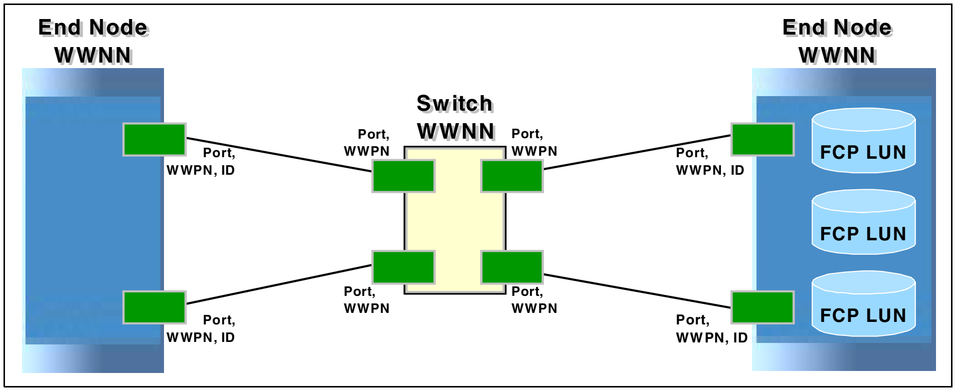


Figure 3-4 World Wide Names

In addition to WWNN and WWPN names, any addressable unit within a node (such as a disk drive) can be assigned a unit name, also know as LUN.

Linux FCP mapping

The format of a map entry is as follows:

```
0x6000 0x00000001: 0x5005076300c38550 0x00000000:0x5249000000000000
    |         |           |                |        |
    |         |           |                |        |
    |         |           |                |        +--FCP_LUN
    |         |           |                +---------- LUN nr for Linux
    |         |           |                               (user asssigned)
    |         |           +--------------------------- WWPN
    |         +---------------------------------------  SCSI ID for Linux
    |                                                     (user assigned)
    +--------------------------------------------------- Device nr for FCP
                                                              channnel
```

- The first element is the zSeries device number that must be defined in the IOCDS, and be attached the FCP channel that is attached to the FCP switch.
- The second element is the SCSI target number; it is user assigned. Normal usage is to start with address 1 and increment by 1 for each new WWPN used.
- The third element is the WWPN of the device containing the LUN, as seen by the FC switch.
- The fourth element is the LUN number to be used by Linux, and is also user assigned.
- The fifth element is the LUN number assigned by the node controller.

Figure 3-5 illustrates how these elements are related to actual Linux device names.

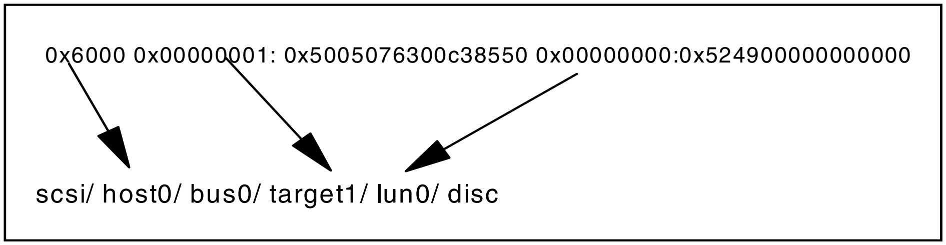

0x6000 0x00000001: 0x5005076300c38550 0x00000000:0x5249000000000000
scsi/ host0/ bus0/ target1/ lun0/ disc

Figure 3-5 FCP to Linux device name mapping

3.1.3 zSeries-specific versus distributed storage controllers

Two types of storage controllers with Fibre Channel interfaces are currently available for zSeries:

- Those supporting the z/Architecture specific FICON interface based on the CCW architecture, with the ECKD command set for disk storage, and similar sets of commands for tapes.
- Those supporting the SCSI-based FCP protocol, with SCSI command sets specific to the device type, such as disk or tape.

Both types of controllers are normally based on the same hardware building blocks. Their individual characteristics are achieved by different firmware handling the host interface. The way the DASD addressing or device mapping is handled, is illustrated in Figure 3-6 and is summarized in our discussion in 3.1.1, “zSeries specific storage attachments” and 3.1.2, “Distributed storage attachments”

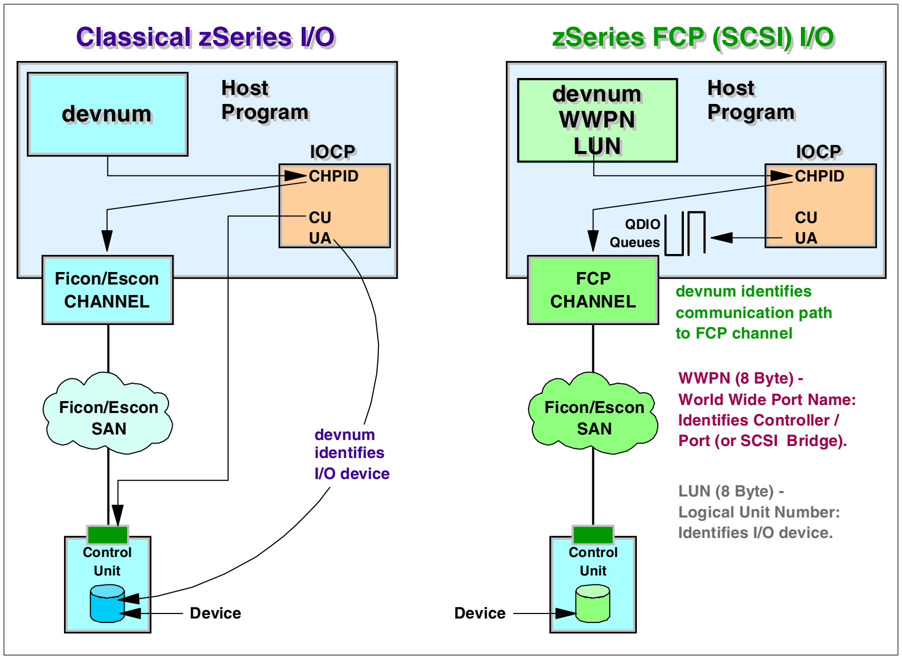

Classical zSeries I/O
devnum
Host Program
IOCP
CHPID
CU
UA
Ficon/Escon CHANNEL
Ficon/Escon SAN
devnum identifies I/O device
Control Unit
Device
zSeries FCP (SCSI) I/O
devnum WWPN LUN
Host Program
IOCP
CHPID
CU
UA
QDIO Queues
FCP CHANNEL
devnum identifies communication path to FCP channel
WWPN (8 Byte) - World Wide Port Name: Identifies Controller / Port (or SCSI Bridge).
LUN (8 Byte) - Logical Unit Number: Identifies I/O device.
Ficon/Escon SAN
Control Unit
Device

Figure 3-6 DASD addressing scheme

3.1.4 zSeries and storage: Summary

To summarize, classical zSeries operating systems such as z/OS™ and z/VM™ were designed for use only with storage controllers that support the I/O protocols defined by z/Architecture. This changed with the advent of Linux for zSeries, since its storage I/O component is oriented toward SCSI protocols. Lacking the capability to access SCSI-based storage devices on a zSeries server system, it was necessary to add specific support to Linux for zSeries to enable it to function in a CCW-based zSeries I/O environment. However, this additional layer in Linux for zSeries is unique to zSeries, and does not allow it to exploit storage subsystems and applications that are dependent on a SCSI attachment. For this reason, an FCP attachment capability has been added to the z800 and z900 systems, allowing the attachment of SCSI-based storage controllers and enabling Linux for zSeries to access these controllers in the Linux-standard manner. The Figure 3-7 shows the differences/format translation between the two access methods.

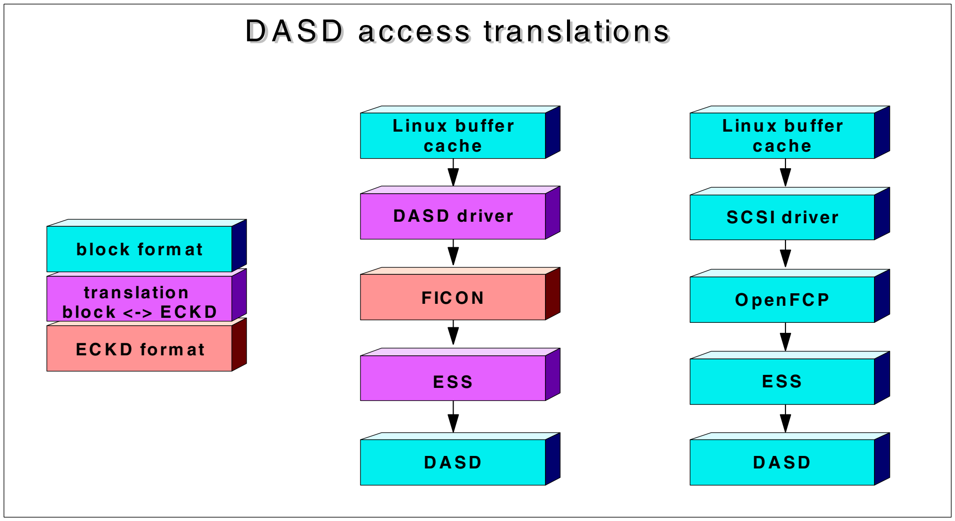


Figure 3-7 DASD access translations FICON verses FCP

3.2 System requirements

Linux for zSeries is supported on the z800 and z900 models and on the predecessor S/390 models G5, G6, and MP3000, which are shown in Figure 3-8.

Figure 3-8 IBM eServer zSeries

3.2.1 Hardware requirements

In this section we review the hardware requirements for running Linux on zSeries; and we successively look at the processor, memory, storage, and network specific requirements.

Processor

The heart of zSeries and S/390 hardware is the Multi-Chip Module (MCM), which contains up to 20 Processing Units (PU), commonly referred to on other platforms as CPUs or engines. All PUs are identical, but can take on different roles in the system. Each PU can be one of:

- A central processor (CP), sometimes called a standard engine
- A System Assist Processor (SAP)
- An Internal Coupling Facility (ICF) for Parallel Sysplex®
- An Integrated Facility for Linux (IFL), sometimes called an IFL engine

The IFL engine is a new type of PU reserved for Linux; it cannot run any other operating systems. It is priced lower than standard engines, and its presence does not affect the power ratings, hence, does not increase the cost of other software installed on that hardware.

Linux needs at least 1 PU configured as CP or IFL, but scales up to 16 processors on zSeries. The PUs can be either shared with other partitions (provided the LPARs are running Linux or z/VM) or dedicated.

Memory

The memory requirements are mostly dependent on the purpose of the server system and type of applications you are going to deploy on this system. For the installation, SuSE recommends the following:

- 128+ MB: For text mode installation
- 256+ MB: For graphical (X11) mode from nfs or smb source
- 512+ MB: For installation using VNC (graphical mode remotely displayed in Java-enabled Web browser) and FTP source

Storage

To hold the operating system, a certain amount of storage has to be in ECKD format since booting (or Initial Program Load - IPL) from Fibre Channel attached storage is not supported.

The types of zSeries DASD are referred to by their machine type and model number. For the 3390s, there are five possible model numbers: 1, 2, 3, 9, and 27. Models 3 and 9 are seen most commonly and are sometimes called a pack or volume. In the past, these DASDs were large physical devices, but today they are emulated by disk arrays with much larger capacities such as the ESS. A 3390 model 3 (or more simply, a 3390-3) has a capacity of approximately 2.8 GB, but when formatted for Linux, that is reduced to about 2.3 GB. A 3390 model 9 has a capacity of about 8.4 GB, but when formatted, that is reduced to about 7.0 GB.

For the operating system files, we need approximately 1.6 GB of DASD storage. Including the requirements for swap and some free space; the equivalent of 1 3390-3 volume is sufficient for a basic installation of SLES 8.

To attach storage via Fibre Channel you need a zSeries FCP channel. A FCP channel requires a FICON card (feature 2315 or 2318), or a FICON Express card (feature 2319 or 2320). This is the same channel hardware used for FICON channels, however, a different firmware load is required.

The type of firmware to be loaded into the FICON/FICON Express card, turning it into either an FCP channel or into one of the FICON type channels (FC or FCP), is controlled by the definition of the channel type for that particular channel in the IOCP (CHPID statement). Thus, by defining FCP type channels in the IOCP, the total number of FICON type channels that can be configured is reduced accordingly.

The two channels residing on a single FICON/FICON Express card can be configured individually, and each can be a different channel type.

For the LPAR installation of SLES 8 you can use the HMC CD-ROM or an FTP server, or if that is not possible, you have to prepare a tape to IPL from during the installation. Thus, you might need access to a tape unit such as a 3480, 3490, or 3590.

Network

Physical networking is typically done through one or more flavors of the Open Systems Adapters (OSA). Please refer to 3.2.3, "Connection requirements" on page 37.

3.2.2 Software requirements

The following MicroCode and software are required for the installation (as described in the SLES installation manual).

Microcode level and APARs/fixes

For installation under VM, you need at least VM/ESA® 2.4. For the 64-bit zSeries release you need z/VM 3.1 or higher. If you want to use Hipersockets under VM, you need z/VM 4.2 or higher and on 2064/z900 you need microcode EC E26949 level 013 or higher.

For the installation of SuSE Linux Enterprise Server 8 on IBM S/390 or zSeries, the following microcode levels and z/VM APARs are required:

OSA-Express QDIO

zSeries 900 GA3

Driver 3G, OSA microcode level 3.0A
MCLs: J11204.007 and J11204.008 (available May 03, 2002)

zSeries 900 GA2

Driver 3C, OSA microcode level: 2.26
MCLs: J10630.013 and J10630.014 (available May 20, 2002)

zSeries 800 GA1

Driver 3G, OSA microcode level 3.0A
MCLs: J11204.007 and J11204.008 (available May 03, 2002)

S/390 Parallel Enterprise Servers G5 and G6

Driver 26, OSA microcode level: 4.25
MCLs: F99904.032 and F99904.033 (available May 16, 2002)

VM/ESA and z/VM

z/VM 4.3

All necessary fixes and enhancements included.

z/VM 4.2

APAR: VM62938, PTF: UM30225
APAR: VM63034, PTF: UM30290

FCP support

FCP and SCSI controllers and devices can be accessed by Linux for zSeries with the appropriate I/O driver support. Linux may run either natively in a logical partition, or as a guest operating system under z/VM Version 4, Release 3.

> **Note:** z/VM Version 4, Release 3 is required to support FCP for Linux guests. However, z/VM itself does not support FCP devices.

3.2.3 Connection requirements

This section describes the connection requirements, such as console and network access needed to install SuSE Linux for zSeries.

Access to the Hardware Management Console (HMC) or a network connection to a virtual console in z/VM is needed to perform the installation. The intended use of the console is solely to launch the Linux installation. After Linux is running, use a Telnet connection directly to Linux for S/390 to logon to a Linux command shell and other applications.

For the Linux system a TCP/IP network connection is required in order to get files from the installation server and to Telnet into the Linux system. The connection can be one of the following:

- OSA (OSA-2, OSA Express)
- CTC (virtual or real)
- ESCON channels
- PCI adapter (emulated I/O, only on MP3000)

All network connections require the correct setup on both Linux for zSeries and the remote system, and a correct routing between both ends.

Since the installation is done via the network, you need an installation server, which can access the installation CDs either via FTP, NFS, or SMB.

Installing SuSE Linux Enterprise Server via non-Linux based NFS or FTP, can cause problems with the NFS/FTP server software. Especially Windows standard FTP server can cause errors, so we generally recommend that you install via SMB on these machines.

To connect to the SuSE Linux Enterprise Server installation system one of the following methods is required:

Telnet or *ssh with terminal emulation*

ssh or Telnet are standard UNIX tools and should normally be present on any UNIX or Linux system. For Windows, there is a Telnet and ssh client called Putty. It is free to use and is included on CD 1 in the directory /dosutils/putty. More information on Putty can be obtained at:
http://www.chiark.greenend.org.uk/~sgtatham/putty.html

VNC client

For Linux, a VNC client called vncviewer is included in SuSE Linux as part of the VNC package. For Windows, a VNC client is included in the present SuSE Linux Enterprise Server. You will find it in `/dosutils/vnc-3.3.3r9_x86_win32.tgz` of CD 1. Alternatively, use the VNC Java client and a Java-enabled Web browser.

X server

You will find a suitable X server implementation on any Linux or UNIX workstation. There are many commercial X-Window environments for Windows. Some of them can be downloaded as free trial versions. A trial version of MI/X (MicroImages X Server) can be obtained at:
http://www.microimages.com/mix

3.3 Configuration used for this redbook

For our tests we used a z800 2066-004 connected to a SAN switch 2032-064 (McData) connected to the ESS 2105-800. The configuration is shown in Figure 3-9.

The z800 was partitioned in to several LPARs, one of them running z/VM with Linux as a guest and one of them running Linux natively. The OSA adapter and FCP card are shared between those LPARs and can be used from either side with the same settings. The relevant settings (in the IOCDS, generated by the IOCP) for the FCP device are:

```
CHPID PATH=(15),SHARED,
      PARTITION=((LINUX1,LINUX2),(LINUX1,LINUX2)),TYPE=FCP
CNTLUNIT CUNUMBR=0600,PATH=(15),UNIT=FCP
IODEVICE ADDRESS=(600,064),CUNUMBR=(0600),UNIT=FCP
```

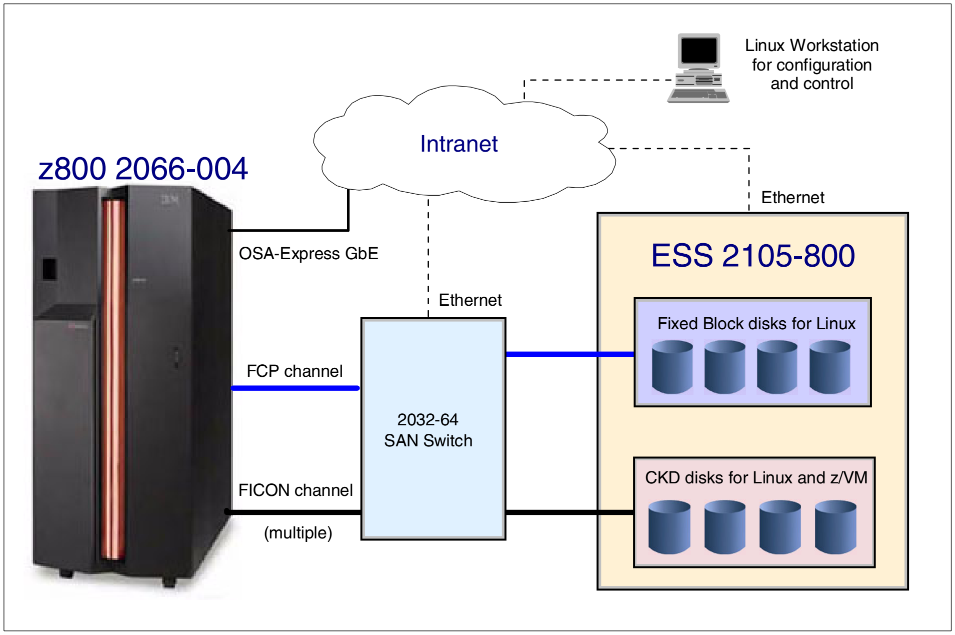


Figure 3-9 Hardware configuration

3.4 Implementing Linux

This section shows the installation of SuSE Linux Enterprise Server Version 8 (SLES 8) in an LPAR and as a z/VM guest and the attachment of the ESS via Fibre Channel. The necessary preparations and installation and customization steps are described in the following manner:

- Two separate sections, 3.4.1 “Initial steps for installation as a z/VM Guest” on page 40, and 3.4.2 “Initial steps for LPAR installation” on page 46 cover the initial steps specific to each installation type;
- A common section, 3.4.3 “Installation with YaST2” on page 51 describes the completion of the SuSE Linux installation using YaST2.
- The remaining section, 3.5 “zSeries and disk storage under Linux” on page 64 explains the necessary steps to attach and configure the ESS storage.

3.4.1 Initial steps for installation as a z/VM Guest

The initial installation of SLES 8 as a z/VM guest mandates the following sequence of tasks:

1. Create a Linux guest.
2. Upload the kernel image, the initial ramdisk, and the parameter file to the VM guest.
3. Create an IPL-script and IPL from virtual reader.
4. Perform the initial network setup on the VM virtual console.

After the initial system is up and running, the installation proceeds using the graphical installation program YaST2. (see 3.4.3, “Installation with YaST2” on page 51).

Create a Linux guest

To run Linux under z/VM, we have to create a new user (guest) in the z/VM directory. In our example, we created a guest with the following settings.

Example 3-1 z/VM directory entry for Linux guest

```
USER LINUXA XXXXX 512M 1G G
   ACCOUNT ITS30000
   IPL CMS PARM AUTOCR
   MACHINE XA 2
   DEDICATE 2C04 2C04 (OSA GbE Adapter - 2C04 thru 2C06)
   DEDICATE 2C05 2C05
   DEDICATE 2C06 2C06
   DEDICATE 0600 0600 (FCP - 0600 thru 063F)
   ......
   DEDICATE 063F 063F
  CONSOLE 0009 3215
  SPOOL 000C 3505 A
  SPOOL 000D 3525 A
  SPOOL 000E 1403 A
  LINK MAINT 0190 0190 RR
  LINK MAINT 019E 019E RR
  LINK MAINT 019F 019F RR
  LINK MAINT 019D 019D RR
  MDISK 0191 3390 1959 50 VMLU1R MR
  MDISK 0202 3390 0001 200 LX3A43 MR
  MDISK 0201 3390 0201 3138 LX3A43 MR
```

Going through these settings, you can see that we have dedicated an OSA Express GbE adapter (2C04-2C06) and a FICON FCP adapter (0600-063F). In addition we have three minidisks: 191 as the home disk for the guest and 201, 202 for the Linux system and swap space.

For detailed information on how to create Linux guests under z/VM please refer to the redbooks *Linux for IBM zSeries and S/390: Distributions*, SG24-6264, and *Linux on IBM zSeries and S/390: Large Scale Linux Deployment*, SG24-6824.

Transfer setup files through FTP

We will now describe how to get the initial installation system up and running within the z/VM guest.

This is done through an FTP transfer.

For that purpose, you need access to an FTP client program on your z/VM guest. The minidisk 592 of user TCPMAINT contains such one. To access the disk, log on to your guest and issue the following commands on your virtual console:

```
LINK TCPMAINT 592 592 rr
ACC 592 592
```

Note: The above commands assume that there is no read password set on TCPMAINTs 592 disk, which is the default. If it is protected, or the settings on your system are different, you need to contact your z/VM administrator to give you access.

There are three files you need to transfer from the installation media to your home disk: the kernel image, the initial ramdisk and the parameter. Make sure that you transfer the kernel and the ramdisk in binary mode and with a record length of 80. The following example shows a extract of the procedure while highlighting the actual input in bold.

Example 3-2 FTP transfer of the three initial files

```
Ready; T=0.01/0.01 14:11:55
ftp gallium.almaden.ibm.com
VM TCP/IP FTP Level 430
Connecting to GALLIUM.ALMADEN.IBM.COM 9.1.38.184, port 21
USER (identify yourself to the host):
stobdr2
>>>USER stobdr2
331 Please specify the password.
Password:
...
Command:
bin
...
cd /iso/s390/cd1/boot
```

```
...
locsite fix 80
...
get vmrdr.ikr sles8.image
...
get initrd sles8.initrd
...
asc
...
get parmfile sles8.parm
...
quit
```

Now the files are stored on your home disk, which you can verify using the **filel** command.

IPL from the virtual reader

The three files are used to IPL from the virtual reader and have to be put (punched) in the reader; you then IPL your system from the virtual reader. The easiest way to do this is to create a small REXX script as shown in Example 3-3.

Example 3-3 Script SLES8 EXEC to IPL from reader

```
/**/
’close rdr’
’purge rdr all’
’spool punch * rdr’
’PUNCH SLES8 IMAGE A (NOH’
’PUNCH SLES8 PARM A (NOH’
’PUNCH SLES8 INITRD A (NOH’
’change rdr all keep nohold’
’ipl 00c clear’
```

Execute the this script and you should see the Linux boot messages appear on the virtual console. Once the initial system has completed the boot process, the network configuration dialog will start automatically.

Initial network and installation setup

The network configuration dialog starts with asking for the type of the network device. In our example we used an OSA Express GbE adapter and network settings as shown in Table 3-1.

Table 3-1 Network settings

OSA device addresses	2C04, 2C05, 2C06
OSA portname	OSA2C00
Hostname	vmlinuxa.itso.ibm.com
IP address	9.12.6.73
Netmask	255.255.254.0
Broadcast address	9.12.7.255
Gateway address	9.12.6.92
DNS server	9.12.6.7
DNS search domain	itso.ibm.com
MTU size	1492

These settings have to be supplied in the dialog. The following example shows a walkthrough (a condensed extract) according to our setup; the actual input is highlighted in bold.

Example 3-4 Initial network setup

```
=                                                                    =
==-    Welcome to SuSE Linux Enterprise Server 8 for zSeries      -==
=                                                                    =

Please select the type of your network device:
0) no network
1) OSA Token Ring
2) OSA Ethernet
3) OSA-Gigabit Ethernet or OSA-Express Fast Ethernet
4) Channel To Channel
5) Escon
6) IUCV
8) Hipersockets
9) Show subchannels and detected devices
Enter your choice (0-9): 3
...
To set up the network, you have to read and confirm the license information
of the network device driver provided by IBM.
Do you want to see the license (Yes/No) ? yes
-------------------------------------------------------------------------------
International License Agreement for Non-Warranted Programs
...
------------------------------------------------------------------------------'
Do you agree with this license (Yes/No) ? yes
```

```
...
Enter the device addresses for the qeth module, e.g. '0x2c04,0x2c05,0x2c06'
(0x2c04,0x2c05,0x2c06): 0x2c04,0x2c05,0x2c06
Please enter the portname(case sensitive) to use(suselin7): OSA2C00
...
eth0 is available, continuing with network setup.

Please enter your full host name, e.g. 'linux.example.com' (linux.example.com):
vmlinuxa.itso.ibm.com
Please enter your IP address, e.g. '192.168.0.1' (192.168.0.1): 9.12.6.73
Please enter the net mask, e.g. '255.255.255.0' (255.255.255.0): 255.255.254.0
Please enter the broadcast address if different from (9.12.7.255): 9.12.7.255
Please enter the gateway's IP address, e.g. '192.168.0.254' (192.168.0.254):
9.12.6.92
Please enter the IP address of the DNS server or 'none' for no DNS (none):
9.12.6.7
Please enter the DNS search domain, e.g. 'example.com' (itso.ibm.com):
itso.ibm.com
Please enter the MTU (Maximum Transfer Unit), leave blank for default: (1500):
1492

Configuration for eth0 will be:
Full host name   : vmlinuxa.itso.ibm.com
IP address       : 9.12.6.73
Net mask         : 255.255.254.0
Broadcast address: 9.12.7.255
Gateway address  : 9.12.6.92
DNS IP address   : 9.12.6.7
DNS search domain: itso.ibm.com
MTU size         : 1492
Is this correct (Yes/No) ? yes

For security reasons you have to set an temporary installation
system password for the user "root".
You'll be asked for it only when you Telnet into this installation
system to limit the access to it and it will be cleared as soon
as you shut down or reset the installation system

Please enter the temporary installation password: XXXXXX
Temporary installation password set.
...
Network Setup finished, running inetd...

You should be able to login via Telnet now, for ssh wait a few seconds,
temporary host keys (only for installation) are being generated now:

Generating /etc/ssh/ssh_host_key.
Generating public/private rsa1 key pair.
```

```
...
```

After the network setup is complete, the creation of the ssh keys is started. Even though it is possible to log in now, you are advised to wait for the system to finish the key generation because additional information about your installation still needs to be supplied.

> **Note:** The root password you provide during this step is valid for the initial installation system only. Since it is only stored in memory, it will be gone after the first re-IPL from DASD, and you will have to provide a password once again.

The next dialogs prompt you for the type of server containing the installation CDs, and for the type of installation (graphical or text based). We used an FTP server as the installation source, and a Linux system running an X server to perform an X-based graphical installation.

Please refer to your Linux or UNIX documentation on how to set up X on your workstation. If you do not have a Linux, or UNIX, or Windows workstation running an X server, you can still use the VNC client provided on the installation CD for a graphical interface. Otherwise, just use an ssh client such as PuTTY, for a text based installation.

Example 3-5 shows the dialog relevant to our installation.

Example 3-5 Choose installation media and type

```
Please specify the installation Source:
 1) NFS
 2) SAMBA
 3) FTP
 0) Abort

Choice: 3

Please enter the IP-Number of the host providing the installation media:
9.1.38.184
...
Please enter the directory of the installation media: /iso/s390/cd1

Is the following correct?
 Installation Source: ftp
 IP-Address: 9.1.38.184
 Directory: /iso/s390/cd1

Yes/No: yes
```

```
Please enter the username for the FTP-access (for anonymous just press enter):
stobdr2
Please enter the password for the FTP-Access (for anonymous just press enter):
XXXXXXX
Is the following correct?

 FTP User: stobdr2
 FTP Password: XXXXXX

Yes/No: yes

Which terminal do want to use?

 1) X-Window
 2) VNC (VNC-Client or Java enabled Browser)
 3) ssh
 Choice: 1

Please enter the IP-Number of the host running the X-Server: 9.43.152.107
...
ramdisk /dev/ram0 freed

>>> SuSE Linux installation program v1.4 (c) 1996-2002 SuSE Linux AG <<<

Starting hardware detection...
                                                                  Searching for
infofile......
Loading data into ramdisk..............
....
 cintegrating the installation system into the ramdisk...
integrating the shared objects of the installation system...
integrating kernel modules of the installation system...
starting yast...
```

Once the installation system is loaded into memory, YaST2 is started automatically and is displayed on the X server. Please proceed with 3.4.3, “Installation with YaST2” on page 51.

3.4.2 Initial steps for LPAR installation

Within LPAR mode you have several options to IPL a system. While for zSeries operating systems like z/OS and z/VM, the IPL from a tape is preferred, the most convenient way to IPL a Linux system is to use the option to load from the HMC CD-ROM drive or from an FTP server. However, you can prepare a Linux IPL tape as well if you do not have access to the HMC CD-ROM or to a FTP server.

Please refer to the documentation of your distribution on how to create an IPL tape.

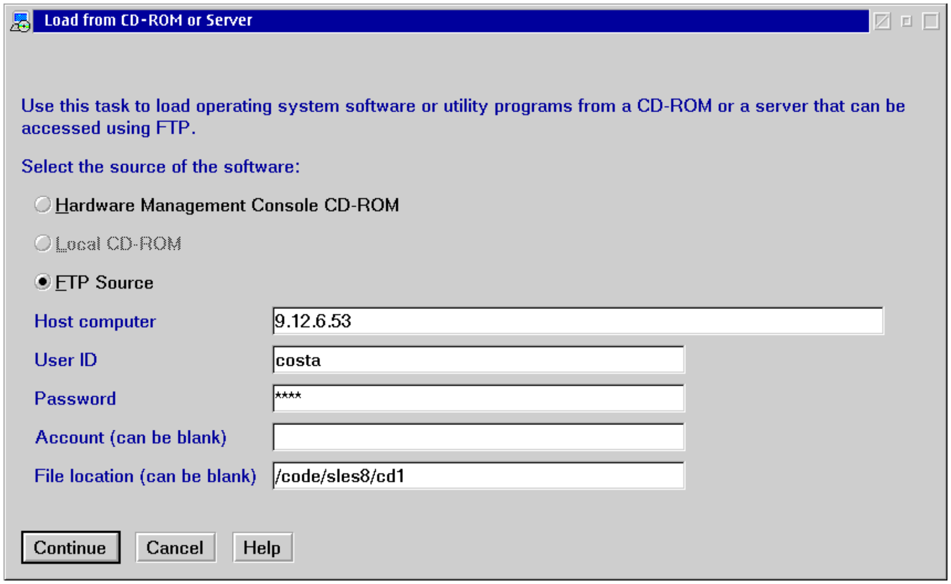
In our example we use an FTP server, which is accessible within the HMC network segment. The Linux installation CDs must be copied on this server. It is recommended to copy each CD in a separate directory, using the same initial path, but separating the different CDs in subdirectories named CD1, CD2, CD3.

Also, note that the first CD has the necessary information to allow the IPL: it contains the *suse.ins* file, which is a tape description file, used to emulate the CD as a tape for the zSeries.

To perform the IPL, log on to your HMC and then select the following:

1. Go into **Defined CPCs** and highlight the CPC you will work on.
2. Change to the **CPC Recovery** menu and double-click **Single Object Operations.**
3. Confirming the action in the next window will bring you on the CPC console.
4. Go into **Images** and highlight the LPAR to IPL.
5. Change to the **CPC Recovery** menu and double-click **Load from CD-ROM or Server.**

You will get a warning that all jobs will be cancelled. Make sure that no system is active or that it is safe to re-IPL the system. Clicking **Yes** opens the window shown in Figure 3-10. You have to provide the hostname or IP address of the FTP server, a valid user ID and password, as well as the path indicating the location of the first CD (i.e: /code/sles8/cd1 in our case).


Load from CD-ROM or Server

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

Select the source of the software:

○ Hardware Management Console CD-ROM

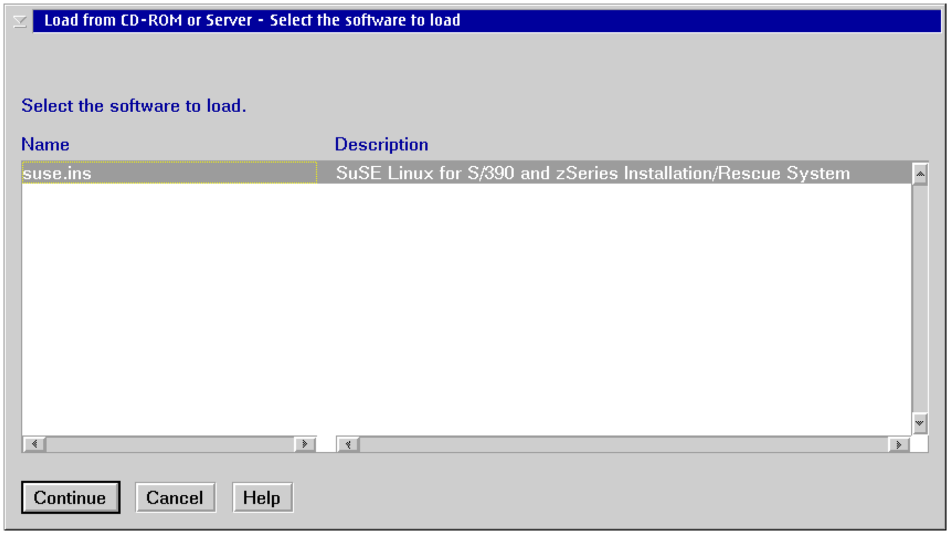
○ Local CD-ROM

● FTP Source

Host computer	9.12.6.53
User ID	costa
Password	****
Account (can be blank)	
File location (can be blank)	/code/sles8/cd1

Figure 3-10 HMC panel to load from FTP source

When you click **Continue**, the system verifies the installation source and provides a list of possible software to load. In the case of our example there is only one choice as shown in the Figure 3-11.


Load from CD-ROM or Server - Select the software to load

Select the software to load.

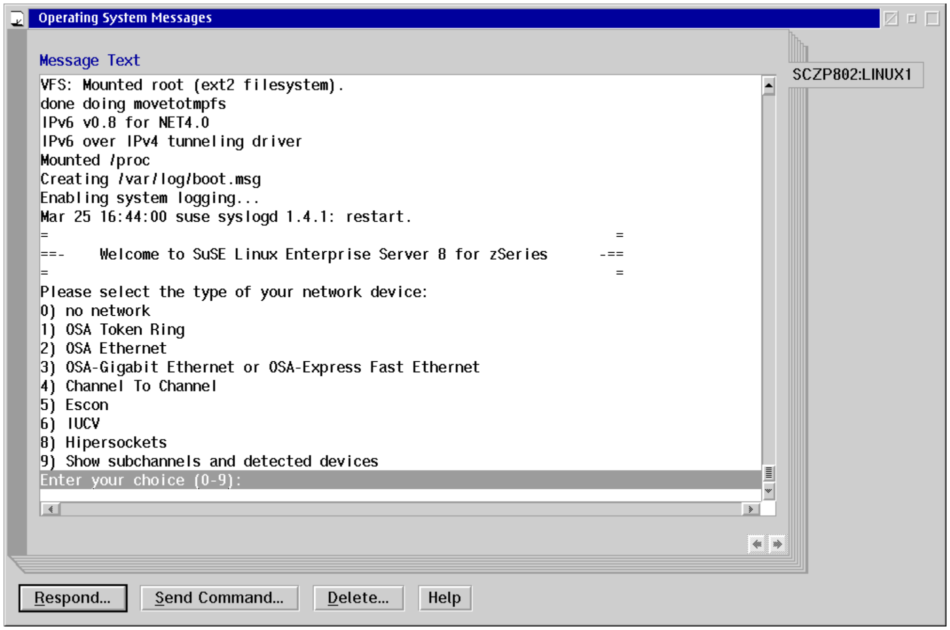
Name	Description
suse.ins	SuSE Linux for S/390 and zSeries Installation/Rescue System

Figure 3-11 Select the source

Do the following:

1. Select the appropriate source and click **Continue**.
2. Acknowledge the action by clicking **Yes** in the confirmation window that pops up.
3. The load process is started; when finished, a dialog window informs you of its completion. Click **OK** on this dialog window.
4. Go back to the HMC by ending your CPC Console session (ending the Single Object Operation).
5. On the HMC go into **CPC Images** and highlight your LPAR.
6. Change to the **CPC Recovery** menu and double-click **Operating System Messages**.

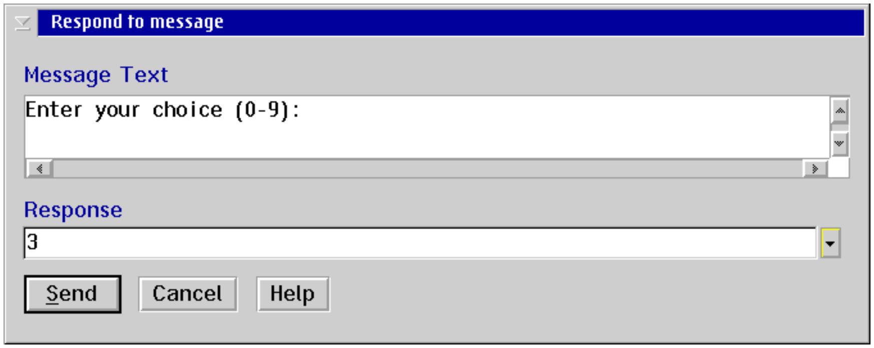
After a moment you will see the boot messages from Linux and the start of the network configuration dialog as shown in Figure 3-12.



```
VFS: Mounted root (ext2 filesystem).
done doing movetotmpfs
IPv6 v0.8 for NET4.0
IPv6 over IPv4 tunneling driver
Mounted /proc
Creating /var/log/boot.msg
Enabling system logging...
Mar 25 16:44:00 suse syslogd 1.4.1: restart.
=                                                            =
==-    Welcome to SuSE Linux Enterprise Server 8 for zSeries     -==
=                                                            =
Please select the type of your network device:
0) no network
1) OSA Token Ring
2) OSA Ethernet
3) OSA-Gigabit Ethernet or OSA-Express Fast Ethernet
4) Channel To Channel
5) Escon
6) IUCV
8) Hipersockets
9) Show subchannels and detected devices
Enter your choice (0-9):
```

Figure 3-12 HMC operating system messages interface

You have to click **Respond** first to be able to enter commands and respond to questions presented in the Message Text pane. This will give you an input window where you can type answers or input commands as shown in Figure 3-13.

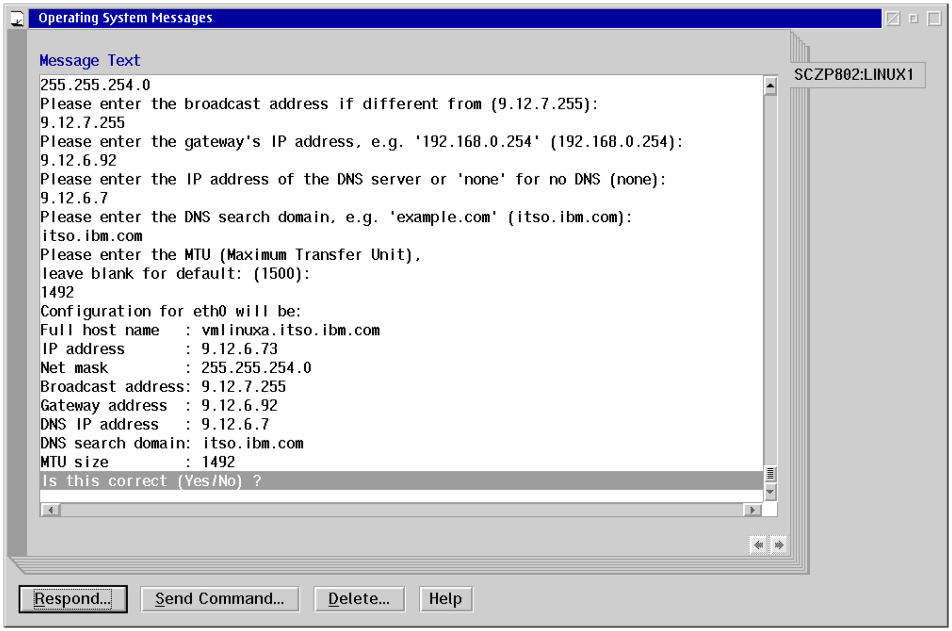


Enter your choice (0-9):

3

Figure 3-13 Enter commands on the HMC

To send the command to the Linux system click **Send**. Proceed in this manner until you have answered all the questions using the same answers as for the z/VM installation.[1] (See “Initial network and installation setup” on page 42) You can see the completed network dialog in Figure 3-14.

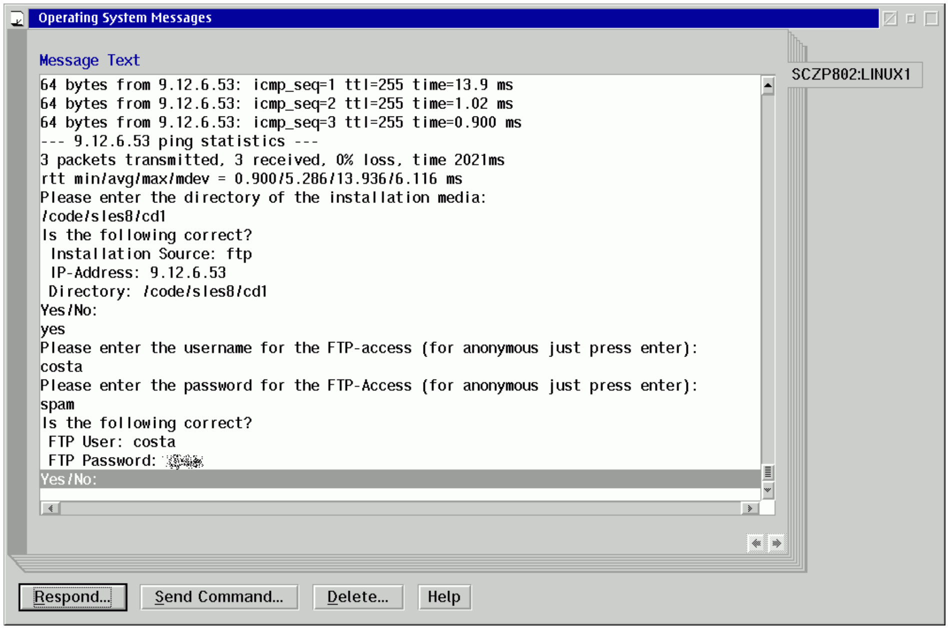


```
255.255.254.0
Please enter the broadcast address if different from (9.12.7.255):
9.12.7.255
Please enter the gateway's IP address, e.g. '192.168.0.254' (192.168.0.254):
9.12.6.92
Please enter the IP address of the DNS server or 'none' for no DNS (none):
9.12.6.7
Please enter the DNS search domain, e.g. 'example.com' (itso.ibm.com):
itso.ibm.com
Please enter the MTU (Maximum Transfer Unit),
leave blank for default: (1500):
1492
Configuration for eth0 will be:
Full host name   : vmlinuxa.itso.ibm.com
IP address       : 9.12.6.73
Net mask         : 255.255.254.0
Broadcast address: 9.12.7.255
Gateway address  : 9.12.6.92
DNS IP address   : 9.12.6.7
DNS search domain: itso.ibm.com
MTU size         : 1492
Is this correct (Yes/No) ?
```

Figure 3-14 Network settings

[1] For our example we used the very same network settings for the Linux LPAR installation as for the installation as a z/VM guest, because the devices such as OSA FCP adapter were the same and the z/VM Linux guest was not active during these tests.

Once the network settings are complete, you need to provide information regarding the installation server as shown in Figure 3-15.



```
64 bytes from 9.12.6.53: icmp_seq=1 ttl=255 time=13.9 ms
64 bytes from 9.12.6.53: icmp_seq=2 ttl=255 time=1.02 ms
64 bytes from 9.12.6.53: icmp_seq=3 ttl=255 time=0.900 ms
--- 9.12.6.53 ping statistics ---
3 packets transmitted, 3 received, 0% loss, time 2021ms
rtt min/avg/max/mdev = 0.900/5.286/13.936/6.116 ms
Please enter the directory of the installation media:
/code/sles8/cd1
Is the following correct?
 Installation Source: ftp
 IP-Address: 9.12.6.53
 Directory: /code/sles8/cd1
Yes/No:
yes
Please enter the username for the FTP-access (for anonymous just press enter):
costa
Please enter the password for the FTP-Access (for anonymous just press enter):
spam
Is the following correct?
 FTP User: costa
 FTP Password: [redacted]
Yes/No:
```

Figure 3-15 Installation server settings

Finally, you will have to specify the type of installation, either graphical or text based, you want to perform. (See “Initial network and installation setup” on page 42)

The initial steps for installation in an LPAR are completed at this point. For the second phase of the installation, using YaST, please proceed to the next section.

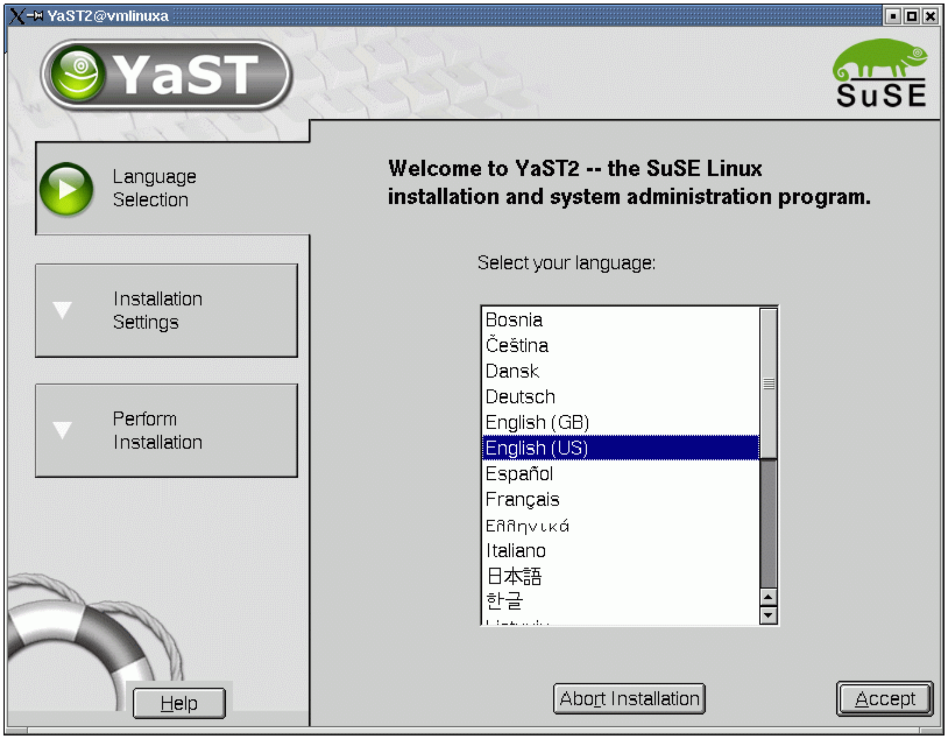
> **Note:** Once the installation with YaST is started, do not exit the program if you have to perform additional actions such as formatting DASD. It would end the entire installation procedure and you would have to do the LPAR IPL all over again. Instead use a parallel Telnet or ssh session to format DASDs or other additional tasks.

3.4.3 Installation with YaST2

After completion of the initial installation steps for either a z/VM or LPAR configuration, the second phase of the installation is performed using SuSEs installation program YaST, which will start automatically. In our example we used the graphical X based installation, but the procedures and menus for a VNC or text based installation are the same, just the appearance, obviously, is different. Please refer to the SuSE installation manual for the other types of YaST installations. The following screen captures were taken from the z/VM installation, so all the settings shown correspond to “Initial steps for installation as a z/VM Guest” on page 40.

YaST starts by displaying the End User License for the SuSE Linux Enterprise Server. You have to click **Accept** in order to proceed.

The first YaST2 installation screen that follows prompts you for the language selection as shown in Figure 3-16.



Welcome to YaST2 -- the SuSE Linux installation and system administration program.

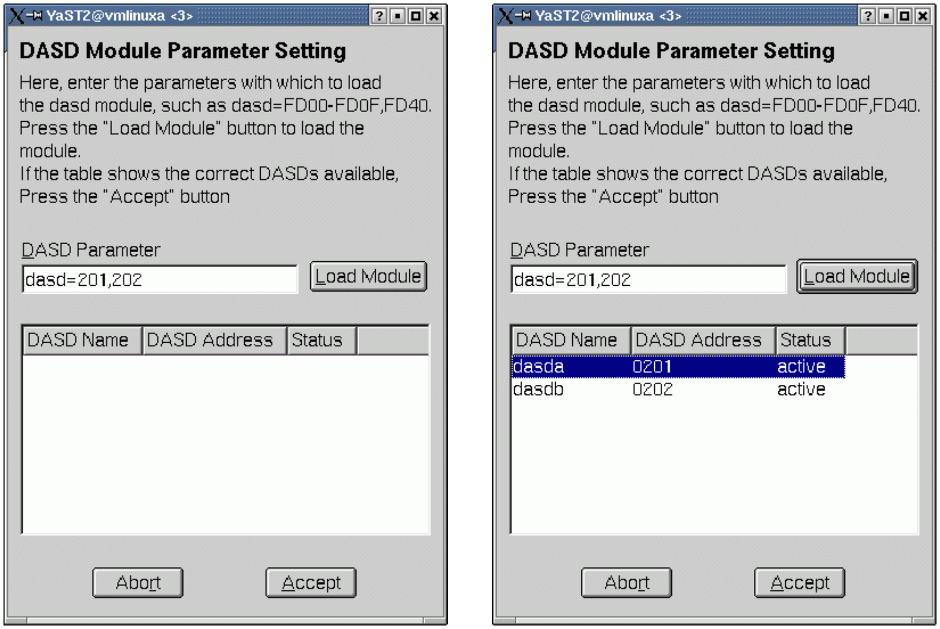
Select your language: English (US)

Figure 3-16 Language selection

Select the appropriate language and click **Accept**.

Now, you need to provide the parameters for the DASD driver to know on what DASDs to install Linux. Enter your numbers for the dasd volumes (in our case the z/VM minidisks 201 and 202; note that individual DASD device numbers are delimited by commas, and ranges of DASD are defined by placing a dash between the lowest and the highest address of the range.)

Click the **Load Module** button; after a successful load the DASDs will show up in the lower half of the window. The process is shown in Figure 3-17. Click **Accept** to proceed.



DASD Module Parameter Setting

Here, enter the parameters with which to load the dasd module, such as dasd=FD00-FD0F,FD40. Press the "Load Module" button to load the module.
If the table shows the correct DASDs available, Press the "Accept" button

DASD Parameter: dasd=201,202

DASD Name	DASD Address	Status

DASD Module Parameter Setting

Here, enter the parameters with which to load the dasd module, such as dasd=FD00-FD0F,FD40. Press the "Load Module" button to load the module.
If the table shows the correct DASDs available, Press the "Accept" button

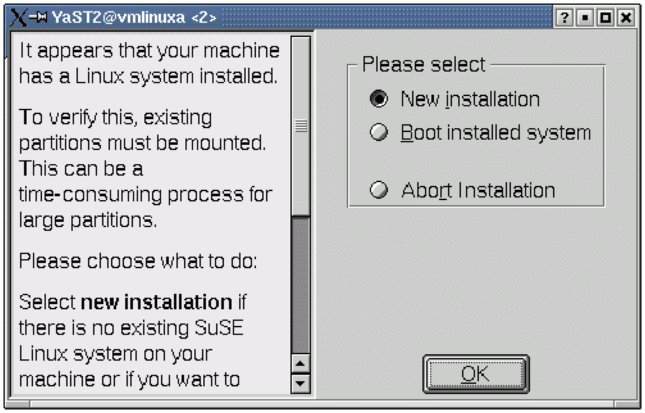
DASD Parameter: dasd=201,202

DASD Name	DASD Address	Status
dasda	0201	active
dasdb	0202	active

Figure 3-17 Load dasd module

In the next window (as shown in Figure 3-18) you have to select if you want to install a completely new system, or if existent, boot an already installed system. In our case we reused some old DASDs, so that YaST recognized that there was a system installed before. Since we want to install a new system, we select **New installation** and click **OK** to continue.

YaST now tries to automatically detect the system settings and set up the installation settings. Because the DASDs we used were not formatted and partitioned with this system before, the automatic assignment of mount points fails and is highlighted (colored red) as an error message that appears under Partitioning, as shown in Figure 3-19.



It appears that your machine has a Linux system installed.

To verify this, existing partitions must be mounted. This can be a time-consuming process for large partitions.

Please choose what to do:

Select **new installation** if there is no existing SuSE Linux system on your machine or if you want to

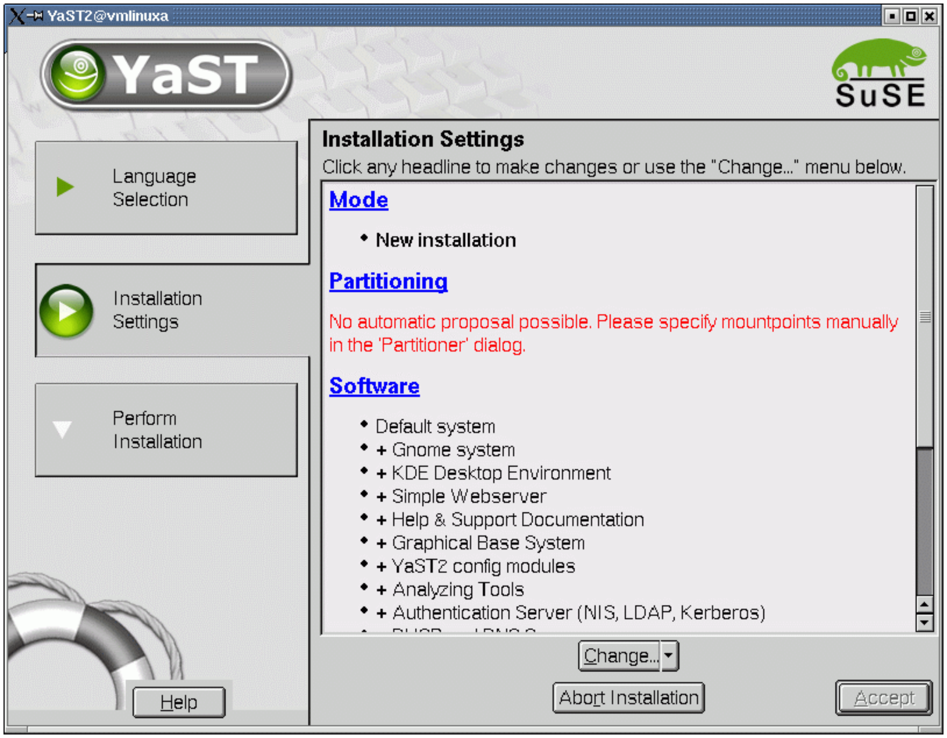
Please select

● New installation

○ Boot installed system

○ Abort Installation

Figure 3-18 Select installation



Installation Settings

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

Mode

- **New installation**

Partitioning

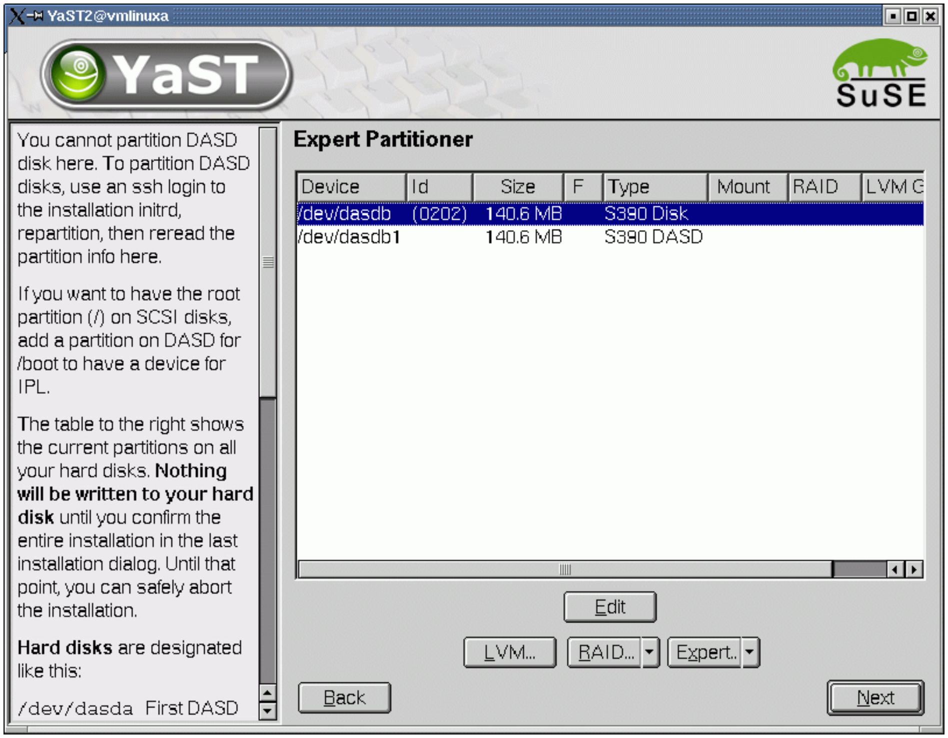
No automatic proposal possible. Please specify mountpoints manually in the 'Partitioner' dialog.

Software

- Default system
- + Gnome system
- + KDE Desktop Environment
- + Simple Webserver
- + Help & Support Documentation
- + Graphical Base System
- + YaST2 config modules
- + Analyzing Tools
- + Authentication Server (NIS, LDAP, Kerberos)

Figure 3-19 Installation settings without partitioning settings

To fix this you have to assign the volumes manually by clicking **Partitioning**. This will bring up the Expert Partitioner as shown in Figure 3-20.



You cannot partition DASD disk here. To partition DASD disks, use an ssh login to the installation initrd, repartition, then reread the partition info here.

If you want to have the root partition (/) on SCSI disks, add a partition on DASD for /boot to have a device 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

Expert Partitioner

Device	Id	Size	F	Type	Mount	RAID	LVM G
/dev/dasdb	(0202)	140.6 MB		S390 Disk			
/dev/dasdb1		140.6 MB		S390 DASD			

Figure 3-20 Expert partitioner without dasda

You will notice, that even if we loaded the DASD module with parameter 201 and 202, only the DASD 202 (which is dasdb in Linux) with one partition shows up here. This is because the DASD 201 was not formatted and partitioned yet within Linux. As described in the upper left of the window, you cannot partition DASD from this graphical dialog; instead, you have to use the command line utilities. Therefore, we open a separate ssh session to access our Linux system and perform the formatting and partitioning.

Note: Do not proceed or perform any action within the Expert Partitioner before the manual formatting and partitioning is completed. You will get error messages and disturb the setup process within YaST.

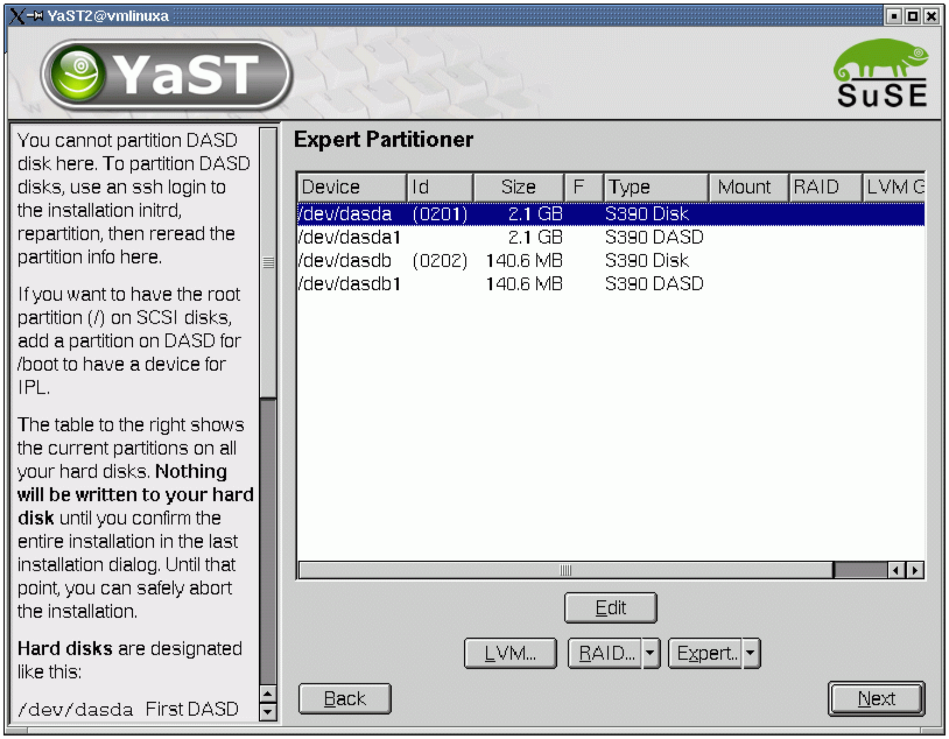
After logging into the system with a parallel ssh session, we used the utilities **`dasdfmt`** to format dasda, and **`fdasd`** to create one partition on it. The format during this stage is the low level format and must not be mistaken with the creation of a file system (e.g. ext2, ext3). Please refer to the SuSE documentation on how to use the utilities, and the different options and parameters.

Example 3-6 shows the process highlighting the input bold.

Example 3-6 Formatting dasd manually

```
SuSE Instsys vmlinuxa:/root # dasdfmt -b 4096 -d cdl -f /dev/dasda
Drive Geometry: 3138 Cylinders * 15 Heads =  47070 Tracks

I am going to format the device /dev/dasda in the following way:
   Device number of device : 0x201
   Labelling device        : yes
   Disk label              : VOL1
   Disk identifier         : 0X0201
   Extent start (trk no)   : 0
   Extent end (trk no)     : 47069
   Compatible Disk Layout  : yes
   Blocksize               : 4096

--->> ATTENTION! <<---
All data of that device will be lost.
Type "yes" to continue, no will leave the disk untouched: yes
Formatting the device. This may take a while (get yourself a coffee).

Finished formatting the device.
Rereading the partition table... ok
SuSE Instsys vmlinuxa:/root #

SuSE Instsys vmlinuxa:/root # fdasd /dev/dasda
reading volume label: VOL1
reading vtoc        : ok

Command action
   m   print this menu
   p   print the partition table
   n   add a new partition
   d   delete a partition
   v   change volume serial
   t   change partition type
   r   re-create VTOC and delete all partitions
   u   re-create VTOC re-using existing partition sizes
   s   show mapping (partition number - data set name)
   q   quit without saving changes
   w   write table to disk and exit

Command (m for help): p

Disk /dev/dasda:
    3138 cylinders,
      15 tracks per cylinder,
      12 blocks per track
    4096 bytes  per block
volume label: VOL1, volume identifier: 0X0201
maximum partition number: 3

             -----------tracks----------
               Device      start      end   length   Id  System
                               2    47069    47068       unused

Command (m for help): n
First track (1 track = 48 KByte) ([2]-47069):
Using default value 2
Last track or +size[c|k|M] (2-[47069]):
Using default value 47069

Command (m for help): w
writing VTOC...
rereading partition table...
```

After the format and partitioning we can close the ssh session and go back to the graphical Expert partitioner. Select **Expert -> Reread the partition table,** and after acknowledging a warning window you should see the newly formatted dasda in the list as shown in Figure 3-21.



You cannot partition DASD disk here. To partition DASD disks, use an ssh login to the installation initrd, repartition, then reread the partition info here.

If you want to have the root partition (/) on SCSI disks, add a partition on DASD for /boot to have a device 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

Expert Partitioner

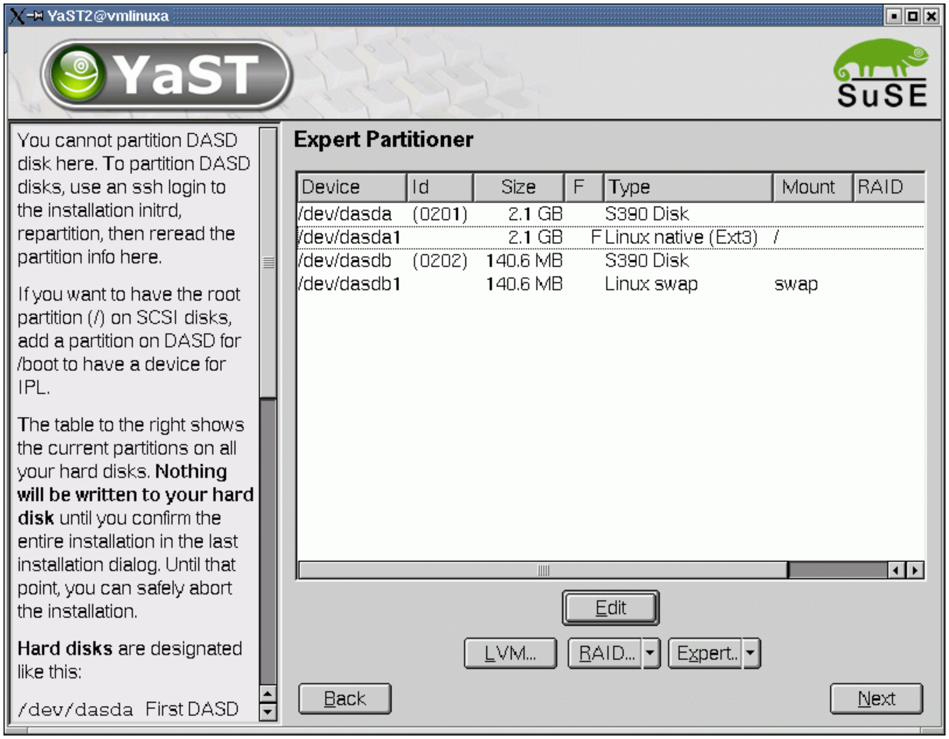
Device	Id	Size	F	Type	Mount	RAID	LVM G
/dev/dasda	(0201)	2.1 GB		S390 Disk			
/dev/dasda1		2.1 GB		S390 DASD			
/dev/dasdb	(0202)	140.6 MB		S390 Disk			
/dev/dasdb1		140.6 MB		S390 DASD			

Figure 3-21 Expert partitioner with dasda

Now we can go ahead and assign mount points and file system formats to the DASD partitions by selecting the partitions and clicking **Edit**. First we select the partition on dasda (**`/dev/dasda1`**) and select **/** as the mount point as well as the the type of file system desired for that partition. In our case we used the **`ext3`** file system (please refer to “File systems” on page 258 in Appendix A for further information on Linux file systems). Click **OK** to save the settings. Finally, we select the partition on dasdb (**`/dev/dasdb1`**) and assign swap as the mount point. This automatically omits the file system option. Click **OK** to save the settings.

> **Important:** In our example we chose to create only one root file system to simplify the installation process. For a real (production) system, we strongly recommend that you create more partitions. This is important not only because of space management, but also for security reasons. By separating for example /tmp, you prevent ordinary users from filling up your root file system. Please refer to the documentation of your distribution for specific recommendations about the partition layout.

Now we have correctly assigned our DASD space. The settings are shown in Figure 3-22. Notice the formatting flag **F** for the partition dasda1 in the F column.



You cannot partition DASD disk here. To partition DASD disks, use an ssh login to the installation initrd, repartition, then reread the partition info here.

If you want to have the root partition (/) on SCSI disks, add a partition on DASD for /boot to have a device 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

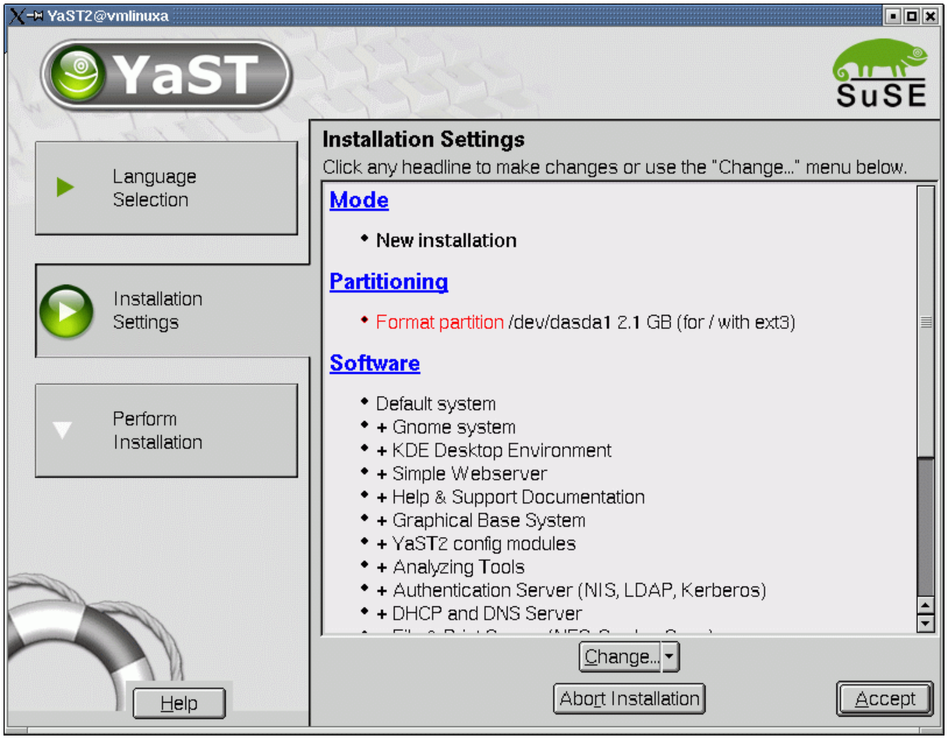
Device	Id	Size	F	Type	Mount	RAID
/dev/dasda	(0201)	2.1 GB		S390 Disk		
/dev/dasda1		2.1 GB	F	Linux native (Ext3)	/	
/dev/dasdb	(0202)	140.6 MB		S390 Disk		
/dev/dasdb1		140.6 MB		Linux swap	swap	

Figure 3-22 Expert partitioner with partition settings

To go back to the Installation Settings screen click **Next**. As you can see in Figure 3-23, we now have the desired action specified for the Partitioning section.

We can proceed with the installation by clicking **Accept**.

Note: You can make changes to the software package selection before clicking **Accept,** by clicking on **Software**. In our example we use the default package selection.



Installation Settings

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

Mode

- New installation

Partitioning

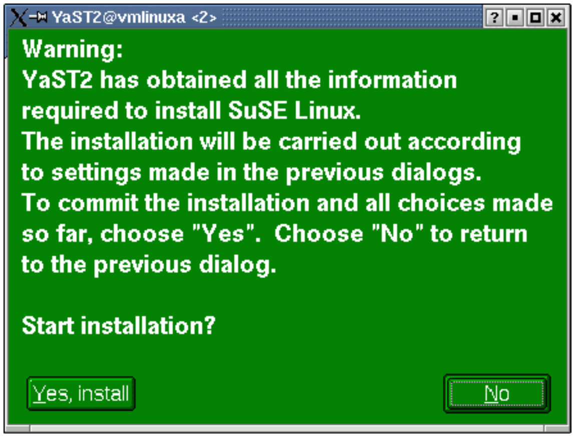
- Format partition /dev/dasda1 2.1 GB (for / with ext3)

Software

- Default system
- + Gnome system
- + KDE Desktop Environment
- + Simple Webserver
- + Help & Support Documentation
- + Graphical Base System
- + YaST2 config modules
- + Analyzing Tools
- + Authentication Server (NIS, LDAP, Kerberos)
- + DHCP and DNS Server

Figure 3-23 Installation settings with partitioning settings

For the system to proceed, you have to acknowledge the start of the installation by clicking **Yes, install,** as shown in Figure 3-24.

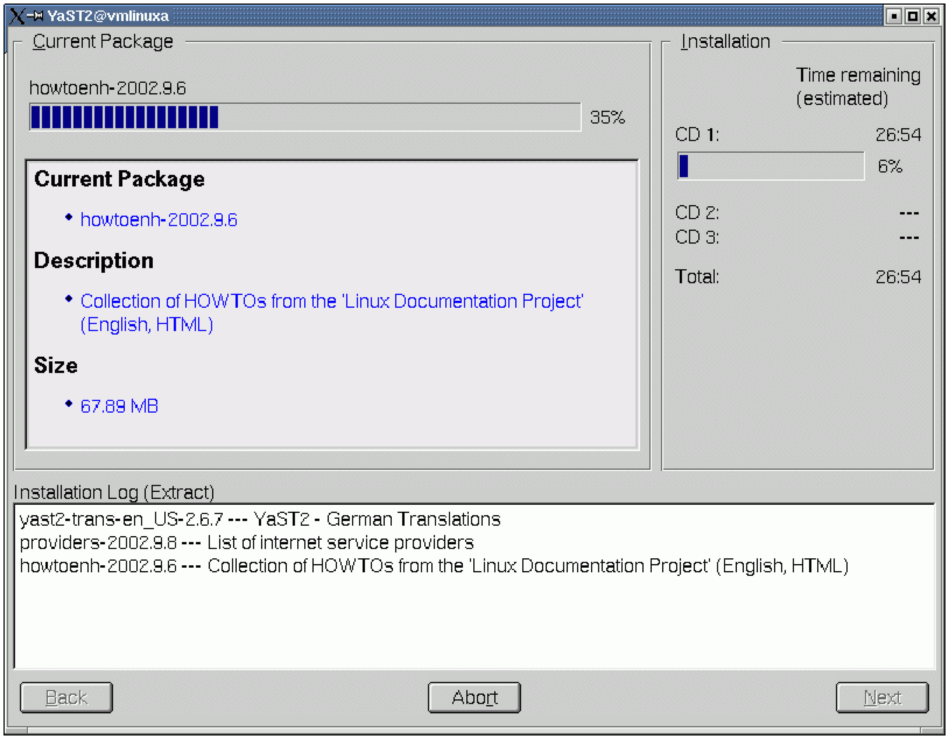


Warning:
YaST2 has obtained all the information required to install SuSE Linux.
The installation will be carried out according to settings made in the previous dialogs.
To commit the installation and all choices made so far, choose "Yes". Choose "No" to return to the previous dialog.

Start installation?

Figure 3-24 Warning before proceed with installation

YaST will now format the DASDs with the file systems you specified, followed by the installation of the software packages from the FTP server to DASD. This process takes a while, but you can watch the progress and the estimated remaining time as shown in Figure 3-25.



Current Package

- howtoenh-2002.9.6

Description

- Collection of HOWTOs from the 'Linux Documentation Project' (English, HTML)

Size

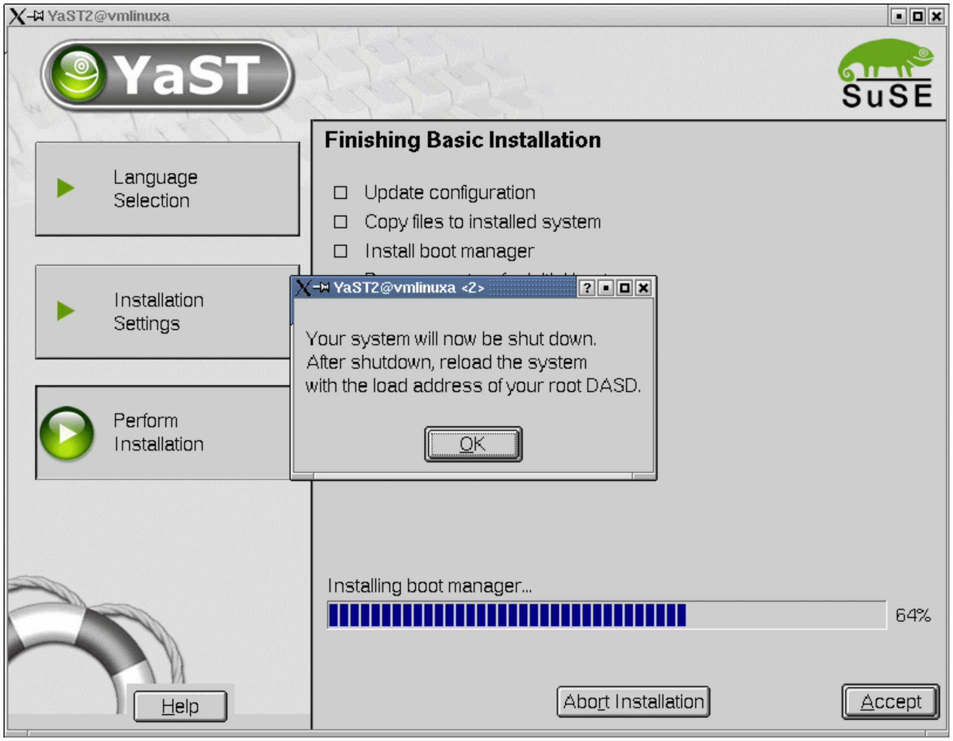
- 67.89 MB

Installation Log (Extract)

```
yast2-trans-en_US-2.6.7 --- YaST2 - German Translations
providers-2002.9.8 --- List of internet service providers
howtoenh-2002.9.6 --- Collection of HOWTOs from the 'Linux Documentation Project' (English, HTML)
```

Figure 3-25 Installation progress

When the packages installation has completed, confirm the re-IPL from DASD by clicking **OK** on the message window as shown in Figure 3-26.



Finishing Basic Installation

□ Update configuration
□ Copy files to installed system
□ Install boot manager

Your system will now be shut down.
After shutdown, reload the system
with the load address of your root DASD.

Installing boot manager...
64%

Figure 3-26 Ready to IPL from dasd for first time

> **Note:** Do not close your X Server or change any of its settings, because the YaST installation will resume after the re-IPL.

After a successful re-IPL from DASD you will see the following message on your system console.

Example 3-7 Message during first IPL from dasd

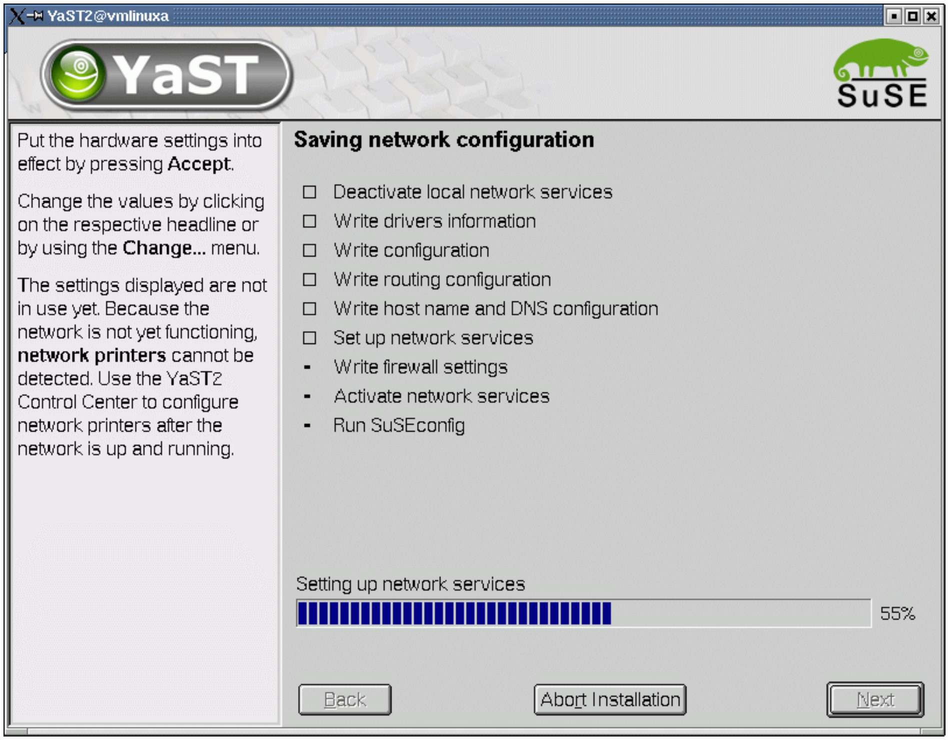
```
***
***          Please return to your X-Server screen to finish installation
***
```

Return to your X server where a dialog asking for the root password should now display. Fill in a carefully chosen password and click **Next** to continue. The next screen allows you to create a normal user account. You can fill in the necessary information and then proceed with **Next**. You can also choose to skip this step and create normal users later, by leaving the fields empty and just clicking **Next**.

At this point, YaST will run several post installation scripts, which will configure the software packages previously installed.

When the scripts have completed, you can review the network settings or add printers by clicking the appropriate links in the next screen. Since the network settings from the installation dialog are carried over automatically and we do not want to add a printer, we just continue by clicking **Next**.

Finally, YaST is saving the network settings and setting up the network services as shown in Figure 3-27.



Put the hardware settings into effect by pressing **Accept**.

Change the values by clicking on the respective headline or by using the **Change...** menu.

The settings displayed are not in use yet. Because the network is not yet functioning, **network printers** cannot be detected. Use the YaST2 Control Center to configure network printers after the network is up and running.

Saving network configuration

- □ Deactivate local network services
- □ Write drivers information
- □ Write configuration
- □ Write routing configuration
- □ Write host name and DNS configuration
- □ Set up network services
- Write firewall settings
- Activate network services
- Run SuSEconfig

Setting up network services

55%

Figure 3-27 Saving the settings

The installation of the SuSE Linux Enterprise Server is complete. The system is a up and running and will allow authorized users to log in.

The next section describes how to add Fibre Channel attached storage to the system.

3.5 zSeries and disk storage under Linux

In this section we focus on the attachment of disk storage via Fibre Channel as stated in the introduction.

> **Important:** A large number of SCSI devices exist in the marketplace. IBM cannot fully test all of these for use in zSeries systems. There will be a distinction between IBM supported devices and devices that have been used and appear to work but are not formally supported for zSeries FCP by IBM. For a complete list of supported devices, consult:
> `http://www-1.ibm.com/servers/eserver/zseries/connectivity/#fcp`

3.5.1 ESS using FCP

This section refers to the physical setup presented in 3.3, “Configuration used for this redbook” on page 38. Here we discuss the following:

- Gather prerequisite information about the SAN setup
- Set up the ESS and configure LUNs for use with Linux for zSeries
- Manually load the required modules and parameters to access the disks
- Make the disks permanently accessible

Prerequisite information

From a software configuration standpoint, you need to collect the following elements of information, as described in 3.1.2 “Distributed storage attachments” on page 29, in order to prepare a Linux system for accessing the Enterprise Storage Server through Fibre Channel:

- Hostname of the server hosting the Linux system
- Device address (and CHPID) of the FCP port attached to Linux
- World Wide Port Name (WWPN) of the FCP port on the zSeries
- FC port on the ESS
- World Wide Port Name of the FC port on the ESS

You can obtain the data from the HMC, the ESS Specialist, and the McData switch; in our case, these were:

Table 3-2 World Wide Port Names

Linux hostname	vmlinuxa
CHPID of the FCP port on the z800	15
WWPN of the FCP port on the z800	50:05:07:64:01:40:01:7d
FC port on the ESS	Bay 3 Slot 1

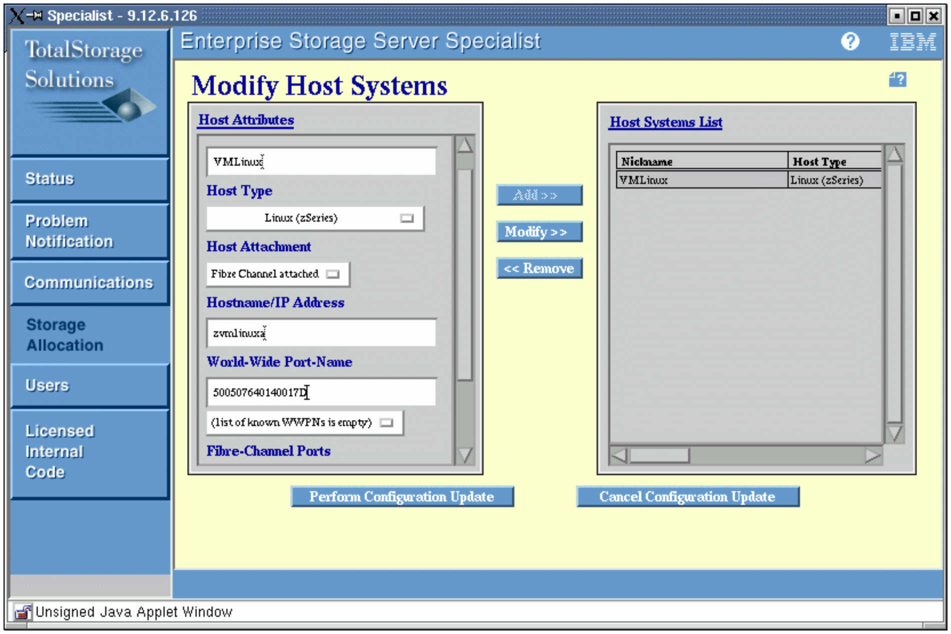
Linux hostname	vmlinuxa
WWPN of the FC port on the ESS	50:05:07:63:00:c8:95:89

In order to gather the necessary data for your environment, please refer to the documentation of your SAN switch management software, and to the Redpaper *Getting Started with zSeries Fibre Channel Protocol*, REDP0205.

Set up the ESS

To set up LUNs for use with your Linux system, you first have to define the system to the ESS. This is done through the ESS Specialist; in this section we assume that you already are familiar with the ESS Specialist. We enumerated all the actions, but did not include a picture for all the screens. You can find more information about the ESS Specialist in Chapter 7. “Configuring ESS for Linux” on page 155.

Log on to the ESS Specialist and select **Storage Allocation -> Open System Storage -> Modify Host Systems** to display the host setup screen. Enter a nickname for the host, select the host type from the pull-down list, fill in the Linux hostname and the WWPN of the FCP port on the z800, and select the **FC port** on the ESS. The complete settings are shown in Figure 3-28.



Modify Host Systems

Host Attributes

VMLinux

Host Type: Linux (zSeries)

Host Attachment: Fibre Channel attached

Hostname/IP Address: zvmlinuxa

World-Wide Port-Name: 500507640140017D

(list of known WWPNs is empty)

Fibre-Channel Ports

Host Systems List

Nickname	Host Type
VMLinux	Linux (zSeries)

Figure 3-28 Add the Linux host

Click **Perform Configuration Update** and the settings you just provided are applied once you confirm.

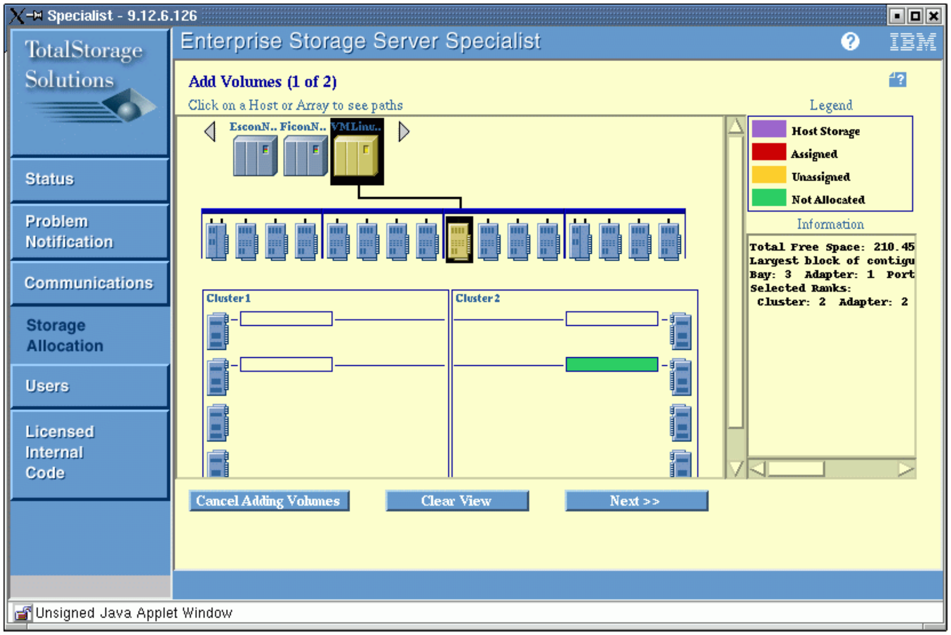
Go back to the Storage Allocation view, and you now see the Linux system shown under the nickname you chose.

Click **View All Storage** to see unassigned storage on the ESS, potentially available for your Linux system.

> **Note:** In our example, the unassigned loop was already formatted for Open System Storage (Fixed Block) and a RAID 5 array was defined on that loop.

Go to **Open System Storage** and select your **Linux host system** from the list.

Click **Add Volumes** to create LUNs as shown in Figure 3-29.



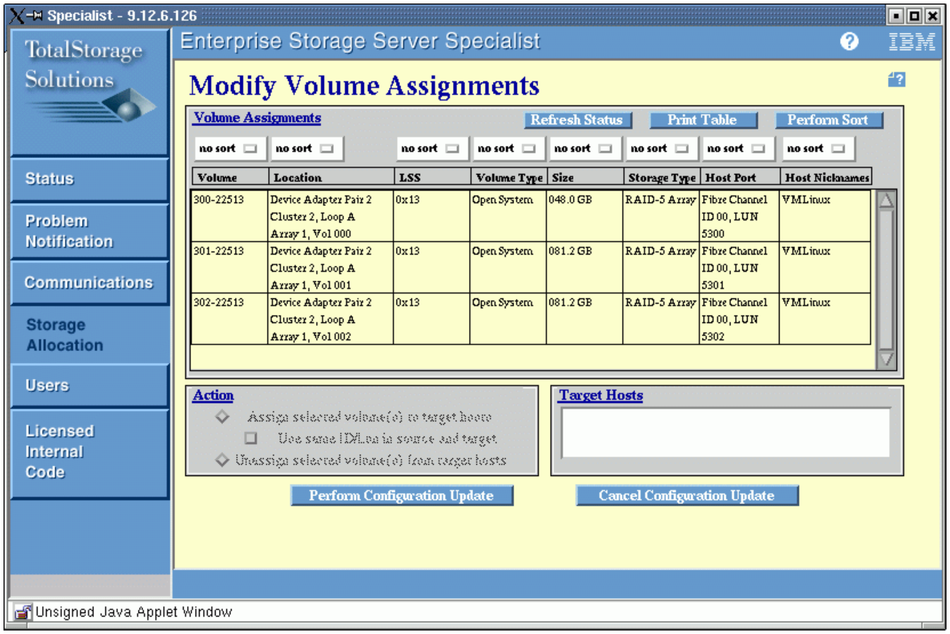
```
Total Free Space: 210.45
Largest block of contigu
Bay: 3  Adapter: 1  Port
Selected Ranks:
 Cluster: 2  Adapter: 2
```

Figure 3-29 Add volumes

Again, select your **Linux host** and the **FC port** to use. This also highlights the connection between them; proceed by clicking **Next**. A second screen appears where you can choose from the available arrays and add volumes of different sizes. To add a volume, select the array and the desired size of the volume, then click **Add** (in our example, we created one 48 GB volume and two 81.2 GB volumes on the available RAID 5 array).

Click **Perform Configuration Update** and after confirmation (separate message window), the volumes are created. A message window informs you that the actual format of the volumes is being performed in the background. You have to wait until the formatting is complete before accessing the new volumes. Select the **Linux host** on the Open System Storage screen to view the newly created volumes.

To view the LUNs of the volumes, go to **Open System Storage** and click **Modify Volume Assignments**. In the list shown in Figure 3-30 you can see the LUNs of the volumes assigned to your Linux system (in the Host Port column).


Specialist - 9.12.6.126
TotalStorage Solutions
Status
Problem Notification
Communications
Storage Allocation
Users
Licensed Internal Code
Enterprise Storage Server Specialist
IBM
Modify Volume Assignments
Volume Assignments
Refresh Status
Print Table
Perform Sort
no sort
Action
Assign selected volume(s) to target hosts
Use same ID/Lun in source and target
Unassign selected volume(s) from target hosts
Target Hosts
Perform Configuration Update
Cancel Configuration Update
Unsigned Java Applet Window

Volume	Location	LSS	Volume Type	Size	Storage Type	Host Port	Host Nicknames
300-22513	Device Adapter Pair 2 Cluster 2, Loop A Array 1, Vol 000	0x13	Open System	048.0 GB	RAID-5 Array	Fibre Channel ID 00, LUN 5300	VMLinux
301-22513	Device Adapter Pair 2 Cluster 2, Loop A Array 1, Vol 001	0x13	Open System	081.2 GB	RAID-5 Array	Fibre Channel ID 00, LUN 5301	VMLinux
302-22513	Device Adapter Pair 2 Cluster 2, Loop A Array 1, Vol 002	0x13	Open System	081.2 GB	RAID-5 Array	Fibre Channel ID 00, LUN 5302	VMLinux

Figure 3-30 View volume assignments to display LUNs

The LUNs for our Linux system are 5300, 5301, and 5302.

Set up the Linux system

You must have the following modules installed to use the FCP device with the ESS:

- qdio - The same module is used for other qdio devices
- scsi_mod - SCSI core
- zfcp - Provides FCP support for zSeries Linux
- sd_mod - SCSI disk support

The modules should be loaded in the order listed. All modules can be loaded with **modprobe**. Further SCSI modules exist for other SCSI devices such as `st` for tape support, `sr_mod` for CD/DVD (read only) support. They have been tested to work with different devices, but will not be discussed in this book. Please refer to the Redpaper *Getting Started with zSeries Fibre Channel Protocol*, REDP-0205.

All modules, except zfcp, can be loaded without parameters. For the zfcp module, parameters for the mapping of the FCP devices on the ESS are required. The parameters needed for each device are:

- The device number of the FCP port on the zSeries
- The SCSI ID starting at 1
- WWPN of the ESS FC port
- The SCSI LUN within Linux starting at 0
- The FCP LUN of the target volume on the ESS

The format of supplying the mapping parameters to the module is as follows:

```
map=”devno SCSI_ID:WWPN SCSI_LUN:FCP_LUN” such as:
map="0x0600 1:0x5005076300c89589 0:0x5301000000000000"
```

To add more then one device to the map, simply add more statements to the parameter line, separated by semicolons. The next example shows the load of the modules (omitting the qdio module, because it is loaded and used already by the OSA Express adapter in our installation) for one volume.

Example 3-8 Load scsi and zfcp modules

```
vmlinuxa:~ # modprobe scsi_mod
vmlinuxa:~ # modprobe zfcp map="0x0600 1:0x5005076300c89589 \
0:0x5301000000000000"

vmlinuxa:~ # cat /proc/scsi/scsi
Attached devices:
Host: scsi0 Channel: 00 Id: 01 Lun: 00
  Vendor: IBM      Model: 2105800          Rev: .101
  Type:   Direct-Access                    ANSI SCSI revision: 03
vmlinuxa:~ # modprobe sd_mod
vmlinuxa:~ # cat /proc/partitions
major minor  #blocks  name     rio rmerge rsect ruse wio wmerge wsect wuse running use aveq

   8     0   79296896 sda 2 6 16 10 0 0 0 0 0 10 10
  94     0    2259360 dasda 24799 32810 460872 116390 34395 47194 654280 738410 0 175040 854800
  94     1    2259264 dasda1 24778 32810 460704 116350 34395 47194 654280 738410 0 175000 854760
  94     4     144000 dasdb 53 105 1264 80 0 0 0 0 0 80 80
  94     5    2403264 dasdb1 32 105 1096 40 0 0 0 0 0 40 40
```

The example shows that a new SCSI volume, sda, is accessible by our system. However, before you can use a new volume, it must be partitioned using **fdisk**

After partitioning the new volume you can create file systems on the partitions either using the command line utilities or using YaST. In our case, we assigned the mount point, formatted the partition with the ext3 file system, and applied the changes. Now the new device is ready for use.

Please refer to Appendix A for details on how to partition disks and create file systems.

Add more devices and make changes permanent

To add more than one device to your SCSI configuration, it is more convenient to write a small script with all the parameters included such as shown in Example 3-9.

Example 3-9 Module load script

```
vmlinuxa:~ # cat scsi.sh
modprobe scsi_mod
modprobe zfcp map="\
0x0600 1:0x5005076300c89589 0:0x5301000000000000;\
0x0600 1:0x5005076300c89589 1:0x5302000000000000;\
0x0600 1:0x5005076300c89589 2:0x5300000000000000"
modprobe sd_mod
```

Make sure that the modules are unloaded before executing the script.

Example 3-10 Module load via script

```
vmlinuxa:~ # sh scsi.sh
vmlinuxa:~ # cat /proc/scsi/scsi
Attached devices:
Host: scsi0 Channel: 00 Id: 01 Lun: 00
  Vendor: IBM      Model: 2105800          Rev: .101
  Type:   Direct-Access                    ANSI SCSI revision: 03
Host: scsi0 Channel: 00 Id: 01 Lun: 01
  Vendor: IBM      Model: 2105800          Rev: .101
  Type:   Direct-Access                    ANSI SCSI revision: 03
Host: scsi0 Channel: 00 Id: 01 Lun: 02
  Vendor: IBM      Model: 2105800          Rev: .101
  Type:   Direct-Access                    ANSI SCSI revision: 03
```

Alternatively, you can also add SCSI devices to an existing configuration via **add_map**. Then you have to make the devices known to the SCSI stack manually as shown in Example 3-11.

Example 3-11 Add devices manually via add_map

```
vmlinuxa:~ # cat /proc/scsi/scsi
Attached devices:
```

```
Host: scsi0 Channel: 00 Id: 01 Lun: 00
  Vendor: IBM      Model: 2105800          Rev: .101
  Type:   Direct-Access                    ANSI SCSI revision: 03
vmlinuxa:~ # echo "0x0600 0x00000001:0x5005076300c89589 \
0x00000001:0x5302000000000000" > /proc/scsi/zfcp/add_map
vmlinuxa:~ # echo "scsi add-single-device 0 0 1 1" > /proc/scsi/scsi
vmlinuxa:~ # cat /proc/scsi/scsi
Attached devices:
Host: scsi0 Channel: 00 Id: 01 Lun: 00
  Vendor: IBM      Model: 2105800          Rev: .101
  Type:   Direct-Access                    ANSI SCSI revision: 03
Host: scsi0 Channel: 00 Id: 01 Lun: 01
  Vendor: IBM      Model: 2105800          Rev: .101
  Type:   Direct-Access                    ANSI SCSI revision: 03
```

To make the devices available permanently (also after a reboot), you have to create a new initial ramdisk containing the necessary modules and parameter information. First, save the module parameters in the configuration file /etc/zfcp.conf.

Example 3-12 Configuration file for zfcp

```
vmlinuxa:~ # cat /proc/scsi/zfcp/map > /etc/zfcp.conf
vmlinuxa:~ # cat /etc/zfcp.conf
0x0600 0x00000001:0x5005076300c89589 0x00000000:0x5301000000000000
0x0600 0x00000001:0x5005076300c89589 0x00000001:0x5302000000000000
0x0600 0x00000001:0x5005076300c89589 0x00000002:0x5300000000000000
```

Next, create a new ramdisk which is done with the utility **mk_initrd** and then run **zipl** to update the IPL record to point to the new ramdisk. This is shown in Example 3-13.

Example 3-13 Create new initial ramdisk and run zipl

```
vmlinuxa:~ # mk_initrd
using "/dev/dasda1" as root device (mounted on "/" as "ext3")

Found ECKD dasd, adding dasd eckd discipline!

Note: If you want to add ECKD dasd support for later mkinitrd
calls where possibly no ECKD dasd is found, add dasd_eckd_mod
to INITRD_MODULES in /etc/sysconfig/kernel

creating initrd "/boot/initrd" for kernel "/boot/kernel/image"
(version 2.4.19-3suse-SMP) (s390)

 - insmod scsi_mod            (kernel/drivers/scsi/scsi_mod.o)
 - insmod qdio                (kernel/drivers/s390/qdio.o)
 - insmod zfcp map="
0x0600 0x00000001:0x5005076300c89589 0x00000000:0x5301000000000000
0x0600 0x00000001:0x5005076300c89589 0x00000001:0x5302000000000000
0x0600 0x00000001:0x5005076300c89589 0x00000002:0x5300000000000000"
(kernel/drivers/s390/scsi/zfcp.o)
 - insmod jbd                 (kernel/fs/jbd/jbd.o)
 - insmod ext3                (kernel/fs/ext3/ext3.o)
 - insmod dasd_mod dasd=$dasd (kernel/drivers/s390/block/dasd_mod.o)
 - insmod dasd_eckd_mod       (kernel/drivers/s390/block/dasd_eckd_mod.o)
 - insmod sd_mod              (kernel/drivers/scsi/sd_mod.o)
 - zfcp support

Run zipl now to update the IPL record!

vmlinuxa:~ # zipl
building bootmap    : /boot/zipl/bootmap
adding Kernel Image : /boot/kernel/image located at 0x00010000
adding Ramdisk      : /boot/initrd located at 0x00800000
adding Parmline     : /boot/zipl/parmfile located at 0x00001000
Bootloader for ECKD type devices with z/OS compatible layout installed.
Syncing disks....
...done
vmlinuxa:~ #
```

Now, all the necessary parameters and modules are accessible at boot time. The next time you re-IPL the system the scsi and zfcp modules are loaded automatically, and the file systems are mounted as defined in /etc/fstab.

3.5.2 Multipath support

Some storage devices allow operating systems to access device storage through more than one physical path, thus extending availability, and sometimes allowing better performance by permitting use of all paths in parallel. To make use of these features, an operating system must be able to detect multipath devices and manage all access in a way that avoids data corruption.

Currently, Linux for zSeries and S/390 supports only the Enterprise Storage Server (ESS) through extensions to the Linux Logical Volume Manager (LVM). To make use of the LVM extensions, any device driver must provide one device node for every path that can be accessed independently (the device driver handles the path switching if the device architecture needs anything special for that purpose).

Block-device managers (such as LVM, MD, or EVMS) take block devices and provide new logical ones after some transformation. This just requires the lower device drivers to provide a separate device for every path so that the manager knows about the paths.

Please refer to the *Device Drivers and Installation Commands* documentation on how to implement multipathing for Linux on zSeries.

4

SuSE Linux on pSeries

This chapter starts with a brief overview of IBM eServer pSeries and its Linux support.

It continues with a focus on the requirements and installation steps pertaining to SuSE Linux Enterprise Server 8 in two ways: first, a native installation on a pSeries server, then installation in an LPAR (available on the p670 and p690 models).

The chapter ends with a review of the tasks involved in the configuration of an Emulex Host Bus Adapter, for attachment to FAStT and ESS.

4.1 Introduction to pSeries and Linux

The pSeries is one of IBM's key solutions to fit a variety of customers needs. The pSeries family presents a full range of servers from entry deskside models to high-end enterprise models with LPAR capabilities.

Linux for pSeries is a key element of the IBM Linux strategy. IBM is working closely with the Linux community to increase performance, scalability, reliability, and serviceability to match the strengths of pSeries servers (see Figure 4-1).

To make it easy to get started with Linux for pSeries, IBM has introduced a number of pSeries Linux ready Express Configurations. These systems represent some of the most popular configurations of the p630 and p650 systems. They are provided without an AIX® license and offer great savings with the ability to add additional features.

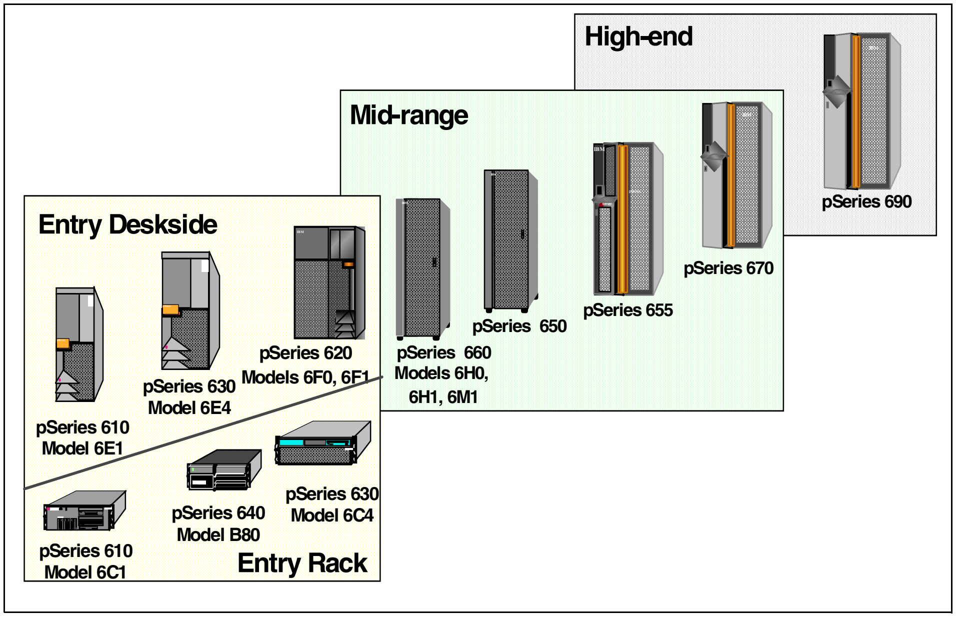


Figure 4-1 pSeries models supporting Linux

Linux for pSeries is especially compelling for solutions requiring a 64-bit architecture or the high-performance floating-point capabilities of the POWER processor.

In addition, the logical partitioning LPAR capabilities of the pSeries make it possible to run one or more instances of Linux along with AIX (i.e. zero or more Linux partitions along with zero or more AIX partitions). This enables the consolidation of workloads from several separate servers onto a single system. Since the partitioning is controlled by the hypervisor firmware and the Hardware Management Console, AIX is never required to run Linux. LPAR capabilities also provide a low-risk way to begin developing and deploying Linux operating system-ready applications as desired while retaining the enterprise-ready capabilities of AIX for mission-critical or highly-scalable workloads. Since Linux does not currently scale to efficiently handle large SMP systems, LPAR also allows large pSeries systems to be partitioned to run Linux workloads.

It should be noted also that with AIX 5L™, IBM introduced the AIX Toolbox for Linux applications. This set of tools allows you to recompile Linux source code under AIX, offering the possibility to run applications developed on Linux while retaining the more advanced features of AIX.

For an overview of what pSeries models are currently supported by what Linux distributions, see:

http://www.ibm.com/servers/eserver/pseries/hardware/linux_facts.pdf

In this redbook, our focus is on SuSE SLES8 / UnitedLinux 1.0.

Note: For the latest information on pSeries and Linux in general, refer to:

http://www.ibm.com/eserver/pseries/linux

4.2 Requirements

Attachment to storage systems shown in this book was still under test at the time of writing. If you plan to use scenarios from this book, please verify first that these configurations have now been certified by IBM.

4.2.1 Hardware requirements

In general, the SuSE Linux Enterprise Server runs on pSeries systems natively. You can find additional information about the supported hardware and technical issues by consulting the following Internet pages:

http://www-i.ibm.com/servers/eserver/pseries/linux

http://www.suse.com/us/business/products/server/sles/i_pseries.html

4.2.2 Software requirements

Take note: At the time of this writing, there was no multipath driver for the Emulex Fibre Channel adapters, nor SDD for ESS available yet. This may have changed since publishing this redbook. Please check the appropriate Internet pages.

The following software components are necessary for installation:

- SLES 8 (UnitedLinux 1.0) for pSeries
- Host adapter device drivers package (provided by SuSE, included in the Service Pack 1 (SP1®) for SLES 8)
- Depending on the IBM TotalStorage device chosen, you require for:

  a. FAStT:

     - Open Build Driver for the Emulex Adapter family (when available)
     - any system able to run the FAStT Storage Manager Client

  b. Enterprise Storage Server ESS:

     - SDD (when announced)

4.2.3 Connection requirements

For the initial setup you will require a network infrastructure to configure the storage components.

Additionally, you need either a console to do a native installation, or access to the HMC when doing an LPAR installation. To facilitate the installation tasks, we suggest that you set up a desktop system with a Telnet client and a VNC client (available on the UnitedLinux CD #1 in the *dosutils* folder). However, the VNC server can be accessed via any Web browser on port 5801.

When setting up your hardware, consider all aspects of required SAN connections, including the length of your cables.

4.3 Configuration used for this redbook

For the configurations illustrated in this redbook, we used the following hardware:

- IBM RS/6000® 270
- IBM pSeries p610 -> config
- IBM pSeries p690

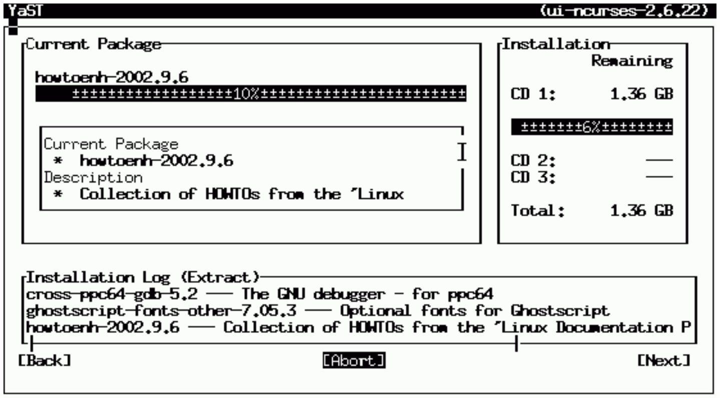
Additionally, we used the following IBM TotalStorage components:

- FAStT200
- FAStT700
- ESS 800
- IBM Fibre Channel switch 2109-F16

4.4 Implementing SuSE Linux Enterprise Server 8

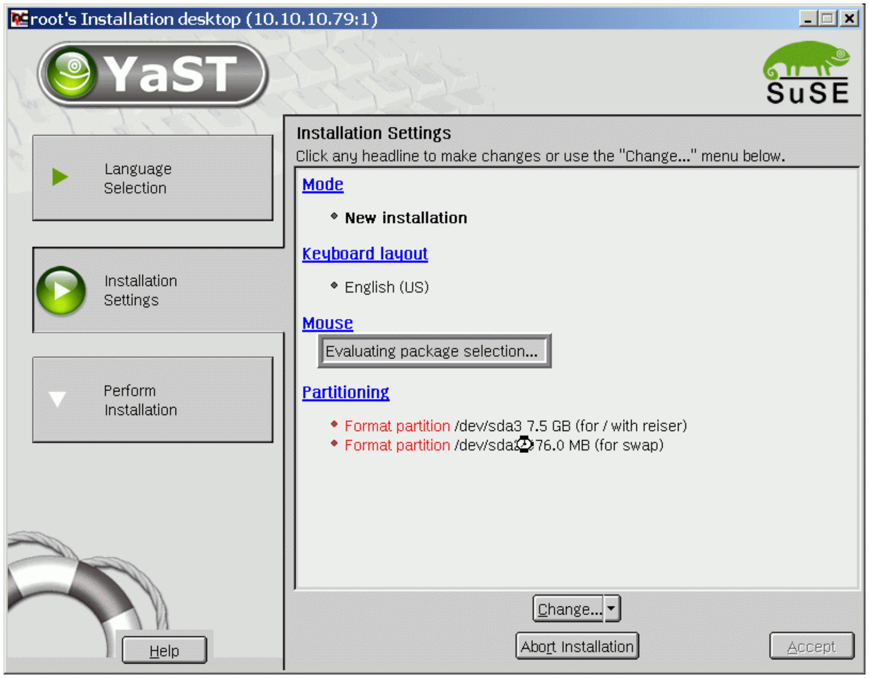
In principle there are four ways to perform the installation:

- Local, full graphical install using the system video adapter
- Serial installation using a terminal console connected to the first serial port of the pSeries system



```
YaST                                                  (ui-ncurses-2.6.22)

 Current Package                                    Installation
                                                          Remaining
  howtoenh-2002.9.6
  ++++++++++++++++++10%++++++++++++++++++++++        CD 1:      1.36 GB

   Current Package                                   +++++++6%++++++++
    *  howtoenh-2002.9.6
   Description                                       CD 2:          --
    *  Collection of HOWTOs from the 'Linux          CD 3:          --

                                                     Total:     1.36 GB

 Installation Log (Extract)
 cross-ppc64-gdb-5.2 -- The GNU debugger - for ppc64
 ghostscript-fonts-other-7.05.3 -- Optional fonts for Ghostscript
 howtoenh-2002.9.6 -- Collection of HOWTOs from the 'Linux Documentation P

[Back]                         [Abort]                              [Next]
```

Figure 4-2 Installation using serial console



Installation Settings

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

Mode

- **New installation**

Keyboard layout

- English (US)

Mouse

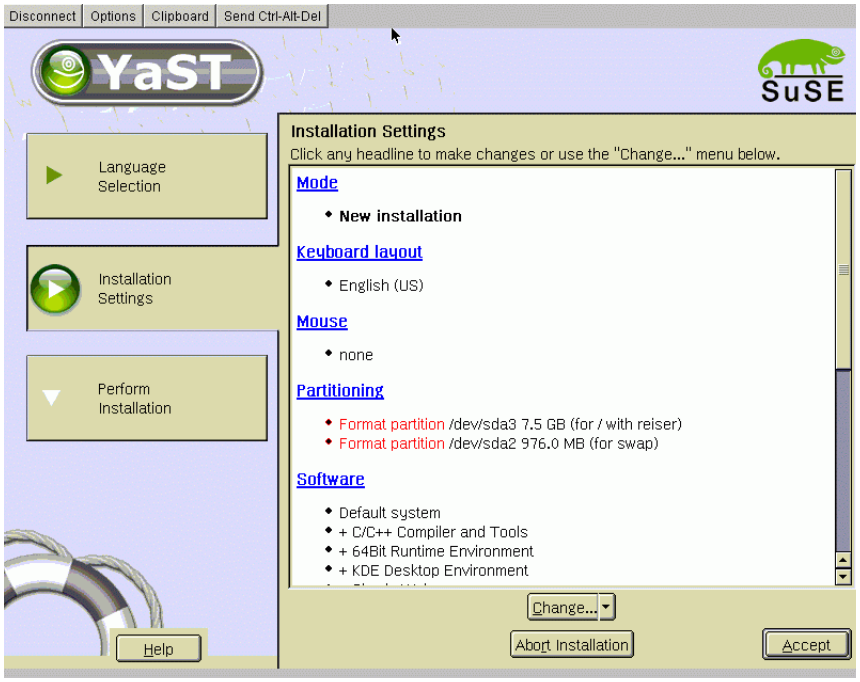
Evaluating package selection...

Partitioning

- Format partition /dev/sda3 7.5 GB (for / with reiser)
- Format partition /dev/sda2 76.0 MB (for swap)

Figure 4-3 Installation using VNC client

- Remote installation using the serial console for hardware settings, and a VNC client for the graphical installation
- Remote installation using the serial console for hardware settings and a Web browser supporting JAVA as VNC client for the graphical installation



Installation Settings

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

Mode

- New installation

Keyboard layout

- English (US)

Mouse

- none

Partitioning

- Format partition /dev/sda3 7.5 GB (for / with reiser)
- Format partition /dev/sda2 976.0 MB (for swap)

Software

- Default system
- + C/C++ Compiler and Tools
- + 64Bit Runtime Environment
- + KDE Desktop Environment

Figure 4-4 Installation using Web browser acting as VNC client by JAVA

The result of the installation is identical whatever mode of installation you use. Simply select the one that is the most convenient, based on your environment.

4.4.1 Installation steps

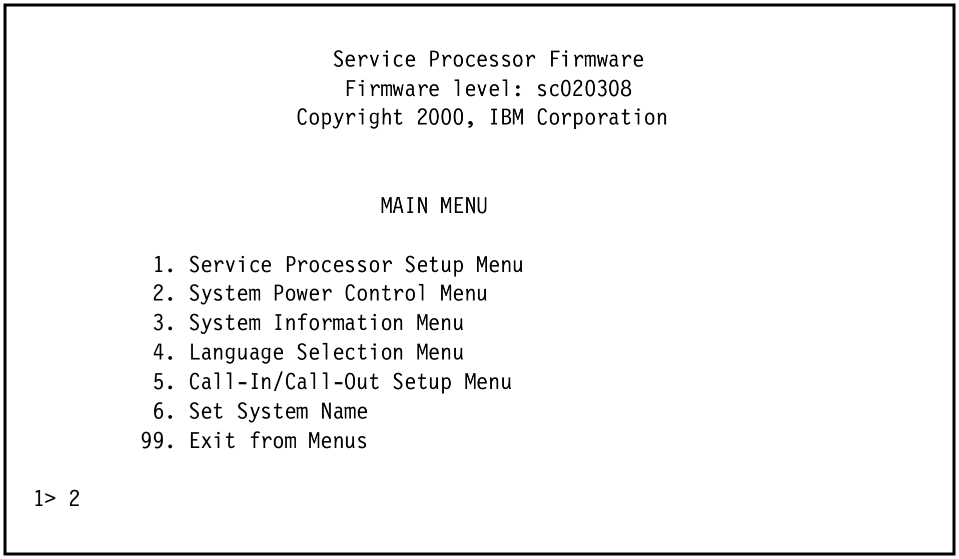
For our experiment, we used the remote installation with a serial console for hardware settings, and a VNC client for the graphical installation.

We set up serial console with the following parameter: VT100, 9600 Baud, 8/1/N and connect it to the Service Processor. Details on how to install a serial console using System Management Services is available from the following pSeries library Web site:

http://www-1.ibm.com/servers/eserver/pseries/library

The paragraphs that follow illustrate the case of a native Linux installation; Linux installation in an LPAR is not different, except for the initial setup of the LPAR addressed at the end of this section.

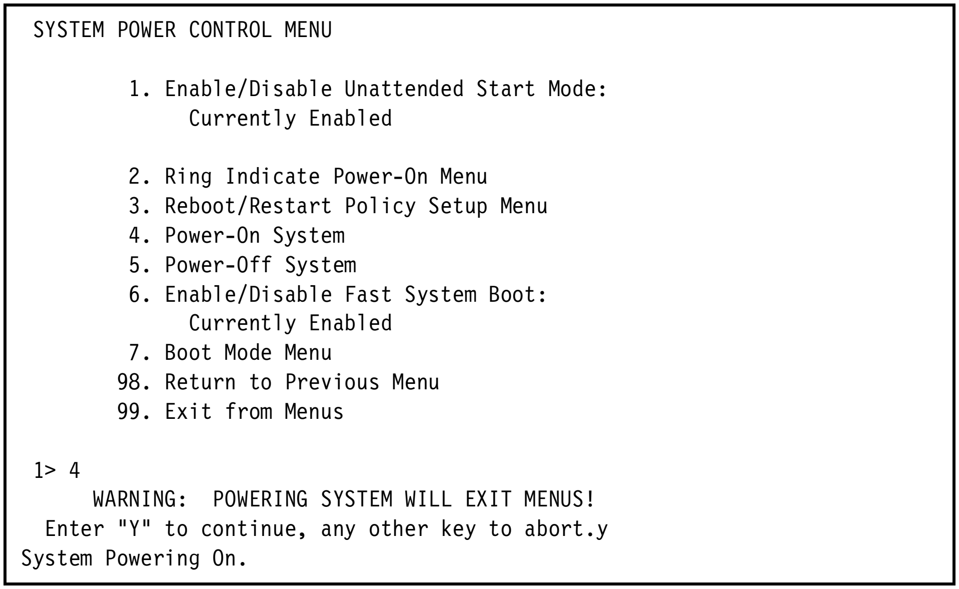
Power on your system. The serial console displays the menu shown in Figure 4-5.



```
                         Service Processor Firmware
                          Firmware level: sc020308
                       Copyright 2000, IBM Corporation


                                 MAIN MENU

           1. Service Processor Setup Menu
           2. System Power Control Menu
           3. System Information Menu
           4. Language Selection Menu
           5. Call-In/Call-Out Setup Menu
           6. Set System Name
          99. Exit from Menus

 1> 2
```

Figure 4-5 Main Menu, choose 2 to proceed

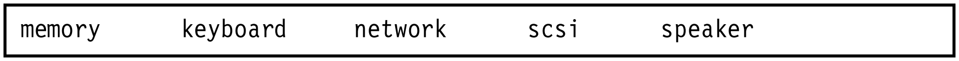
Enter 2 to access the System Power Control Menu shown in Figure 4-6.



```
 SYSTEM POWER CONTROL MENU

         1. Enable/Disable Unattended Start Mode:
              Currently Enabled

         2. Ring Indicate Power-On Menu
         3. Reboot/Restart Policy Setup Menu
         4. Power-On System
         5. Power-Off System
         6. Enable/Disable Fast System Boot:
              Currently Enabled
         7. Boot Mode Menu
        98. Return to Previous Menu
        99. Exit from Menus

 1> 4
      WARNING:  POWERING SYSTEM WILL EXIT MENUS!
  Enter "Y" to continue, any other key to abort.y
System Powering On.
```

Figure 4-6 Power Control Menu

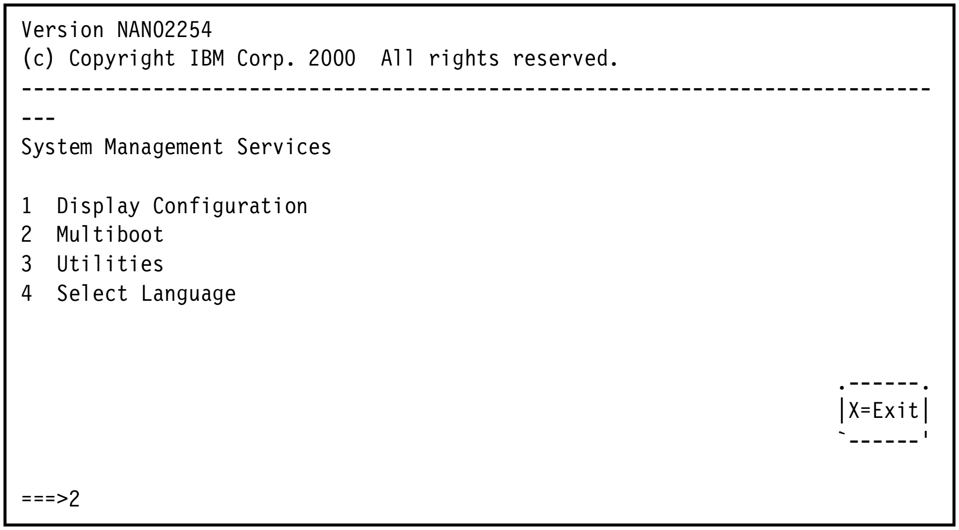
Watch your console closely for the line shown in Figure 4-7. As soon as it displays, and depending on your system architecture, press either F1 (for PPC with 64bit) or 1 (for PP with 32bit) to enter the System Management Services.



```
memory      keyboard      network      scsi      speaker
```

Figure 4-7 End of system boot, press 1/F1

In the System Management Services (Figure 4-8) select Multiboot, by entering `2`.

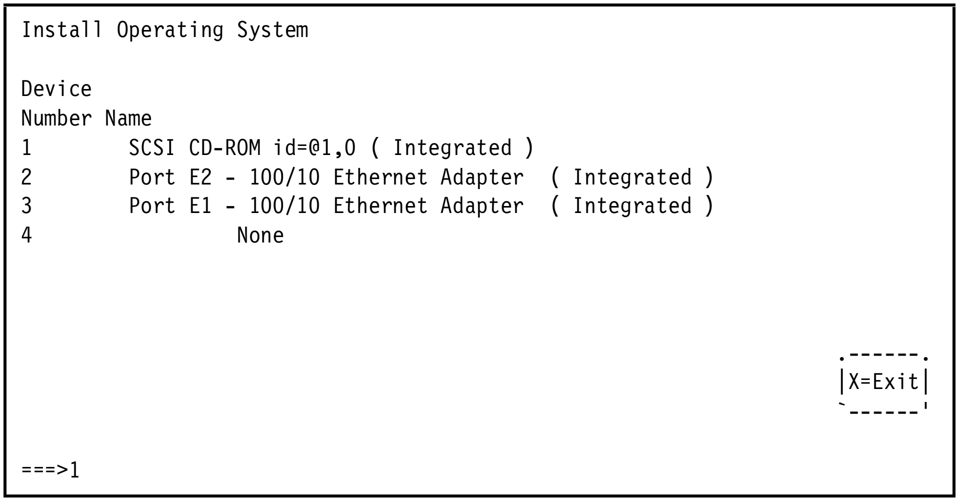


```
Version NAN02254
(c) Copyright IBM Corp. 2000  All rights reserved.
-----------------------------------------------------------------------------
---
System Management Services

1  Display Configuration
2  Multiboot
3  Utilities
4  Select Language

                                                                     .------.
                                                                     |X=Exit|
                                                                     `------'

===>2
```

Figure 4-8 Select boot device

This triggers the system to look for boot devices. This operation can last for several minutes because of time out values in effect (while the operation is in progress, the bottom corner of the screen shows the device currently being probed).

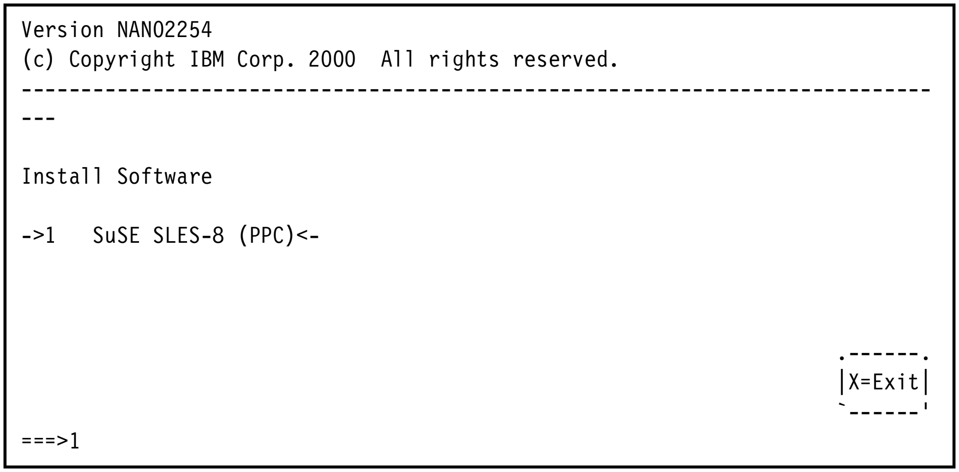
When the scan is complete, a menu similar to the one shown in Figure 4-9 is displayed. Insert your installation CD, and enter the number corresponding to the CD-ROM device.



```
Install Operating System

Device
Number Name
1       SCSI CD-ROM id=@1,0 ( Integrated )
2       Port E2 - 100/10 Ethernet Adapter  ( Integrated )
3       Port E1 - 100/10 Ethernet Adapter  ( Integrated )
4               None




                                                                 .------.
                                                                 |X=Exit|
                                                                 `------'

===>1
```

Figure 4-9 Choose CD drive for installation CD boot

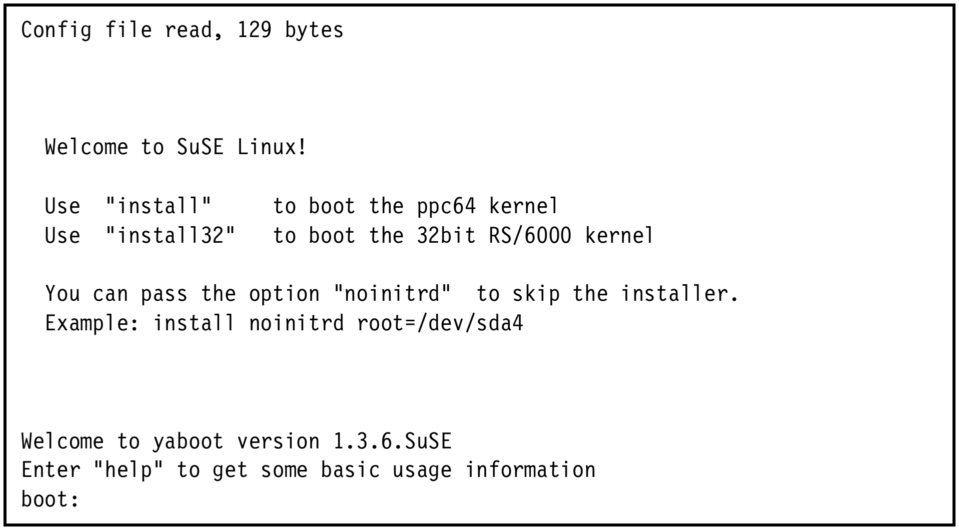
The next menu presents a choice of software to install. Select **Linux** (the ID string of the CD is displayed) by entering the corresponding number (see Figure 4-10).



```
Version NAN02254
(c) Copyright IBM Corp. 2000  All rights reserved.
-----------------------------------------------------------------------------
---

Install Software

->1   SuSE SLES-8 (PPC)<-



                                                                 .------.
                                                                 |X=Exit|
                                                                 `------'
===>1
```

Figure 4-10 Choose 1 to trigger SLES install

The system starts booting from CD. The companion piece to the boot manager LILO, called yaboot for the PPC architecture prompts you, as shown in Figure 4-11.



```
Config file read, 129 bytes



  Welcome to SuSE Linux!

  Use  "install"     to boot the ppc64 kernel
  Use  "install32"   to boot the 32bit RS/6000 kernel

  You can pass the option "noinitrd"  to skip the installer.
  Example: install noinitrd root=/dev/sda4



Welcome to yaboot version 1.3.6.SuSE
Enter "help" to get some basic usage information
boot:
```

Figure 4-11 yaboot "boot"-prompt

In the boot entry field at the bottom of the screen, enter the command corresponding to the appropriate kernel type for your platform (**`install`** for 64-bit, **`install32`** for 32-bit).

To perform the installation using the VNC Client you must specify some additional parameters. In our case we used the following command line (please adapt the parameters for your environment):

```
boot: install vnc=1 vnc_password=yourchoice hostip=10.10.10.79 netdevice=eth0
insmod=pcnet32
```

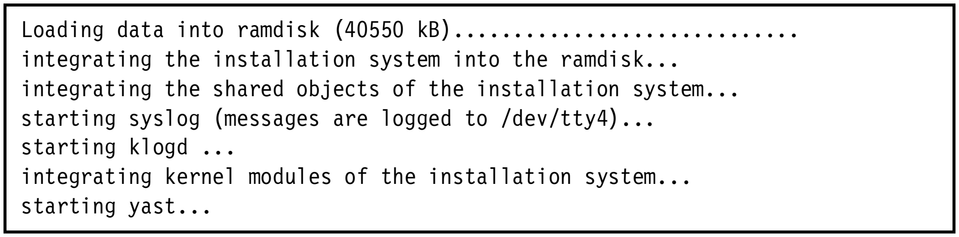
Table 4-1 lists all commands and possible parameters.

Table 4-1 Standard parameters for boot-prompt

install install32	selected kernel
vnc=1	start vnc-server
vnc_password=xyz	defines the password xyz that is mandatory for client access
dhcp=1	network configuration is provided by existing DHCP server
hostip=a.b.c.d	sets the IP address of the host you are installing to a.b.c.d

netmask=a.b.c.d	use a.b.c.d as netmask
gateway=a.b.c.d	use a.b.c.d as gateway
netdevice=eth0	use network device eth0
insmod=pcnet32	install module pcnet32 automatically

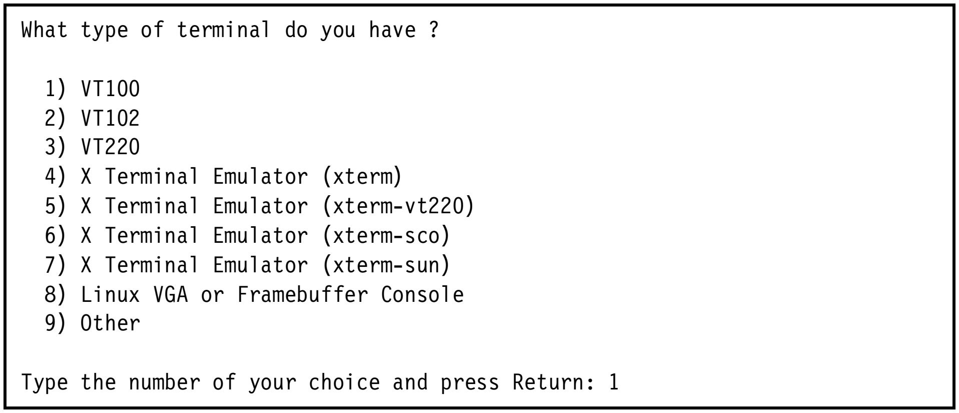
As a result of the command, yaboot loads the installation system from a CD to a ramdisk.



```
Loading data into ramdisk (40550 kB)............................
integrating the installation system into the ramdisk...
integrating the shared objects of the installation system...
starting syslog (messages are logged to /dev/tty4)...
starting klogd ...
integrating kernel modules of the installation system...
starting yast...
```

Figure 4-12 Loading installation system from CD

Next, the system prompts you for the type of terminal (Figure 4-13). In our case it was VT100.



```
What type of terminal do you have ?

  1) VT100
  2) VT102
  3) VT220
  4) X Terminal Emulator (xterm)
  5) X Terminal Emulator (xterm-vt220)
  6) X Terminal Emulator (xterm-sco)
  7) X Terminal Emulator (xterm-sun)
  8) Linux VGA or Framebuffer Console
  9) Other

Type the number of your choice and press Return: 1
```

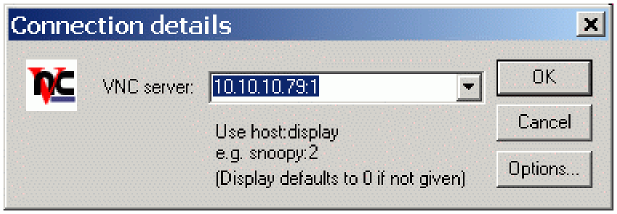
Figure 4-13 Type of terminal selection

Now, the actual SuSE Linux installation program, YaST2, starts. A VNC server also starts in our case, because we specified the parameter “vnc=1” in the boot prompt (Figure 4-11 on page 83). The VNC server displays the information you need to access your host, either through a Web browser or using a VNC client (see Figure 4-14 on page 85).

```
Please wait while YaST2 will be started

OK
starting VNC server...
a log can be found in /tmp/vncserver.log ...

***
***           You can connect to 10.10.10.78, display :1 now with vncviewer
***           Or use a Java capable browser on  http://10.10.10.78:5801/
***

(When YaST2 is finished, close your VNC viewer and return to this window.)
```

Figure 4-14 Start of VNC server

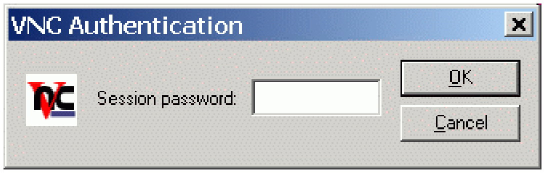
Switch to the VNC client machine, and launch the VNC client application after entering the values you gathered from the VNC server screen.



VNC server: 10.10.10.79:1

Use host:display
e.g. snoopy:2
(Display defaults to 0 if not given)

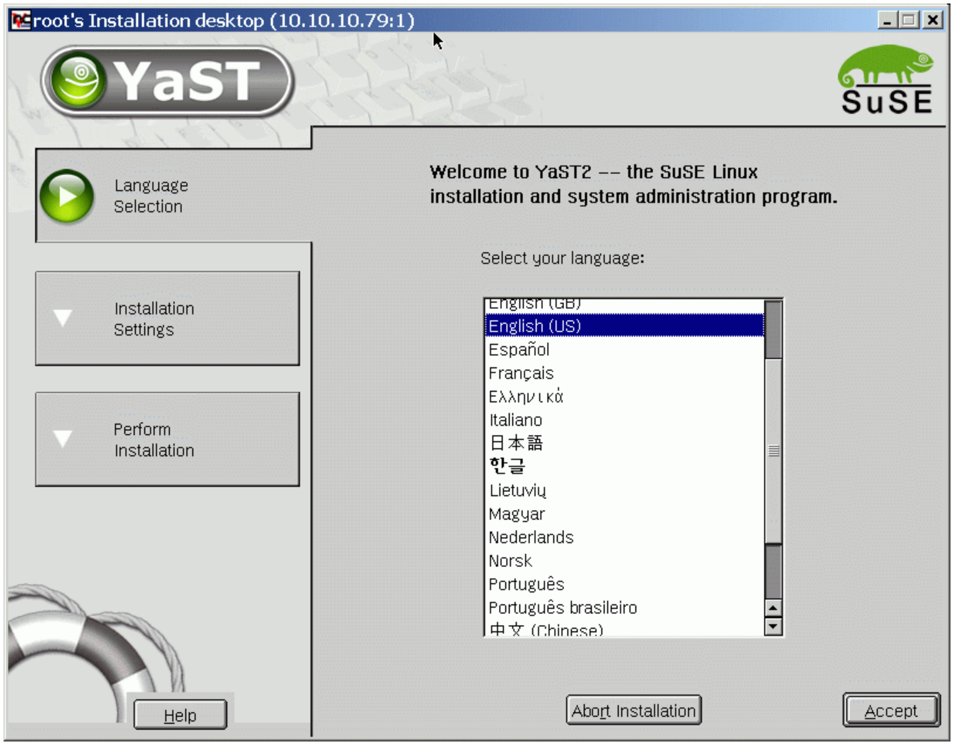
Figure 4-15 Enter IP address as specified in boot parameter



Session password:

Figure 4-16 Enter the password you have defined

If all settings are correct, the VNC client launches a window, which displays the output of YaST2 (see Figure 4-17).



root's Installation desktop (10.10.10.79:1)

Welcome to YaST2 -- the SuSE Linux installation and system administration program.

Select your language:

Figure 4-17   YaST2 in VNC window

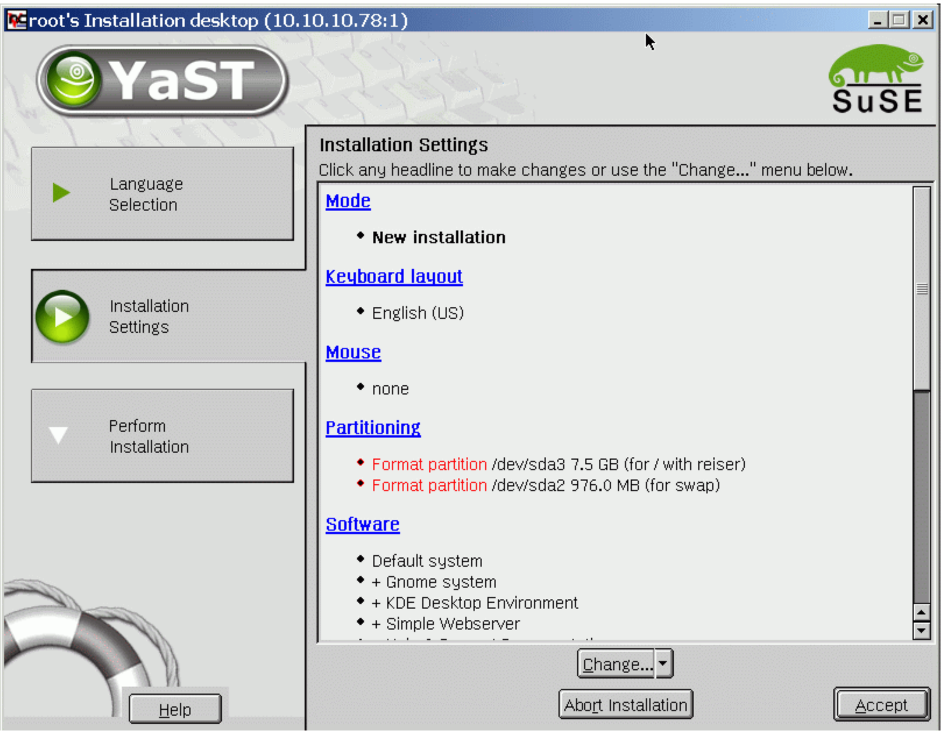
The installation steps that follow are trivial and very similar to other types of installation (introduced at the beginning of this section) or even comparing to other platforms, like zSeries and xSeries. Start by selecting a language, and you will get a window as shown in Figure 4-18 where YaST2 allows you to override default settings with your own (if needed). To change the default settings simply click the blue underlined items (Mode, Keyboard Layout, Mouse, Partitioning or Software). Each item leads to a new submenu where you can specify and apply the changes.

The particularities of your environment and its intended usage dictate specific settings, like partitions for instance, and are not detailed in this book. However, since we are using an Emulex adapter for the purpose our experiment, we review here the specifics of installing the Emulex device driver and the software packages it requires.

The Emulex device driver comes as so called “open build” which basically means that you have to compile sources to get a loadable module. An open build driver, like the Emulex driver, differs from other source code drivers as it consists out of two parts: the lower level driver component, which is delivered as object code, and the higher level driver, which comes as source code. To make a new device driver, the high level driver source must be compiled and the resulting object is linked with the lower level driver object into a single loadable module.

To do this you have to install the following software packages:

- The kernel source
- Compiler (remember that you need the cross compiler for 64 bits)
- All binutils and libraries required by the compiler´


root's Installation desktop (10.10.10.78:1)

Installation Settings

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

Mode

- New installation

Keyboard layout

- English (US)

Mouse

- none

Partitioning

- Format partition /dev/sda3 7.5 GB (for / with reiser)
- Format partition /dev/sda2 976.0 MB (for swap)

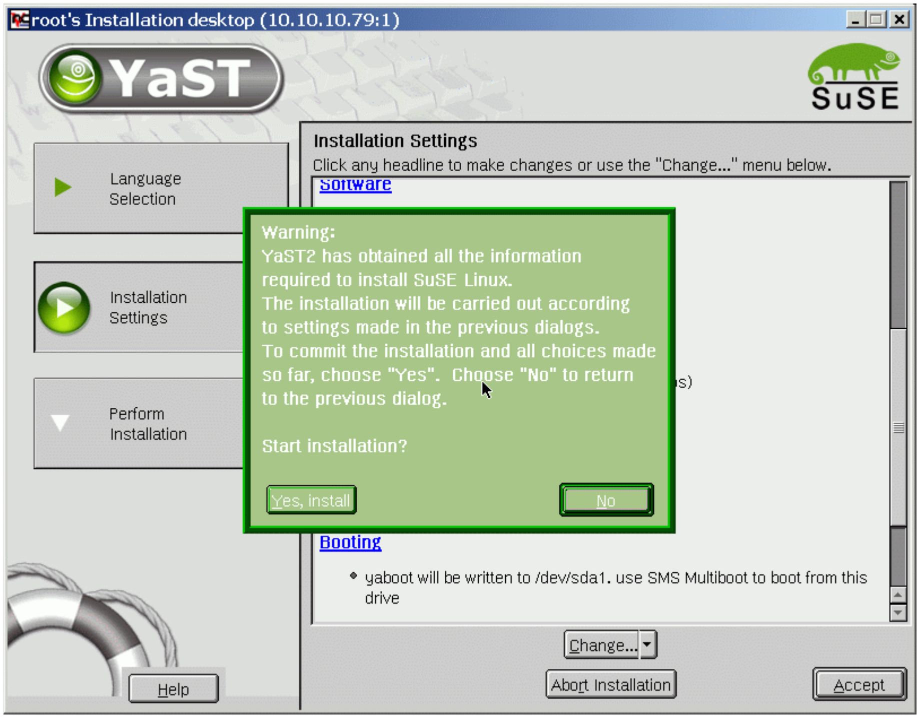
Software

- Default system
- + Gnome system
- + KDE Desktop Environment
- + Simple Webserver

Figure 4-18 Installation settings

After you have made all your changes, click **Accept** (Figure 4-18) to save the new settings.

YaST2 asks you for a final confirmation (Figure 4-19). After you accept, the installation process starts, partitioning the disk(s). Please check your settings before you confirm; specially disk partitioning, as it will wipe out any former data. For more information and advice on disk partitioning, refer to Appendix A., “Storage from the OS view” on page 253.


root's Installation desktop (10.10.10.79:1)

Installation Settings

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

Warning:
YaST2 has obtained all the information required to install SuSE Linux.
The installation will be carried out according to settings made in the previous dialogs.
To commit the installation and all choices made so far, choose "Yes". Choose "No" to return to the previous dialog.

Start installation?

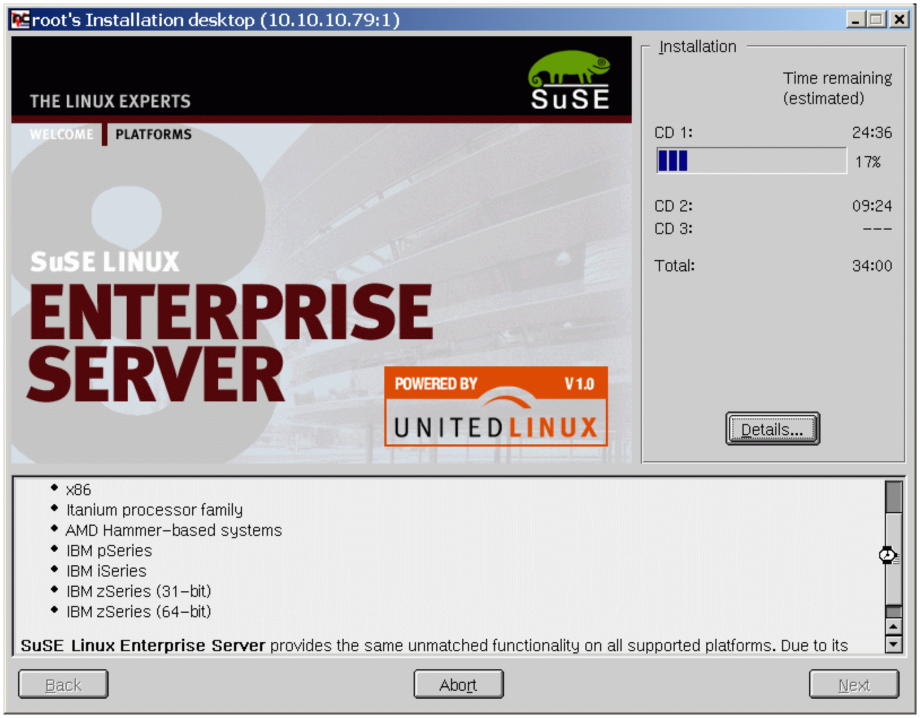
Booting

- yaboot will be written to /dev/sda1. use SMS Multiboot to boot from this drive

Figure 4-19 Start of Linux installation

Once the disk is ready, the installation of the software packages begins (Figure 4-20).

During the installation of the selected software packages, the system can occasionally prompt you to insert another CD.



- x86
- Itanium processor family
- AMD Hammer–based systems
- IBM pSeries
- IBM iSeries
- IBM zSeries (31–bit)
- IBM zSeries (64–bit)

SuSE Linux Enterprise Server provides the same unmatched functionality on all supported platforms. Due to its

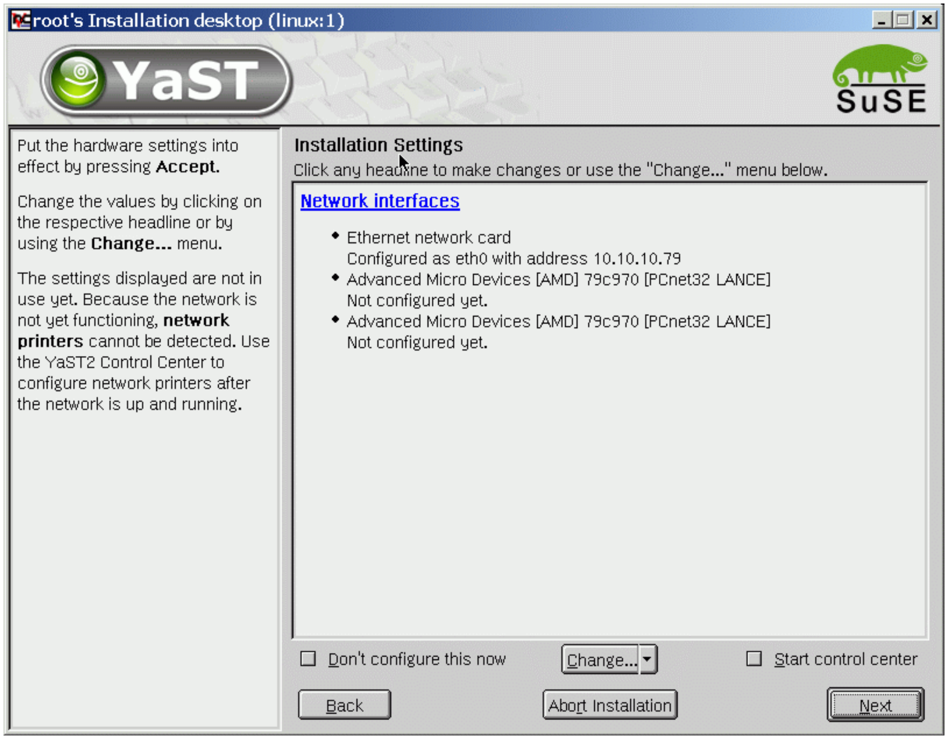
Figure 4-20 Package installation

Following the installation of all software packages, you are ready to finalize your configuration.

First, make choice of a root password, create users as desired, and wait until the corresponding configuration scripts are completed.

Finally, you can configure the network interfaces. The default entry is DHCP, which is generally not the right selection for a server system. To change this default, select **Network Interfaces** to get a configuration menu (see Figure 4-21).

Note: All settings can be changed at anytime by starting YaST2 (either from a command shell or in a graphical environment). YaST2 offers the dialog to apply and commit the changes.


root's Installation desktop (linux:1)

Put the hardware settings into effect by pressing **Accept**.

Change the values by clicking on the respective headline or by using the **Change...** menu.

The settings displayed are not in use yet. Because the network is not yet functioning, **network printers** cannot be detected. Use the YaST2 Control Center to configure network printers after the network is up and running.

Installation Settings

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

Network interfaces

- Ethernet network card
  Configured as eth0 with address 10.10.10.79
- Advanced Micro Devices [AMD] 79c970 [PCnet32 LANCE]
  Not configured yet.
- Advanced Micro Devices [AMD] 79c970 [PCnet32 LANCE]
  Not configured yet.

Figure 4-21 Installation settings of the hardware

Note: When you install your system via onboard video, you can configure your video settings within the same Installation Settings dialog. (Figure 4-21 in that case shows an additional underlined item for Video.)

Congratulations, your Linux installation is now complete.

4.4.2 Linux installation in an LPAR

We assume that you have the appropriate knowledge of handling high-end pSeries systems running in a partitioned environment. Otherwise, the redbook *The Complete Partitioning Guide on IBM eServer pSeries Servers*, SG24-7039, and the LPAR section of the *IBM eServer pSeries 690 Availability Best Practices* white paper might be a good start. See:

http://www.ibm.com/servers/eserver/pseries/hardware/whitepapers/p690_avail.html

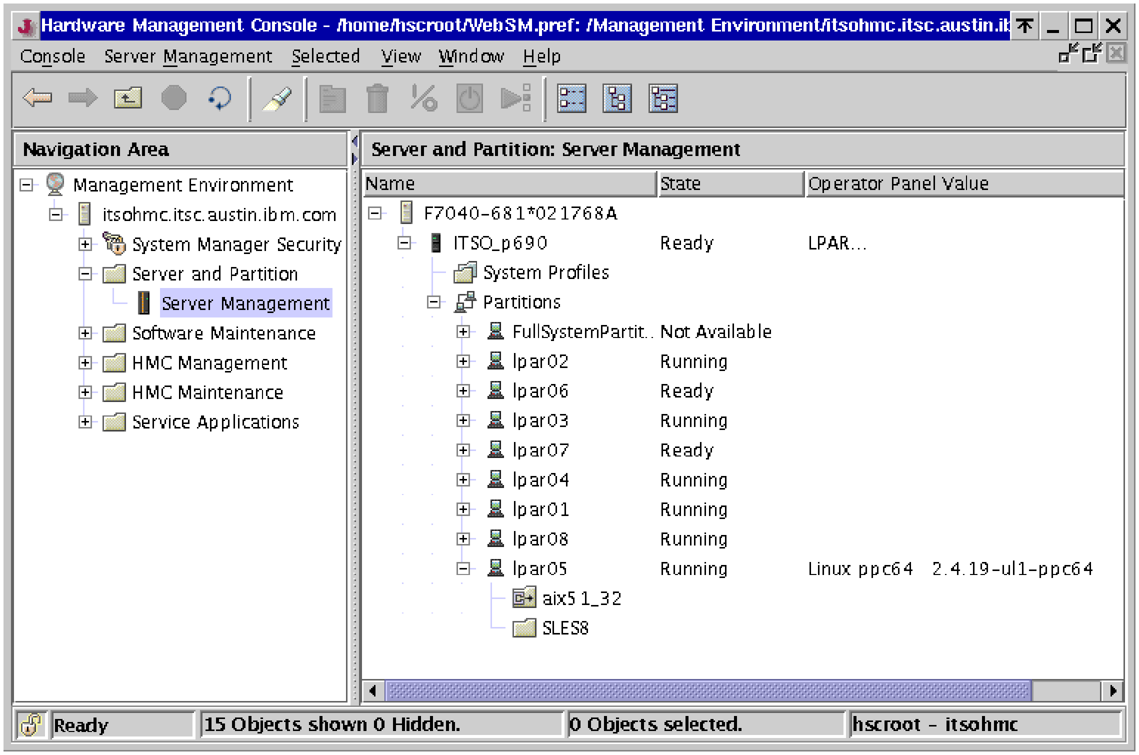
If not configured properly, any LPAR (including any that runs Linux) could impact the availability of the whole Regatta system transcending the LPAR itself.

There are many different hardware and software levels on the p670 and p690. These involve actual hardware parts, firmware upgrades, and software on the HMC. To successfully run Linux in a partition on the p670/p690, the system must be configured following the high availability configuration guide, and the system's firmware and HMC levels must be at GA2 or a higher level.

For additional details, and up-to-date information on supported hardware and software configurations, please consult the Release Notes® available on the pSeries Linux Web site:

`http://www.ibm.com/servers/eserver/pseries/linux/linux_release_notes.pdf`

For our tests, we use a p690 system with eight partitions defined (see Figure 4-22).

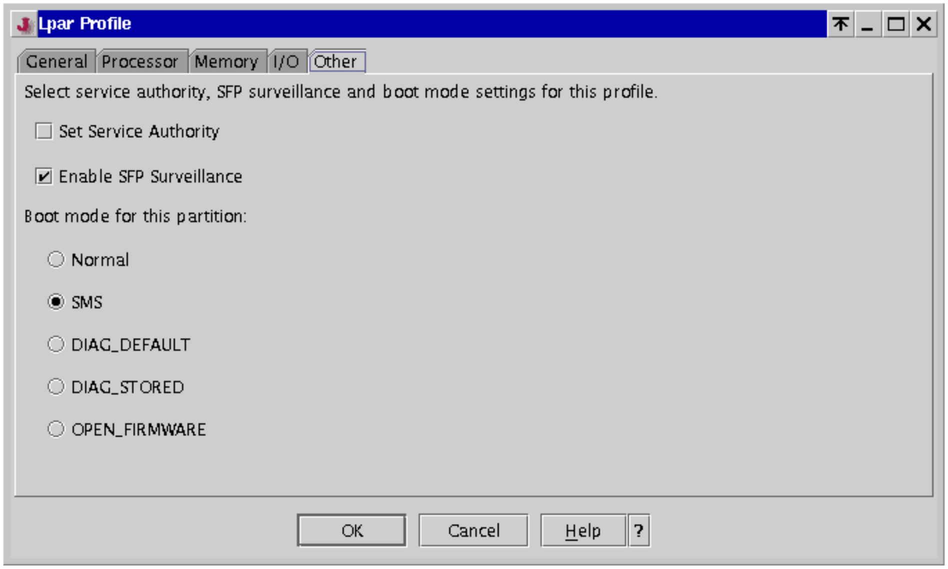


Name	State	Operator Panel Value
F7040-681*021768A		
ITSO_p690	Ready	LPAR...
System Profiles		
Partitions		
FullSystemPartit..	Not Available	
lpar02	Running	
lpar06	Ready	
lpar03	Running	
lpar07	Ready	
lpar04	Running	
lpar01	Running	
lpar08	Running	
lpar05	Running	Linux ppc64 2.4.19-ul1-ppc64
aix51_32		
SLES8		

Ready | 15 Objects shown 0 Hidden. | 0 Objects selected. | hscroot - itsohmc

Figure 4-22 HMC partition setup

The Linux partition, LPAR05, has one processor minimum, two as desired, and four as maximum defined. Similarly, we have 1 GB of memory set as minimum, 2 GB as desired, and 4 GB as maximum.[1] The boot mode is set to SMS (see Figure 4-23). The SCSI adapter controlling the CD drive is allocated to the Linux partition.



Select service authority, SFP surveillance and boot mode settings for this profile.

□ Set Service Authority

☑ Enable SFP Surveillance

Boot mode for this partition:

○ Normal

● SMS

○ DIAG_DEFAULT

○ DIAG_STORED

○ OPEN_FIRMWARE

Figure 4-23   LPAR boot mode

Basically, installation on a LPAR is identical to the native Linux installation via serial console. The differences are that:

You must not forget activate the partition.

When you reach the screen displayed in Figure 4-13, you must select option **9) Other,** and then **VT320** for the terminal type.

Finally, you must update the boot configuration by selecting the appropriate boot device in SMS Multiboot and changing the LPAR boot mode (see Figure 4-23) to normal.

[1] minimum are the resources needed at least, desired are assigned if available and minimum requirements are met

4.4.3 Post-installation

Note: At the time of writing this book no multipath capabilities were available yet. Neither did the Emulex device driver support multiple paths (required for FAStT) nor was a SDD (required for ESS) available. The supported driver release might be different in terms of installation and distribution.

At this time there is also no PPC 32-bit support.

SLES 8 contains out-of-the-box, the Emulex device driver V4.20. Prior to the installation of the new version, any previous version of the driver must be uninstalled; to do this, enter the following command from a commands shell:

```
rpm -qa | grep emulex
```

to get the name of the currently installed driver package. You will get a response like `emulex-4.20-63`. To remove the package, enter the command:

```
rpm -e emulex-4.20-63.
```

After Linux is fully installed you have to install the package containing the Emulex Open Build device driver.

Install the Emulex driver package you have obtained from SuSE as shown in Figure 4-24. When released the new Emulex driver package will part of an update CD or available on the SuSE support pages.

The package consists out of the following components:

- The source code for storage and networking (not covered in this book)
- Tools
- Library for the tools
- Shell scripts for installation
- Make files

We are assuming that the kernel sources, compiler, and libraries are already installed. To compile the code successfully, make sure you meet the following prerequisites:

- `/usr/src/linux` links to the kernel source; if not either establish the link or edit the makefile by editing the `BASEINCLUDE` variable.
- Synchronize the versions of driver and kernel using the command.
- `# cp /boot/vmlinuz.version.h /lib/modules/<kernel-version>/build/include/linux/version.h`
- Uninstall any existing Emulex driver package.

- Update the kernel and source to the required level.

Tip: The pre-release driver we used, required to build a new kernel image. Otherwise, the driver failed to load because of unresolved symbols. However, only the build is necessary, not the installation of the new kernel.

```
linux:/emulex/suse # rpm -Uvh emulex-4.21-0.src.rpm
emulex
##################################################
linux:/emulex/suse # rpm -Uvh km_emulex-4.21-0.ppc.rpm
km_emulex
##################################################
linux:/emulex/suse # rpm -Uvh emulex-4.21-0.ppc.rpm
emulex
##################################################
linux:/emulex/suse # cd /usr/src/kernel-modules/emulex/
linux:/usr/src/kernel-modules/emulex # ls -l
total 1285
drwxr-xr-x    3 root     root          584 Mar 26 07:41 .
drwxr-xr-x    3 root     root           72 Mar 26 07:41 ..
-rw-r--r--    1 root     root         1492 Feb 19 00:05 Install.sh
-rw-r--r--    1 root     root         3739 Feb 21 22:48 Makefile
-rw-r--r--    1 root     root         3294 Feb 19 00:05 Makefile.kernel
-rw-r--r--    1 root     root         3826 Feb 21 22:48 Makefile.module
-rw-r--r--    1 root     root         3433 Feb 19 00:05 README
-rw-r--r--    1 root     root         1227 Feb 19 00:05 Remove.sh
-rwxr-xr-x    1 root     root        99812 Feb 21 22:48 dfc
-rw-r--r--    1 root     root       294091 Feb 19 00:05 fcLINUXfcp.c
-rw-r--r--    1 root     root        13378 Feb 19 00:05 fcLINUXlan.c
drwxr-xr-x    2 root     root          368 Mar 26 07:41 include
-rwxr-xr-x    1 root     root        17716 Feb 21 22:48 libHBAAPI.so
-rw-r--r--    1 root     root        27778 Feb 19 00:05 libdfc.a
-rwxr-xr-x    1 root     root       108273 Feb 21 22:48 libemulexhbaapi.so
-rw-r--r--    1 root     root        22686 Feb 19 00:05 lpfc.conf.c
-rw-r--r--    1 root     root       136007 Feb 19 00:05 lpfc.conf.defs
-rw-r--r--    1 root     root         2362 Feb 19 00:05 lpfc_tar.spec
-rw-r--r--    1 root     root       401333 Feb 19 00:05 lpfcdriver
-rwxr-xr-x    1 root     root       146062 Feb 21 22:48 lputil
```

Figure 4-24 Installation of the Emulex driver packages

Change to the directory where the driver sources reside after copy. You can build the driver with the command **`make build`**

To copy the driver to the appropriate directory, you enter **`make install`**

To test the module, load it using the command **insmod lpfcdd**

In order to load the module with the **modprobe** command you have to do the following:

1. Add the line `alias scsi_hostadapter lpfcdd` to the file `/etc/modules.conf`
2. Run the command **depmod -a** to make the module dependencies.
3. Change the kernel append to `append=”max_scsi_luns=128”` (this allows the kernel to support up to 128 LUNs, default is 0!)

To start your system with the driver available at boot time, you have to build an appropriate initial ramdisk:

1. Add the name of the module to the file /etc/systemconfig/kernel
2. Rebuild your ramdisk with the command **mk_initrd**
3. Edit the file `/etc/lilo.conf` and add the configuration including the name of the initial ramdisk (see Example 4-1). Do not forget to launch `lilo` to commit the changes.

Example 4-1 Adding new initial ramdisk to lilo.conf

```
# Generated by YaST2

default=linux
timeout=100
boot=/dev/sda1
activate

image=/boot/vmlinuz-kernel_version
label=new_label
root=/dev/sda8
initrd=/boot/new_image_filename
read-only
append="max_scsi_luns=128"
```

After reboot, the driver is loaded. For details on how to make the storage usable for Linux, please refer to Appendix A, “Storage from the OS view” on page 253.

5

Linux on xSeries and BladeCenter

This chapter covers Linux installation on xSeries and BladeCenter in conjunction with external storage.

First, it provides the step-by-step procedure for installing Red Hat Enterprise Linux Advanced Server (AS) on IBM xServers and BladeCenter. Then, following the same structure, we cover UnitedLinux 1.0 as part of the SuSE Linux Enterprise Server (SLES 8) distribution.

Following the basic installation of both products, this chapter reviews required post-installation tasks and caveats for specific server types.

The chapter ends with instructions and special considerations for attaching external disk storage, and discusses the SDD multipath driver required for ESS, while differences between the Red Hat and SuSE implementations are highlighted.

5.1 Introduction to xSeries, BladeCenter

IBM xSeries™ servers leverage IBM Enterprise X-Architecture™'s state-of-the-art, industry-leading technologies that deliver mainframe-class power, and enterprise scalability and availability at very attractive prices. IBM is working closely with leading Linux distributors — Red Hat and the founders of UnitedLinux including the SCO Group, SuSE Linux AG, and Turbolinux, Inc., to offer tested and validated configurations for the full line of xSeries servers to ensure maximum performance and functionality across the full line of xSeries servers.

Blade servers are a relatively new technology that has captured industry focus because of its modular design, which can reduce cost with a more efficient use of valuable floor space, and its simplified management, which can help to speed up such tasks as deploying, repositioning, updating, and troubleshooting hundreds of Blade servers. All this can be done remotely with one graphical console using IBM Director systems management tools. In addition, Blade servers provide improved performance by doubling current rack density. By integrating resources and sharing key components, not only will costs be reduced, but also, availability will be increased. The BladeCenter is ideally placed for High Performance Computing clusters and Web farms. For more details, see the following Redpaper *The Cutting Edge: The IBM eServer BladeCenter*, REDP3581.

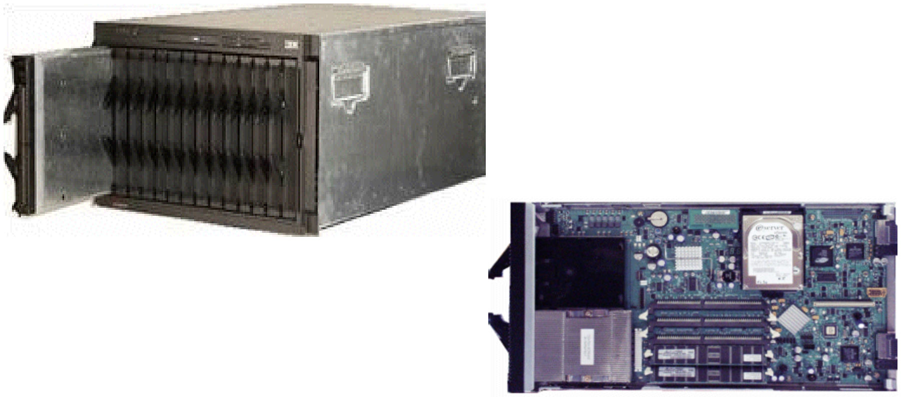


Figure 5-1 IBM BladeCenter chassis and Blade

Both Red Hat Enterprise Linux and SuSE Linux Enterprise Server provide enterprise class features that enable Linux-based solutions to be deployed across the widest range of enterprise IT environments.

5.2 Requirements

Prior to installation, we must first consider the system requirements. Especially, the impact of the attached networks and storage on overall performance has to be considered. Some recommendations to follow are:

- Design the storage subsystem carefully (RAID-level, stripe size, number of hard drives versus capacity). See *Tuning IBM eServer xSeries Servers for Performance*, SG24-5287.
- Check for appropriate cache policy usage on FAStT (see *IBM TotalStorage FAStT700 and Copy Services*, SG24-6808).

Prior to installation, all firmware and BIOS levels must be verified and updated if necessary. To get the latest information and downloads, visit the following site:

http://www.pc.ibm.com/support/

On the xSeries pages, you will find the latest drivers and firmware releases, as well as documentation about installation for specific models.

Remember to also check compatibility of you disk systems by visiting:

http://www.storage.ibm.com/proven/index.html

5.2.1 Hardware requirements

The xSeries servers currently range from single processor to 16-way processor systems. Memory ranges from 128 MB to 64 GB. Requirements will be dependant on workload: for example, a large database would be well suited to a system with plenty of memory and CPU power, whereas a Web server would be better with more networking resources. Many options are available and supported in IBM xSeries servers. You can view the latest list at the following site:

http://www.pc.ibm.com/us/compat/index.html

Red Hat and Suse also have a list of tested and supported hardware with their distribution of Linux, respectively at:

http://hardware.redhat.com/hcl or
http://www.suse.com/us/business/certifications/certified_hardware

For attaching to external storage, you will need a Host Bus Adapter. You can find the list of supported adapters by going to:

http://www.storage.ibm.com

After selecting your storage type, look for interoperabilty matrix.

For the purpose of this book, we used an IBM FAStT FC2-133 (note that although the name seems to indicate it is for FAStT, it can be used to attach to other storage systems, e.g., ESS).

5.2.2 Software requirements

The following is a list of software used to set up Fibre Channel storage with the xSeries and BladeCenter.

- Red Hat Enterprise Linux (release AS if setting up a high availability cluster), or SuSE Linux Enterprise Server 8 (SLES8)
- IBM FAStT Host Bus Adapter drivers

If you attach to FAStT, you will need:

- Storage Manager
- IBM FAStT Management Suite for Java (MSJ)

If you attach to ESS, you will need:

- Subsystem Device Driver (SDD)
- ESS Specialist

To configure the BladeCenter FC switch modules you will need:

- BladeCenter SAN utility

5.2.3 Connection requirements

A network switch is required for systems management and to configure the Fibre Channel storage. The servers may be connected to the storage via fibre switches or hubs, or if there is only one to four servers they can be connected directly to the FAStT with the correct amount of mini-hubs.

All connections have to be planned and set up carefully. Cable length considerations are also important. The topic is beyond the scope of this book and we assume that you are familiar with SAN and its implications. For an introduction, please refer to *Introduction to Storage Area Networks*, SG24-5470. Additionally, we assume that ESS and FAStT are set up with redundant paths from the host system.

5.3 Configurations used for this redbook

In preparation of this redbook, we used equipment located at two different sites.

- **Site 1**: (We want to cover High Availability clustering for Red Hat, see Chapter 6, “Red Hat Cluster Manager” on page 139, so we included two servers.)

  x440 (Node #1)

  - 2x 1.5GHz CPUs, Hyper Threading disabled
  - 4096MB RAM
  - 2x 36.4GB Ultra3 10 Krpms HDDs
  - 2x Intel 10/100 Ethernet Adapter
  - 2x FC-2 HBA, BIOS v1.29
  - 1x ServeRAID4Lx, BIOS v5.10

  x440 (Node #2)

  - 2x 1.4GHz CPUs, Hyper Threading disabled
  - 2048MB RAM
  - 2x 36.4GB Ultra3 10Krpms HDDs
  - 2x Intel 10/100 Ethernet Adapter
  - 2x FC-2 HBA, BIOS v1.29
  - 1x ServeRAID4Lx, BIOS v5.10

  BladeCenter

  - 1x Management Module
  - 2x Ethernet Switch Modules
  - 2x Fibre 2 Port Switch Modules

    HS20 Blade #1:

    - 2.4GHz CPU
    - 512MB RAM
    - 40GB IDE HDD
    - 2312 Fibre Expansion Card

    HS20 Blade #2:

    - 2.4GHz CPU
    - 512MB RAM
    - 40GB IDE HDD
    - 2312 Fibre Expansion Card

  STORAGE

  - 1x FAStT700
  - 1x EXP700
  - 14x 18.2GB HDDs

We are using direct connect with the x440's for the storage as we only have access to one 2109-F16 switch.

MISC

- 1x 4port KVM (Keyboard/Video/Mouse) switch
- 2x PDU's (Power Distribution Units)
- 1x 16port Netgear Fast Ethernet switch
- 1x 9513 Flat Panel Monitor
- 1x Enterprise rack

- **Site2**

  x440

  - x440 with two INTEL Xeon 1.6 GHz MP
  - 2 GB of memory
  - 2x IBM FC2-133 Fibre Channel Host adapter in slot 5 and 6
  - 2x 18 GB hard disk drives

  Storage

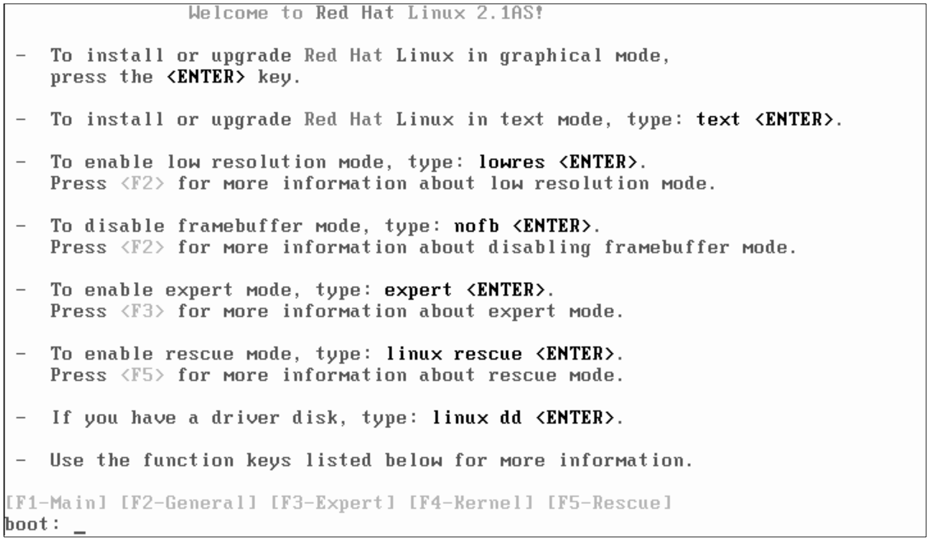
  - 2x FAStT700
  - 4x EXP700 with 14 36 GB HDDs
  - 2x IBM SAN switch 2109-F16
  - 1x ESS Model 800

5.4 Implementing Linux

In this section, we review the tasks pertaining to the installation of Red Hat Enterprise Linux on X Series and BladeCenter servers, then SuSE Linux Enterprise server on X Series and BladeCenter successively.

5.4.1 Installation steps for Red Hat on xSeries

Boot your server from CD #1 of the Red Hat Enterprise Linux distribution, after making sure that your system is set to boot from CD-ROM (this may require changes to the BIOS settings). The screen shown in Figure 5-2 is displayed.



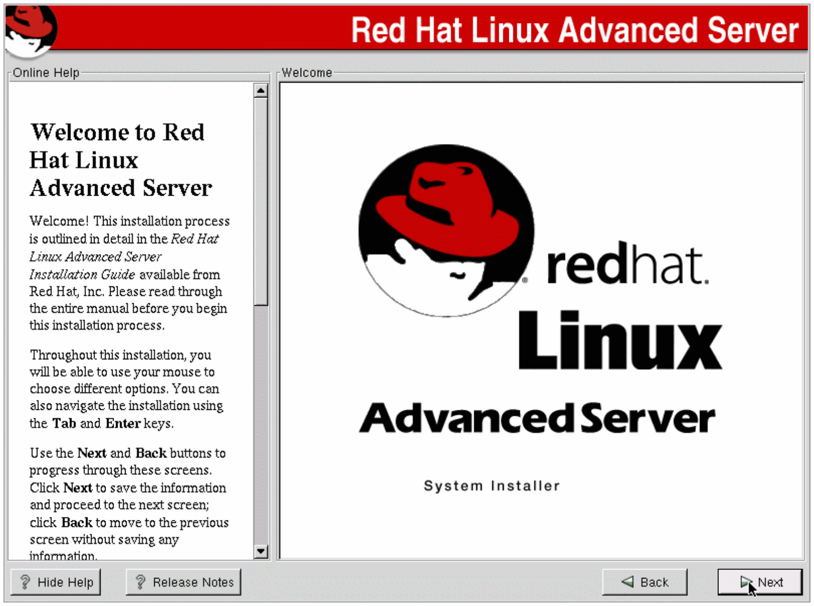
```
               Welcome to Red Hat Linux 2.1AS!

 -  To install or upgrade Red Hat Linux in graphical mode,
    press the <ENTER> key.

 -  To install or upgrade Red Hat Linux in text mode, type: text <ENTER>.

 -  To enable low resolution mode, type: lowres <ENTER>.
    Press <F2> for more information about low resolution mode.

 -  To disable framebuffer mode, type: nofb <ENTER>.
    Press <F2> for more information about disabling framebuffer mode.

 -  To enable expert mode, type: expert <ENTER>.
    Press <F3> for more information about expert mode.

 -  To enable rescue mode, type: linux rescue <ENTER>.
    Press <F5> for more information about rescue mode.

 -  If you have a driver disk, type: linux dd <ENTER>.

 -  Use the function keys listed below for more information.

[F1-Main] [F2-General] [F3-Expert] [F4-Kernel] [F5-Rescue]
boot: _
```

Figure 5-2 Initial boot screen

Press Enter at the initial boot window to begin the default graphical install.

A series of screens let you successively select the desired language and input peripherals. If you cannot find an exact match for your mouse, select the generic type that most closely resembles yours. If you are using a two-button mouse and you wish to emulate a three-button mouse, check the **Emulate three buttons** box. This is suggested for X-windows users, but is also handy in terminal mode for cutting and pasting as well. The third or middle button is emulated by pressing both buttons at the same time.

Read the text in the Welcome screen and click **Next** to continue (Figure 5-3).



Welcome to Red Hat Linux Advanced Server

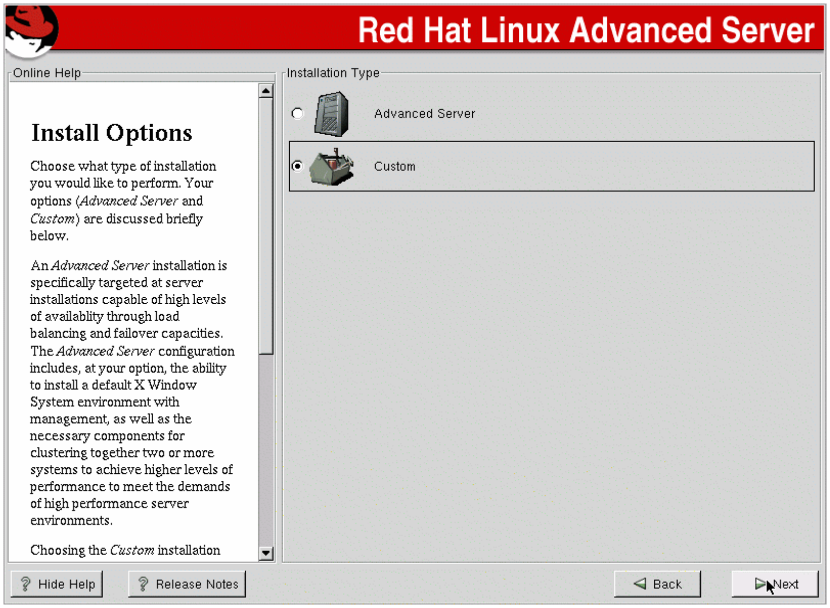
Welcome! This installation process is outlined in detail in the *Red Hat Linux Advanced Server Installation Guide* available from Red Hat, Inc. Please read through the entire manual before you begin this installation process.

Throughout this installation, you will be able to use your mouse to choose different options. You can also navigate the installation using the **Tab** and **Enter** keys.

Use the **Next** and **Back** buttons to progress through these screens. Click **Next** to save the information and proceed to the next screen; click **Back** to move to the previous screen without saving any information.

Figure 5-3 Welcome screen

Select a **Custom** installation and click **Next** (Figure 5-4).



Install Options

Choose what type of installation you would like to perform. Your options (*Advanced Server* and *Custom*) are discussed briefly below.

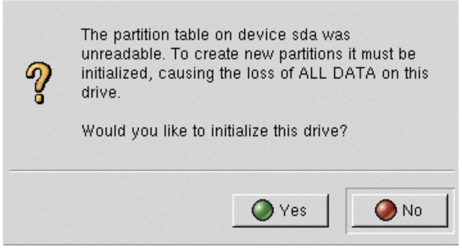
An *Advanced Server* installation is specifically targeted at server installations capable of high levels of availablity through load balancing and failover capacities. The *Advanced Server* configuration includes, at your option, the ability to install a default X Window System environment with management, as well as the necessary components for clustering together two or more systems to achieve higher levels of performance to meet the demands of high performance server environments.

Choosing the *Custom* installation

Figure 5-4 Install Options screen

Be cautious if an installation has already been done on the drives. If you have just created new logical drives on a ServeRAID™ controller, the installer will prompt you to initialize the disk (Figure 5-5).

Note: Initialization of disks causes 100% data loss!



The partition table on device sda was unreadable. To create new partitions it must be initialized, causing the loss of ALL DATA on this drive.

Would you like to initialize this drive?

Figure 5-5 Initialize disk message

We have chosen a **Custom** installation. Then to set up the partitions, Disk Druid, (a GUI tool similar to fdisk) is launched. You need to create the partitions. A type (Linux, Swap,...) has to be chosen. After all partitions are defined, the mount point of the individual disks has to be defined.

For example, you can choose four partitions on an empty drive with 18 GB capacity:

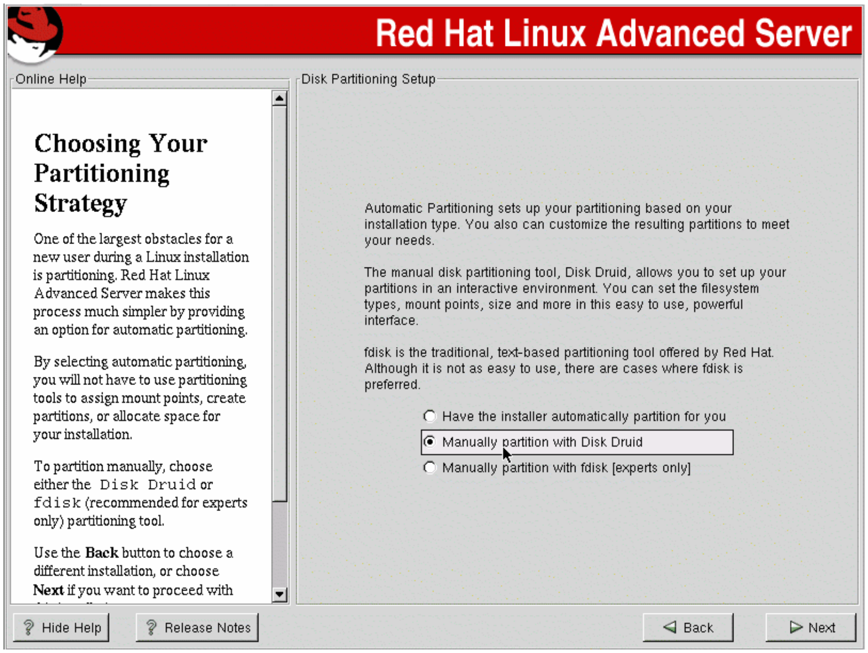
- For 20 MB, primary, select **Linux, mount point /boot**
- For 1 to 2 times the physical memory, select **SWAP**.
- For 3000MB, select **Linux, mount point /usr**
- For the rest select **Linux, mount point /** (your root file system)

Please consult the *Red Hat Linux Advanced Server Installation Guide* and Appendix A., “Storage from the OS view” on page 253 of this book for additional information on partitions and file systems.

Selecting **Have the installer automatically partition for you** creates the following partitions, and mount points:

- Swap is determined by the amount of memory in your system and the amount of space available on your hard drive. If you have 128MB of RAM, then the swap partition created can be 128MB - 256MB (twice your RAM), depending on how much disk space is available. (The maximum swap partition size possible is 2047MB)
- 47 MB /boot
- 2096 MB /

In a server environment, it is recommended that you select **Manually partition with Disk Druid** (Figure 5-6) and create your own partitions.



Choosing Your Partitioning Strategy

One of the largest obstacles for a new user during a Linux installation is partitioning. Red Hat Linux Advanced Server makes this process much simpler by providing an option for automatic partitioning.

By selecting automatic partitioning, you will not have to use partitioning tools to assign mount points, create partitions, or allocate space for your installation.

To partition manually, choose either the `Disk Druid` or `fdisk` (recommended for experts only) partitioning tool.

Use the **Back** button to choose a different installation, or choose **Next** if you want to proceed with

Automatic Partitioning sets up your partitioning based on your installation type. You also can customize the resulting partitions to meet your needs.

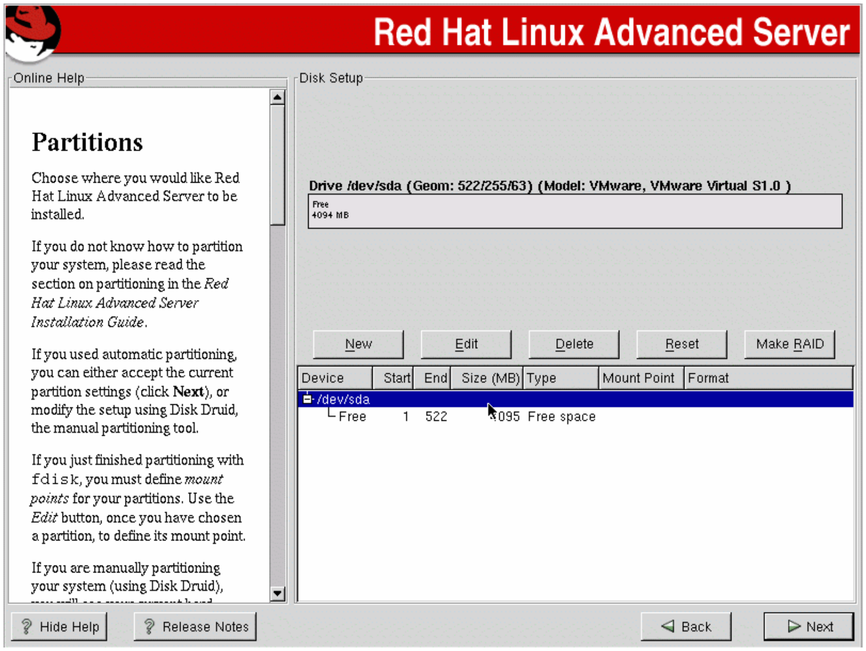
The manual disk partitioning tool, Disk Druid, allows you to set up your partitions in an interactive environment. You can set the filesystem types, mount points, size and more in this easy to use, powerful interface.

fdisk is the traditional, text-based partitioning tool offered by Red Hat. Although it is not as easy to use, there are cases where fdisk is preferred.

○ Have the installer automatically partition for you
● Manually partition with Disk Druid
○ Manually partition with fdisk [experts only]

Figure 5-6 Manual Partitioning screen

To add a partition, click the **New** button (Figure 5-7).



Partitions

Choose where you would like Red Hat Linux Advanced Server to be installed.

If you do not know how to partition your system, please read the section on partitioning in the *Red Hat Linux Advanced Server Installation Guide*.

If you used automatic partitioning, you can either accept the current partition settings (click **Next**), or modify the setup using Disk Druid, the manual partitioning tool.

If you just finished partitioning with `fdisk`, you must define *mount points* for your partitions. Use the *Edit* button, once you have chosen a partition, to define its mount point.

If you are manually partitioning your system (using Disk Druid),

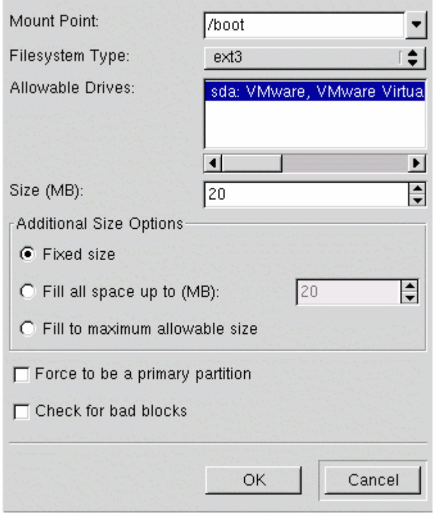
Drive /dev/sda (Geom: 522/255/63) (Model: VMware, VMware Virtual S1.0)

Free
4094 MB

Device	Start	End	Size (MB)	Type	Mount Point	Format
/dev/sda						
Free	1	522	4095	Free space		

Figure 5-7   Partitioning with Disk Druid

You will be prompted by the following window (Figure 5-8).



Mount Point:	/boot
Filesystem Type:	ext3
Allowable Drives:	sda: VMware, VMware Virtua
Size (MB):	20

Additional Size Options

● Fixed size

○ Fill all space up to (MB): 20

○ Fill to maximum allowable size

□ Force to be a primary partition

□ Check for bad blocks

Figure 5-8   Adding a new partition

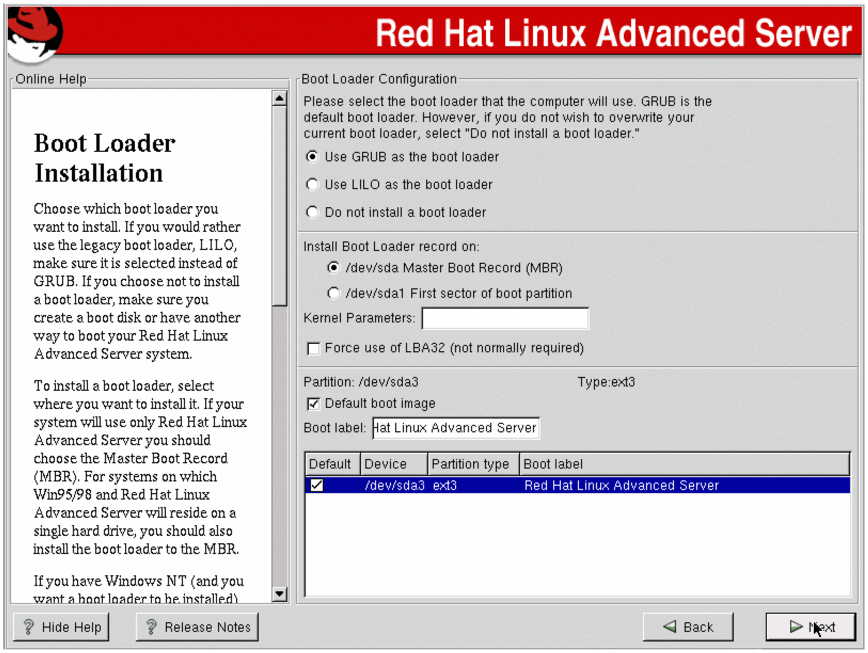
To select the mount point, click the drop-down button next to the **Mount Point** text box. This will present you with a menu of mount points such as `/boot`, `/usr`, etc. You do not need to specify a mount point for a *Swap* partition.

Type in the size of the partition in megabytes.

To select the partition type, click the drop-down button next to the **Filesystem Type** text box. This will present you with a menu of file system types such as `ext3`, `swap`, and so on. For your Swap partition, use `swap`. For all other Linux partitions use `ext3`.

> **Restriction:** If you plan to attach to ESS and use the SDD driver, do not select ext3, as it is not supported. Instead, use ext2, which is basically ext3 without journaling.

Next, the Boot Loader is installed (Figure 5-9).



Red Hat Linux Advanced Server

Online Help

Boot Loader Installation

Choose which boot loader you want to install. If you would rather use the legacy boot loader, LILO, make sure it is selected instead of GRUB. If you choose not to install a boot loader, make sure you create a boot disk or have another way to boot your Red Hat Linux Advanced Server system.

To install a boot loader, select where you want to install it. If your system will use only Red Hat Linux Advanced Server you should choose the Master Boot Record (MBR). For systems on which Win95/98 and Red Hat Linux Advanced Server will reside on a single hard drive, you should also install the boot loader to the MBR.

If you have Windows NT (and you want a boot loader to be installed)

Boot Loader Configuration

Please select the boot loader that the computer will use. GRUB is the default boot loader. However, if you do not wish to overwrite your current boot loader, select "Do not install a boot loader."

● Use GRUB as the boot loader

○ Use LILO as the boot loader

○ Do not install a boot loader

Install Boot Loader record on:

● /dev/sda Master Boot Record (MBR)

○ /dev/sda1 First sector of boot partition

Kernel Parameters:

□ Force use of LBA32 (not normally required)

Partition: /dev/sda3 Type:ext3

☑ Default boot image

Boot label: Hat Linux Advanced Server

Default	Device	Partition type	Boot label
☑	/dev/sda3	ext3	Red Hat Linux Advanced Server

Figure 5-9 Boot Loader installation screen

Select the boot loader that the server will use. GRUB is the default.

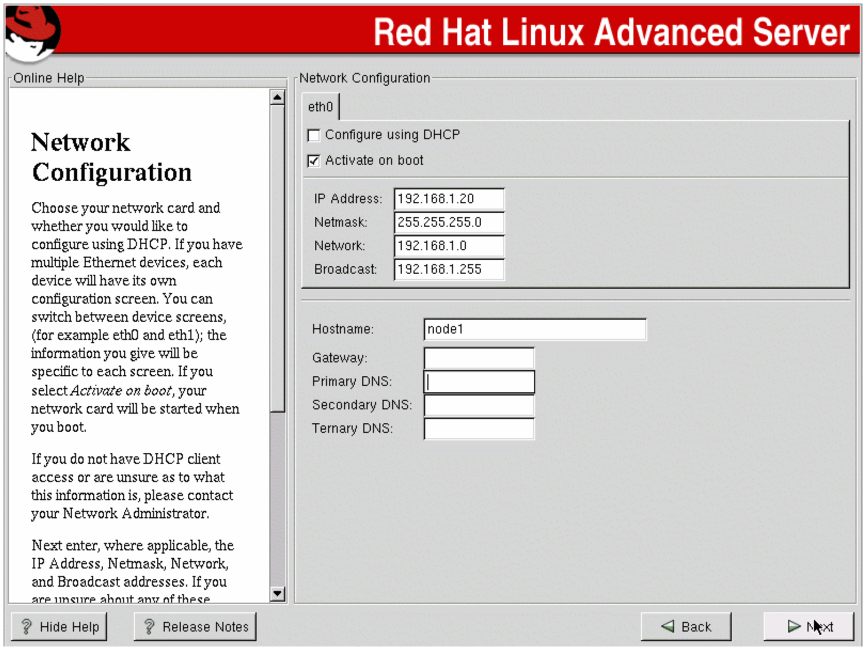
If you choose to install a boot loader (GRUB or LILO), then you must determine where it will be installed. If GRUB is the only boot manager you use, then choose **Master Boot Record (MBR)**. If you use the OS/2® boot manager, then GRUB has to be placed on the boot partition (if you have a partition with `/boot` otherwise `/`). We recommend installing GRUB on the Master Boot Record.

After making your selections, click **Next**.

If you selected the GRUB boot loader you will be asked if you wish to use a boot loader password.

For highest security we, recommended setting a password, but this is not necessary for more casual users. If used, enter a password and confirm the password. Click **Next** to continue.

At the next step, the network is configured (Figure 5-10).



Network Configuration

Choose your network card and whether you would like to configure using DHCP. If you have multiple Ethernet devices, each device will have its own configuration screen. You can switch between device screens, (for example eth0 and eth1); the information you give will be specific to each screen. If you select *Activate on boot*, your network card will be started when you boot.

If you do not have DHCP client access or are unsure as to what this information is, please contact your Network Administrator.

Next enter, where applicable, the IP Address, Netmask, Network, and Broadcast addresses. If you are unsure about any of these

□ Configure using DHCP
☑ Activate on boot

IP Address: 192.168.1.20
Netmask: 255.255.255.0
Network: 192.168.1.0
Broadcast: 192.168.1.255

Hostname: node1
Gateway:
Primary DNS:
Secondary DNS:
Ternary DNS:

Figure 5-10 Network Configuration screen

Each network adapter will have a corresponding tab. For each adapter, you should check the appropriate box to indicate whether the adapter:

- Will be configured via DHCP
- Should be activated on boot

If you are not using DHCP (highly likely for a server), you have to enter the following:

- IP address

- Netmask, Network, and Broadcast are automatically entered, but can be edited if different.
- Hostname
- Gateway
- Primary DNS
- Secondary DNS (optional)
- Ternary DNS (optional)

At this point, a firewall can be configured. The proper configuration of a firewall would be beyond the scope of this book. In our example, we used the preselected **Medium** settings. Select your security level, and click **Next**.

Note: A properly configured firewall can greatly increase the security of your system. For more details of the three security levels see the *Red Hat Linux Advanced Server Installation Guide*.

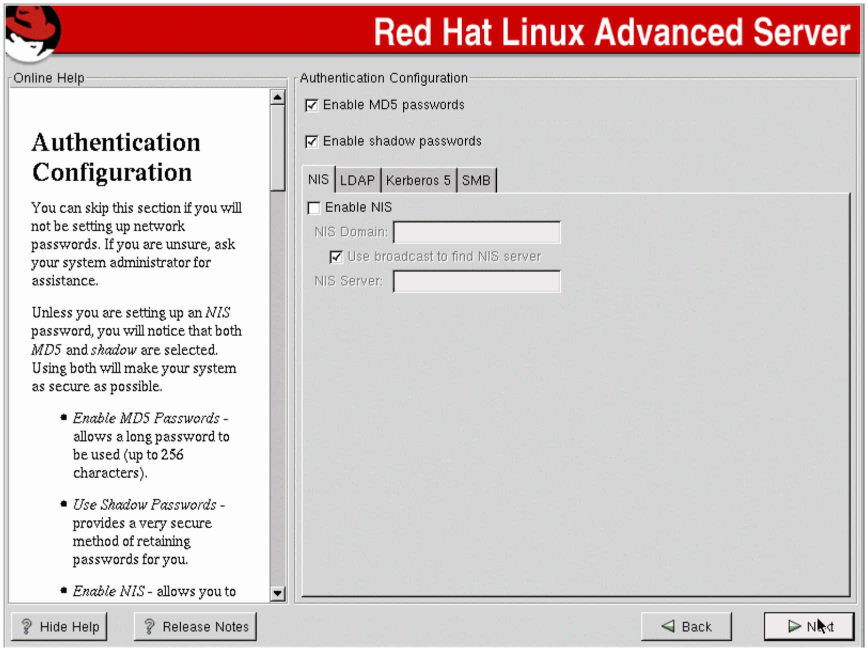
Next, is the Language Support Selection. Red Hat Linux Advanced Server supports multiple languages. You must select a language to use as the default language. Choosing only one language significantly saves on disk space.

Now, the time zone has to be configured. Select your time zone by clicking the appropriate location on the world map, or by selecting a location from the list shown in the dialog. Click **Next** on the screens to proceed.

The next window takes you to the account configuration.

First, you set the Root Password. Please do this carefully to prevent unauthorized use! Additionally, you can add other user accounts as well. Click **Next**.

Next, the authentication configuration is done (Figure 5-11).



Authentication Configuration

You can skip this section if you will not be setting up network passwords. If you are unsure, ask your system administrator for assistance.

Unless you are setting up an *NIS* password, you will notice that both *MD5* and *shadow* are selected. Using both will make your system as secure as possible.

- *Enable MD5 Passwords* - allows a long password to be used (up to 256 characters).
- *Use Shadow Passwords* - provides a very secure method of retaining passwords for you.
- *Enable NIS* - allows you to

☑ Enable MD5 passwords

☑ Enable shadow passwords

☐ Enable NIS

NIS Domain:

☑ Use broadcast to find NIS server

NIS Server:

Figure 5-11 Authentication Configuration screen

Clicking **Enable MD5 passwords** allows passwords up to 256 characters, instead of the standard 8 characters or less.

Clicking **Enable shadow passwords** replaces the standard `/etc/passwd` file with `/etc/shadow`, which can only be read by root user.

Clicking **Enable NIS** will allow you to add the Linux system to an existing NIS domain via a specific server or broadcast.

Clicking **Enable LDAP** will allow authentication via an LDAP directory server.

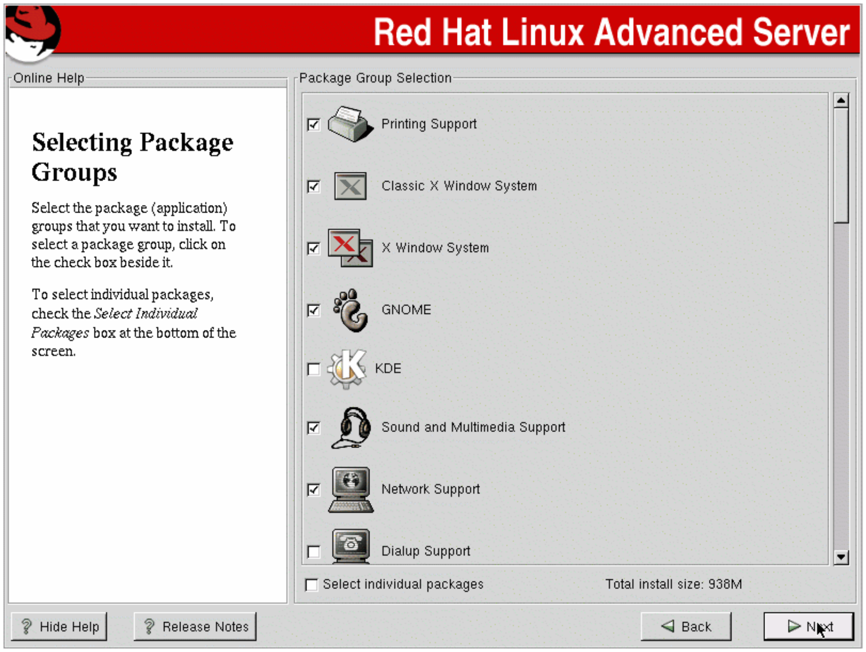
Clicking **Enable Kerberos** will allow authentication via a Realm, KDC, or Admin Server.

Clicking **SMB** will allow PAM to use an SMB server to authenticate users.

By default, **Enable MD5 passwords** and **Enable shadow passwords** are selected.

After making your selection, click **Next**.

The Selecting Package Group dialog box allows you to choose the packages you wish to install. Please keep in mind that for any changes involving driver compilation, the kernel sources have to be installed too (Figure 5-12).



Selecting Package Groups

Select the package (application) groups that you want to install. To select a package group, click on the check box beside it.

To select individual packages, check the *Select Individual Packages* box at the bottom of the screen.

☑ Printing Support

☑ Classic X Window System

☑ X Window System

☑ GNOME

□ KDE

☑ Sound and Multimedia Support

☑ Network Support

□ Dialup Support

□ Select individual packages

Total install size: 938M

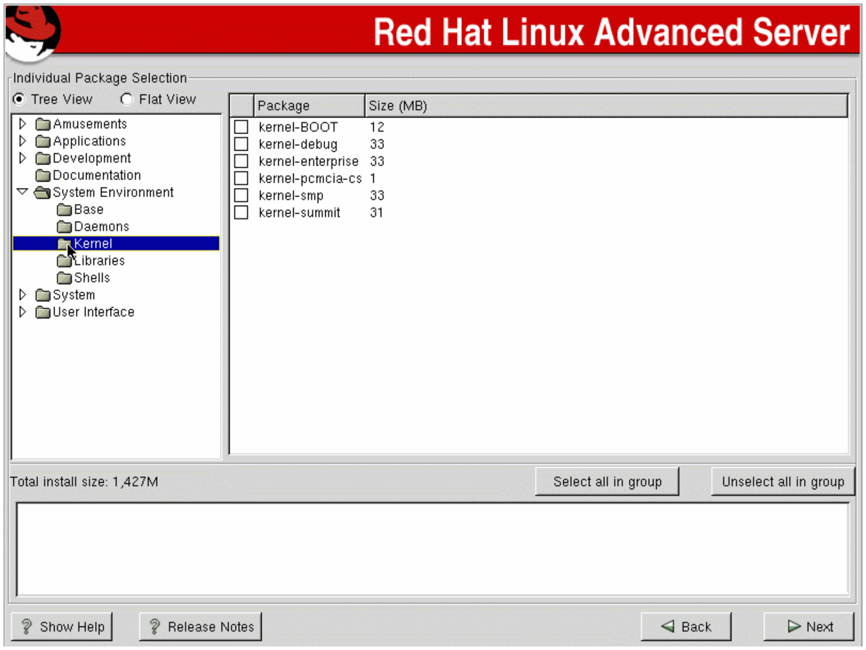
Figure 5-12 Selecting Package Group screen

For the custom installation, you can select groups of packages such as `Mail/WWW/News Tools`, or individual packages within these groups.

Scroll down the list and select the groups of packages you would like to install by clicking the check box next to them.

If you wish to select individual packages, click the **Select individual packages** check box. You should also check this box if you wish to install the supplied summit kernel (Figure 5-13).[1]

[1] If you install the summit kernel by selecting the individual packages, you will have to edit /etc/grub.conf in order to boot the kernel, or see it in the GRUB menu. A preferred method is to use the kernel-summit RPM post-install. The RPM script will automatically edit /etc/grub.conf for you. See 5.5.1 “Configuring for the Summit kernel (Red Hat)” on page 121.



● Tree View ○ Flat View

	Package	Size (MB)
□	kernel-BOOT	12
□	kernel-debug	33
□	kernel-enterprise	33
□	kernel-pcmcia-cs	1
□	kernel-smp	33
□	kernel-summit	31

Total install size: 1,427M

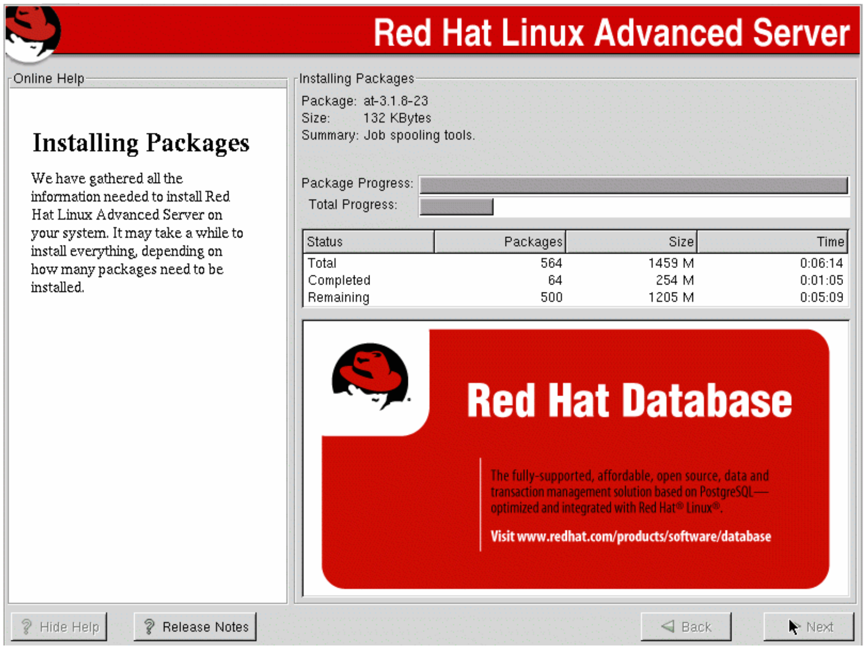
Figure 5-13 Individual Package selection screen

Next, you can configure your video. It is useful to know which video card or chipset you have in your system, however, you can have the installation probe your system to see if it finds any hardware it recognizes.

You can also select the **Skip X Configuration** check box if you wish to configure X after installation or not at all.

Now all required settings are made and you will get a screen telling you that the installation is about to begin. Click **Next** to begin copying the operating system files to your hard disk.

A screen similar to the one in Figure 5-14 keeps you informed over the progress of the installation.



Installing Packages

We have gathered all the information needed to install Red Hat Linux Advanced Server on your system. It may take a while to install everything, depending on how many packages need to be installed.

Package: at-3.1.8-23
Size: 132 KBytes
Summary: Job spooling tools.

Package Progress:
Total Progress:

Status	Packages	Size	Time
Total	564	1459 M	0:06:14
Completed	64	254 M	0:01:05
Remaining	500	1205 M	0:05:09

Red Hat Database
The fully-supported, affordable, open source, data and transaction management solution based on PostgreSQL—optimized and integrated with Red Hat® Linux®.
Visit www.redhat.com/products/software/database
Hide Help
Release Notes
Back
Next

Figure 5-14 Copying files screen

Once the file copy is complete, you are asked whether you want to have a boot disk created. Insert a diskette if you require the boot disk, otherwise click the **Skip boot disk creation** check box.

If you choose to use X you will configure your monitor and desktop settings on the two next screens. Select the **monitor model** or the **generic model** that most closely resembles it and click **Next**.

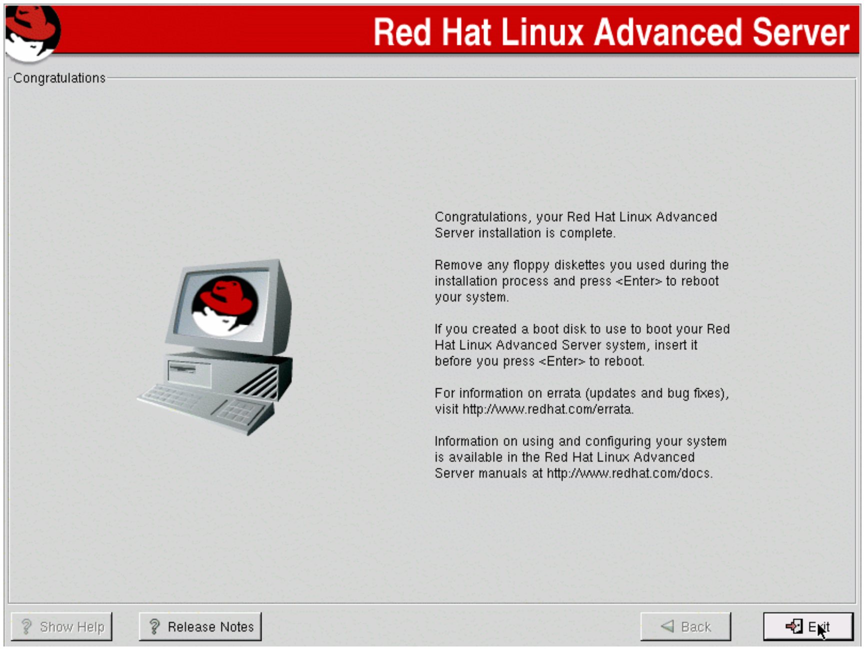
Important: Wrong monitor settings can damage your monitor!

Next, you can choose your default desktop and whether you would like to use a text or graphical login.

Now the installation is complete and it is time to reboot. Please remove the CD, and if generated, remove the boot disk (Figure 5-15).

If you intend to install on Red Hat on BladeCenter, please proceed to the next section.

For post-installation tasks an issues, please proceed to 5.5, “Post-installation” on page 121.



Congratulations, your Red Hat Linux Advanced Server installation is complete.

Remove any floppy diskettes you used during the installation process and press <Enter> to reboot your system.

If you created a boot disk to use to boot your Red Hat Linux Advanced Server system, insert it before you press <Enter> to reboot.

For information on errata (updates and bug fixes), visit http://www.redhat.com/errata.

Information on using and configuring your system is available in the Red Hat Linux Advanced Server manuals at http://www.redhat.com/docs.

Figure 5-15 End of installation

5.4.2 Installation steps for Red Hat on BladeCenter

Ensure that all BIOS and firmware are at the latest levels and any switches are installed prior to OS installation. See the Installation and User's Guides for the HS20 Fibre Channel Expansion Card, BladeCenter 2-Port Fibre Channel Switch Module, and Management Module for more details.

It is possible to install the operating system over the network. Using PXE boot allows multiple systems to be installed in parallel.

The following instructions assume you are installing Red Hat Enterprise Linux on a local IDE disk and not SCSI or external Fibre Channel storage. We will also be using CDs to install as we are only installing two servers. In the real world you may be installing to fourteen servers in a single BladeCenter chassis. In this situation, it would make more sense to perform a network installation. This would allow you to install to the fourteen servers simultaneously.

Note: At the time of writing, Red Hat Enterprise Linux was only supported by IBM on the BladeCenter with IDE disks. Booting from SCSI was at that time under test. Check the ServerProven® site for updates:

http://www.pc.ibm.com/us/compat/index.html

To install Red Hat Linux Advanced Server 2.1 on a BladeCenter you need a special set of boot disks to allow the installer to recognize the USB CD-ROM and USB floppy drives. The boot disks also allow graphical installs.

In addition, you need a driver disk to make use of the LSI SCSI controller, as well as the included Broadcom gigabit network adapters.

Obtain the Red Hat Enterprise Linux boot diskette and driver diskette from the following site:

http://redhat.com/support/partners/ibm/ibm_netfinity.html

Create the boot disk: `boot-AS-2.1-bladecenter-<version>.img` using `dd` under Linux or `rawrite` under Microsoft Windows.

Create the driver disk: `AS-2.1-dd.img` using `dd` under Linux or `rawrite` under Microsoft Windows. This disk is required for networking and if using SCSI disks.

Select the media tray and KVM on the first system and boot the system from the diskette. Type `linux dd` when prompted for a graphical installation or `text dd` for a text install.

Insert the driver diskette when prompted and press Enter.

Proceed with installation; see 5.4.1 “Installation steps for Red Hat on xSeries” on page 102 for more details and screen shots.

Note: Installation on the BladeCenter will give you an additional prompt asking if the installation method will be using `Local CD-ROM` or `Harddisk`, select **Local CD-ROM.**

For post-installation tasks and issues, consult 5.5, “Post-installation” on page 121.

5.4.3 Installation steps for SLES8 on xSeries

Power on your server and insert the SuSE SLES 8 CD in the CD drive, after making sure that your system is set to boot from CD-ROM (this may require changes to the BIOS settings). The screen shown in Figure 5-16 is displayed.

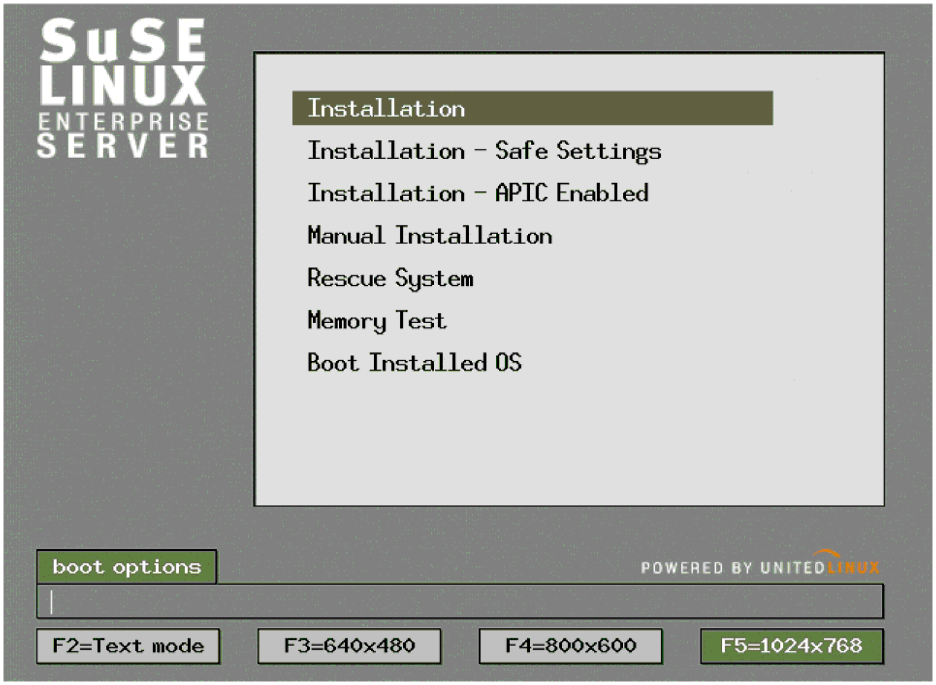


Figure 5-16 Boot prompt

The boot options field at the bottom of the screen allows you to specify additional options for loading the kernel. On older systems, for instance, you may want to disable DMA access to the CD-ROM by specifying `ide=nodma` as a boot option.

The push buttons at the very bottom of the screen allow you to choose text or graphical installation modes.

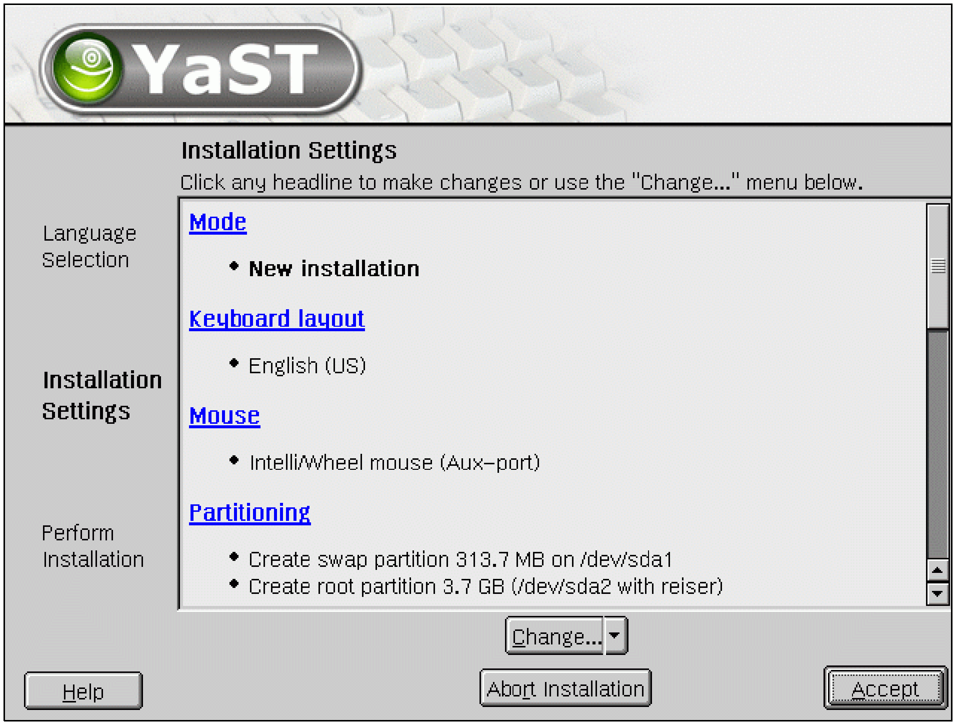
Tip: If your system does not boot after changes to the bootloader, please choose **Rescue System**, pick manually the module required to connect to your storage, and boot. The rescue system allows you to make the required repairs like writing a new MBR.

Press Enter or wait until the kernel is loaded automatically.

Once the kernel is loaded, the YaST installation utility starts and displays the licence agreement, followed (upon your acceptance) by the language selection screen.

The installation utility is now probing your hardware, loading (if available) the appropriate modules for the devices found. It is important to note that SuSE will only install drivers required to boot: from a disk driver standpoint, this avoids the problem potentially caused by having different drivers attach differently to the SCSI implementation; loading a second driver may cause a shift in the numbering of the SCSI devices.

After completing the hardware detection cycle, YaST presents you the screen shown in Figure 5-17. You can review and change **Installation settings** and make sure they fit your actual requirements. In particular, you will most likely change the boot disk partitioning and the software packages selection.



Installation Settings

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

Mode

- **New installation**

Keyboard layout

- English (US)

Mouse

- Intelli/Wheel mouse (Aux-port)

Partitioning

- Create swap partition 313.7 MB on /dev/sda1
- Create root partition 3.7 GB (/dev/sda2 with reiser)

Figure 5-17 SLES 8 Installation Settings screen

Please refer to section “Disk partitions” on page 255 in Appendix A for more information on disk partitioning under Linux. There are some system requirements that must be respected when you partition the disk, but it is also important to make sure that you select a file system suited for the type of application and storage system you are using.

When going through the software package selection, be careful to include all required sources (e.g. ncurses), kernel sources, compiler, and libraries. These packages contain the development environment you will need to compile the device drivers for the HBAs. After you have made your changes, click the **Accept** button (see Figure 5-17) to set and save the new settings.

YaST informs you that all settings are saved and that the installation is about to begin. Before clicking the **Yes** button, make sure one more times that you have set the appropriate settings; once you accept, the installation process starts with partitioning and data on disk, if any, will be lost.

The process continues with the installation of the selected packages from the SLES 8 and the UnitedLinux CDs.The system prompts you whenever it requires another CD in the set.

The process is complete when you are prompted to reboot; please remember to remove any bootable CD and diskette.

After reboot, you are now ready to configure your system. You begin the configuration by defining the root password, then you can eventually create additional users.

YaSt now displays suggested defaults for your desktop settings. Select **Change** if you want to adjust or select your own video, mouse, and keyboard setup. Once you accept, the values are committed.

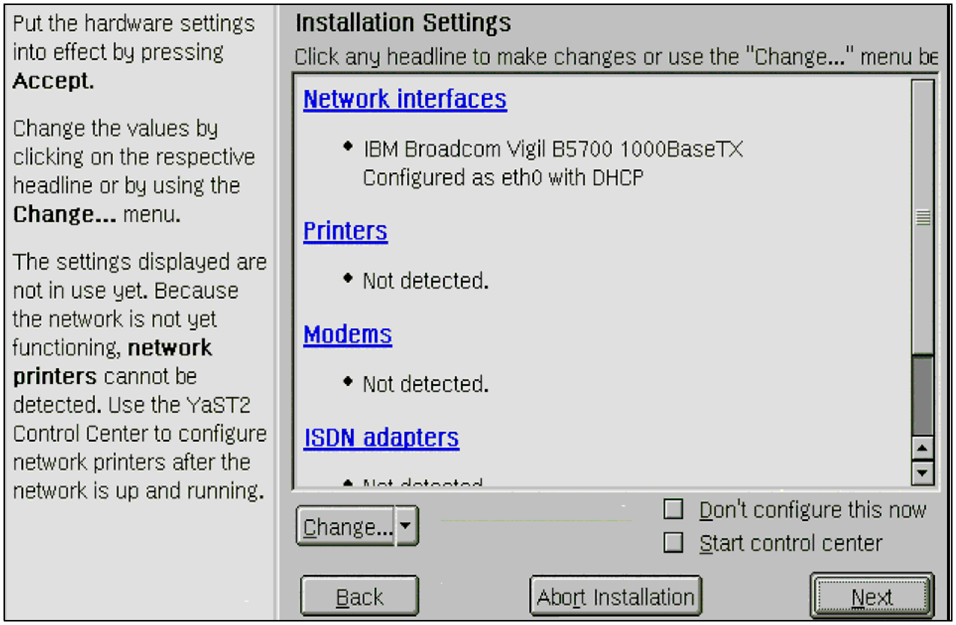
YaST now probes your system for additional peripherals like printers.

On the next screen (shown in Figure 5-18) you can review and change the settings for printer and all communication devices including the network, firewall, and server services. By default, the detected network adapters are set for DHCP. If you want a static address, which is usually the case for a server, click the **Network Interfaces** headline to make the appropriate changes.

Clicking **Next** saves the network configurations and applies all changes to the system (see Figure 5-19).

The system now reboots and starts the configured services with the settings selected.

The installation is complete.



Put the hardware settings into effect by pressing **Accept.**

Change the values by clicking on the respective headline or by using the **Change...** menu.

The settings displayed are not in use yet. Because the network is not yet functioning, **network printers** cannot be detected. Use the YaST2 Control Center to configure network printers after the network is up and running.

Installation Settings

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

Network interfaces

- IBM Broadcom Vigil B5700 1000BaseTX Configured as eth0 with DHCP

Printers

- Not detected.

Modems

- Not detected.

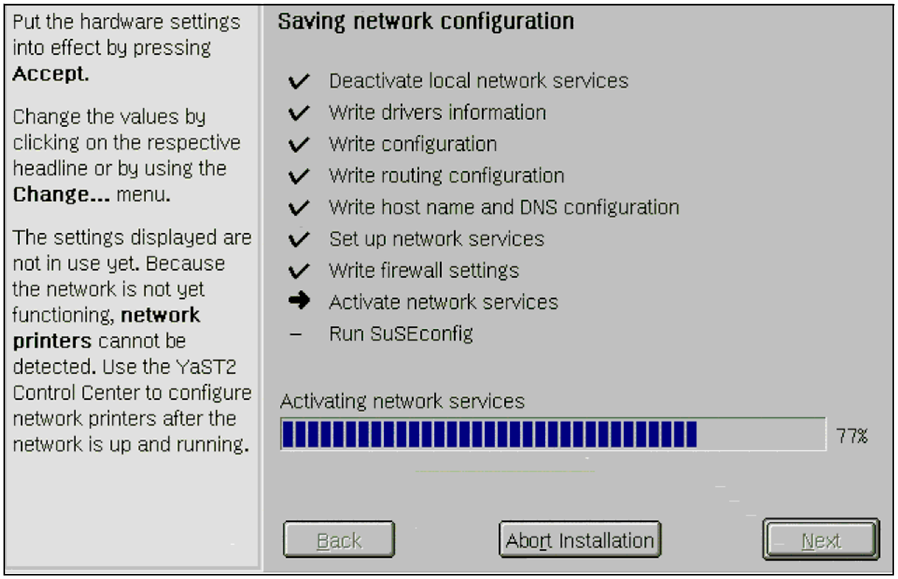
ISDN adapters

- Not detected.

□ Don't configure this now

□ Start control center

Figure 5-18 Network and peripherals setup



Put the hardware settings into effect by pressing **Accept.**

Change the values by clicking on the respective headline or by using the **Change...** menu.

The settings displayed are not in use yet. Because the network is not yet functioning, **network printers** cannot be detected. Use the YaST2 Control Center to configure network printers after the network is up and running.

Saving network configuration

- ✓ Deactivate local network services
- ✓ Write drivers information
- ✓ Write configuration
- ✓ Write routing configuration
- ✓ Write host name and DNS configuration
- ✓ Set up network services
- ✓ Write firewall settings
- ➜ Activate network services
- – Run SuSEconfig

Activating network services 77%

Figure 5-19 Apply network configuration

5.4.4 Installing SuSE Linux on BladeCenter

The Installation of UnitedLinux/SLES on the BladeCenter servers is very similar to 5.4.3 “Installation steps for SLES8 on xSeries” on page 117. There are a few particularities to be aware of though.

Mouse during installation

The mouse will not work initially when the installer starts. Use the keyboard to start the installation until you get to the Installation Settings screen. Press Alt+C to change the settings. Select **Mouse** from the menu, scroll down the list and select **USB Mouse**. Press Alt+T to test and then click **Accept** when finished.

Single processor or hyper-threading

If the server has more than one processor or has hyper-threading enabled, select **Booting,** and add the parameter `acpi=oldboot` in the Kernel Boot Parameters field.

5.5 Post-installation

This section addresses post-installation tasks that are critical, and thus required depending on your hardware or Linux distribution.

5.5.1 Configuring for the Summit kernel (Red Hat)

This only applies to Red Hat.

IBM’s new high-end Intel based servers take advantage of the new IBM Summit chipsets. Summit is the name for a set of technologies inspired by proven mainframe capabilities. Summit enables mainframe inspired capabilities in the areas of availability, scalability, and performance. Enhanced features include:

- Memory mirroring, redundant bit steering, chipkill memory
- L4 cache providing high speed communications between memory and CPU subsystems
- High performance PCI-X allows adapters significantly better memory access
- Increased scalability

We are using x440 servers, which include the Summit chipset. If you are using an x360 or a server without the Summit chipset, then you can move to the next section.

If you have an SMP system and you have booted the SMP or enterprise kernel, you will find that the loading of Linux will hang. You should boot into the system BIOS and enable hyper-threading:

Press F1 when prompted at the system BIOS. From the menu select **Advanced Setup** -> **CPU Options**. Enable **Hyper-Threading Technology**.

Once enabled you will be able to boot with the SMP or Enterprise kernels. The Summit kernel should be used on servers with this technology except the x360. Once the Summit kernel is installed hyper-threading can be switched off again if not required.

The summit kernel, source, and patch are supplied with the distribution CDs if required.

Obtain the latest supported summit kernel, headers, and source from the following site:
http://www.rhn.redhat.com
You will need to register your licensed copy of Red Hat Enterprise Linux to obtain updates for this OS.

Install the updates:

```
rpm -ivh kernel-summit-<version>.rpm
rpm -Uvh kernel-headers-<version>.rpm
rpm -Uvh kernel-source-<version>.rpm
rpm -ivh kernel-<version>.src.rpm
```

An entry will be added to `/etc/grub.conf` , which will allow you to boot from the Summit kernel. Edit the `default` field to boot from this kernel by default.

When using the summit kernel, it is also necessary to install a patch that is supplied with the `kernel<version>.src.rpm` file. Type the following commands to install the patch:

```
cd /usr/src/linux-<version>
patch -p1 < /usr/src/redhat/SOURCES/linux-<version>-summit.patch
cp -f configs/kernel-<version>-i686-summit.congif.config
```

5.5.2 Updating the kernel on BladeCenter (Red Hat)

This only applies to Red Hat. Obtain the latest supported kernel, headers, and source from the following site:

http://www.rhn.redhat.com

You will need to register your licensed copy of Red Hat Enterprise Linux to obtain updates for this OS.

Install the updates:

```
rpm -ivh kernel-<version>.rpm
rpm -Uvh kernel-headers-<version>.rpm
rpm -Uvh kernel-source-<version>.rpm
rpm -ivh kernel-<version>.src.rpm
```

An entry will be added to `/etc/grub.conf` , which will allow you to boot from the new kernel. Edit the `default` field to boot from this kernel by default.

5.5.3 Issues on BladeCenter

This section applies to Red Hat and SuSE.

Floppy after installation

You may be unable to mount the floppy after installation due to it being a USB device.

Red Hat

The floppy resides on the last SCSI device. For example, if you have LUNs configured on the fibre storage (/dev/sda, /dev/sdb), the floppy will reside on /dev/sdc. It is possible to manually mount the floppy using this information, or /etc/fstab can be edited with the following where X=device letter:

```
/dev/sdX      /mnt/floppy      auto     noauto,user     0,0
```

Create a directory for the floppy under `/mnt`

```
mkdir /mnt/floppy
```

The system is now configured to mount the floppy at boot time. To mount the floppy now type the following:

```
mount /mnt/floppy
```

SuSE

Type the following to load the USB storage module:

```
modprobe usb-storage
```

The floppy resides on the last SCSI device. For example, if you have LUNs configured on the fibre storage (/dev/sda, /dev/sdb), the floppy will reside on /dev/sdc. It is possible to manually mount the floppy using this information, or /etc/fstab can be edited with the following where X=device letter:

```
/dev/sdX     /media/floppy     auto     noauto,user,sync     0 0
```

Create a directory for the floppy under `/media`

```
mkdir /media/floppy
```

The system is now configured to mount the floppy at boot time. To mount the floppy now type the following:

```
mount /media/floppy
```

CD-ROM after installation

You may be unable to mount the CD-ROM after installation due to it being a USB device.

Red Hat

The CD-ROM will fail to mount if the media tray had been previously switched to another Blade system. Edit /etc/fstab to change /dev/cdrom to /dev/scd0:

```
/dev/scd0     /mnt/cdrom     iso9660     noauto,owner,kudzu,ro     0,0
```

The system is now configured to mount the CD-ROM at boot time. To mount the CD-ROM, now type the following:

```
mount /mnt/cdrom
```

If this system is rebooted and the media tray is moved to another system then this process may need to be repeated.

SuSE

The CD-ROM device will be known as /dev/sr0, so remove the device /dev/cdrom and create a link from sr0 to cdrom. Type the following:

```
rm /dev/cdrom; ln -s /dev/sr0 /dev/cdrom
```

Edit /etc/fstab to include the following line of text:

```
/dev/cdrom     /media/cdrom     auto     ro,noauto,user,exec     0 0
```

Type `mount /dev/cdrom /media/sr0` to manually mount a CD now.

Mouse after installation - SuSE only

The mouse may not work after the server restarts. Type the following to regain mouse controls:

```
/sbin/modprobe mousedev
```

Type the following to avoid having to repeat this every time the server reboots:

```
echo “/sbin/modprobe mousedev” >> /etc/init.d/boot.local
```

5.6 Setup for Fibre Channel attachment

In this section we discuss the setup required for Fibre Channel attachment for Red Hat Enterprise Linux and Suse Linux Enterprise Server.

First, install the Host Bus Adapters in your server.

Next, replace and update the BIOS on the adapter and apply the proper settings.

- For a Red Hat installation, you must now replace the default driver that was automatically installed by an appropriate driver for your storage (ESS or FAStT).
- For a Suse installation, no default driver was installed, and you must now install the appropriate driver for your storage (ESS or FAStT).

Finally, install SDD if you attach to ESS, or configure your paths if you attach to FAStT.

As described in 5.3 “Configurations used for this redbook” on page 101, we used IBM FC2 Host Bus Adapters (HBA). Although this HBA is also known as the IBM FAStT FC2, it does not mean that it is exclusively used for attachment to a FAStT. Note also that the device driver supports all IBM HBAs labelled with FAStT (FC1 and FC2).

The use of the IBM FC-2 adapter and the configuration described here after apply whether you will attach to FAStT or ESS. The only difference is that the ESS requires the single path driver for the HBA.

5.6.1 Install the adapters

If the IBM FC2 HBAs are not already installed, shut down your system and install the adapters. At this stage it is best to either leave the cards disconnected from the SAN, or have the SAN switched off. Otherwise, if you boot your server to the OS, it is now possible that the sequence of the SCSI devices could change, in which case the boot would fail. This situation will be corrected later in this chapter.

Red Hat system

If you are running an “out-of-the-box” installation of Red Hat Enterprise Linux, then KUDZU probes for new hardware during boot. All detected hardware is compared with entries in the KUDZU database at /etc/sysconfig/hwconfig. If this file does not exist for any reason, then /etc/modules.conf, etc/sysconfig/network-scripts, and /etc/X11/XF86Config are scanned instead. If new hardware is detected, a dialog offers three choices:

Configure: Include the adapter(s) into your configuration.

Ignore: Ignore the new hardware in the future.

Do Nothing: Do nothing, continue to boot.

Select **Configure**: A driver will be installed for the hardware.

The standard driver delivered with the distribution is not capable of multipathing. If you attach to ESS, you can use this single path driver. If you attach to FAStT, you need to replace it with an available multipath driver.

SuSE system

SuSE ignores the adapters, and you can simply proceed with the installation of the new device drivers.

5.6.2 Prepare the Fibre Channel Host Bus Adapter cards

Obtain the latest IBM FAStT BIOS update file:

`http://www.pc.ibm.com/support/`

Create the bootable diskette and use it to boot the server. You should end up with a DOS `C:\` prompt. The update will do all like HBAs in the server at the same time. (Dissimilar HBAs need to be updated separately.)

- **For xSeries**

  Type the following to update the HBA BIOS:

  ```
  flasutil /f /l
  ```

  Type the following to set the defaults:

  ```
  flasutil /u
  ```

  Remove the boot diskette and restart the server.

- **BladeCenter**

  The `/i` option is required for the BladeCenter to identify the vendor. Type the following to update the HBA BIOS:

  ```
  flasutil /f /l /i
  ```

  Type the following to set the defaults:

  ```
  flasutil /u /i
  ```

  Remove the boot diskette and restart the server.

Settings need to be changed in the HBA configuration for use with Linux. These settings should be changed *after* the BIOS update. Watch for the IBM FAStT BIOS and press Ctrl-Q when prompted to enter the setup utility.

Select the adapter to change. Select **Host Adapter Settings.**

- **Loop reset delay** - Change this setting to 8.
- **Adapter Port Name** - Note this number as it will be required later for storage partitioning.

Select the advanced adapter settings:

- **LUNs per target** - This should be 0 (zero)
- **Enable target reset** - This should be yes
- **Port down retry count** - Change this to 12

Repeat this process for any other HBAs in the server, save the settings, and exit to restart the server.

5.6.3 Installing the Fibre Channel Host Bus Adapter driver

Obtain the latest IBM FAStT HBA driver from IBM. The BladeCenter driver will be a separate package from the xSeries driver.

As mentioned earlier, two different packages exist for the IBM FAStT HBA family:

- A fail-over (or multipath) device driver for FAStT attachment
- A singlepath driver for ESS attachment

The procedure to install the sources is identical for both drivers.

Copy the `i2xLNX-<driver version>-fo-dist.tgz` file to a folder on your system and type the following to extract the files:

```
tar zxvf *.tgz
```

This also creates an `./i2x00-<driver version>` directory.

Change to this directory and run the **drvrsetup** script to extract the source:

```
cd i2x00-<version>
sh drvrsetup
```

Table 5-1 shows the contents of the driver archive folder.

Table 5-1 Contents of the driver archive

drvrsetup	Script file to copy driver source files included in the driver source tgz file
i2xLNXsrc-<driver version>.tgz	driver source archive.
libinstall	Script file to install/setup HBA API library.

libremove	Script file to remove HBA API library.
qlapi-<api lib version>-rel.tgz	Compressed binary distribution file for API library.
ipdrvrsetup	Script file to copy IP driver source files included in the IP driver source tgz file
qla2xipsrc-<IP driver version>.tgz	Compressed binary distribution file for IP driver sources

For Red Hat, the drivers require that the source headers be prepared before compilation. Change to the kernel source directory:

```
cd /usr/src/linux2.4
```

Change the `Extraversion` field in the `Makefile` to correspond with your kernel version. The default setting with Red Hat Enterprise Linux is `custom`.

Run **`make menuconfig`** to load the Linux Kernel Configurator tool.

For SuSE you may need to run **`make cloneconfig`** to pull in the current kernel configuration prior to running **`make menuconfig`**. Check the processor setting is correct (SMP, cyclone support for x440 (not required for x360, not necessary but suggested for x440 with one CEC; this features introduces a generalized time to all CECs), exit and save the configuration. Run **`make dep`** to rebuild the kernel dependencies.

The fibre HBA driver source can now be compiled. Change back to the directory containing the driver source. The following commands will generate `qla2200.o` and `qla2300.o` modules:

For a uni-processor system type:

```
make all
```

For an SMP system type:

```
make all SMP=1
```

For SuSE systems add `OSVER=linux`, for example:

```
make all SMP=1 OSVER=linux
```

The BladeCenter Fibre Channel Expansion cards use the `qla2300.o` module. After removing the existing drivers, copy the required driver module to the following location:

```
/lib/modules/’uname -r’/kernel/drivers/addon/qla2200/
```

5.6.4 Loading the FAStT HBA driver

The following sections explain how to load the FAStT HBA driver for Red Hat and SuSE respectively.

Red Hat

Edit /etc/modules.conf to ensure that the IBM FAStT modules are loaded after any local SCSI module. Note in the example that the ServeRAID module, ips, is loaded before the IBM FAStT modules. Your modules.conf file should look similar to the following:

```
alias eth0 bcm5700
alias scsi_hostadapter ips
alias scsi_hostadapter2 qla2300
alias scsi_hostadapter3 qla2300
```

The following line should also be added in order to support more than one SCSI device per adapter (Red Hat Enterprise Linux supports up to 40 LUNs by default):

```
options scsi_mod max_scsi_luns=128
```

Save modules.conf and exit. Run **depmod -a** to update the modules.dep file.

You can load the fibre HBA module manually by typing the following:

```
modprobe qla2300
```

Alternatively, you can have it loaded at boot time by having it load as part of the ramdisk image. To build a new ramdisk image type the following:

```
mkinitrd /boot/<newname>.img <kernel version>
```

In our case, we typed the following:

```
mkinitrd /boot/initrd-storage-2.4.9-e.12summit.img 2.4.9-e.12summit
```

Edit /etc/grub.conf so that the boot loader knows to use the new ramdisk image. Edit the relevant kernel entry to load to new ramdisk image. At this point, you can also edit the default entry to load your new kernel and ramdisk image as default. It would be advisable to test that it boots successfully before changing the default.

Shut down the server and connect the SAN or bring the SAN online. Once online, bring the server back up. (As an alternative to bringing the server down you could unload the IBM FAStT module, (**modprobe -r qla2300**) attach the SAN, and reload the IBM FAStT module (**modprobe qla2300**).)

If you will attach to FAStT, you may proceed to 5.8, “Configuring FAStT” on page 137.

SuSE

Edit `/etc/sysconfig/kernel` and add qla2x00 to the string behind INITRD_MODULES where x represents either 2 or 3:

```
INITRD_MODULES=”aic7xxx reiserfs qla2300” .
```

Run **`depmod -a`** next.

The script **`/sbin/mk_initrd`** uses this string as input. By launching **`mk_initrd`** , a new version of the initial ramdisk is built in the /boot folder. Even though this is convenient, you might want to built a new ramdisk rather than replacing the existing one.

Before you restart you should add the kernel parameter max_scsi_luns=128 to the `grub.conf` or `lilo.conf` file. This is necessary because the kernel does not allow the full range support of LUNs (most desktop versions support no LUN out of the box). If you are using LILO run **`/sbin/lilo`** to commit the changes.

You can load the fibre HBA module manually by typing the following:

```
modprobe qla2300
```

Shut down the server and connect the SAN or bring the SAN online. Once online bring the server back up. (As an alternative to bringing the server down, you could unload the IBM FAStT module, (**`modprobe -r qla2300`**) attach the SAN, and reload the IBM FAStT module (**`modprobe qla2300`**)).

If you will attach to FAStT, you may proceed to 5.8, “Configuring FAStT” on page 137.

5.6.5 Attaching to ESS

Linux distributors do not offer a multipath driver that can be used when attaching to ESS as is the case with FAStT. However, one can achieve redundant pathing and load balancing by using a multipath extension called Storage Device Driver (SDD), available from IBM. The particularity and advantage of SDD is that it operates *independently* of the HBA device driver (SDD is a pseudo device driver). Therefore, the following applies for any HBA compatible with the ESS.

The SDD introduces a new tree in the `/dev` file system with the name `/vpath`. This is one reason why SDD does not support LVM. In LVM the supported components of the `/dev` directory are hard coded.

SDD actually consists of a driver and a server daemon. The server daemon runs in the background at all times (it starts automatically after the SDD driver package is installed). It scans all paths at regular intervals for an indication of failure. A path can have the states OPEN, CLOSE DEAD, DEAD, and INVALID. The SDD server reclaims paths and changes their state according to Table 5-2.

Table 5-2 SDD path probing and reclaiming

Path status before probe	Path status while probing	New path status
INVALID, CLOSED DEAD	OPEN	OPEN
CLOSED DEAD	CLOSE	CLOSE
DEAD	OPEN	OPEN
OPEN	not operational	DEAD
CLOSE	not working	CLOSE DEAD

Installing the SDD driver

Before you install SDD, please make sure that you meet all the hardware and software requirements. For up-to-date information, consult:

http://www.storage.ibm.com/disk/ess/supserver.htm

Note: At the time of writing this redbook SDD was still under testing for the Linux kernel we used.

Before installing SDD, you must configure the Fibre Channel adapters and the adapter drivers that are attached to the Linux host system, as explained in 5.6.2, “Prepare the Fibre Channel Host Bus Adapter cards” on page 126, and 5.6.3, “Installing the Fibre Channel Host Bus Adapter driver” on page 127.

In addition, before you install SDD, you must configure your ESS for multiport access for each LUN (use the ESS Specialist for that purpose). SDD requires a minimum of two independent paths that share the same LUN to use the load-balancing and path failover protection features.

Strictly speaking, note that a host system with a single Fibre Channel adapter connected through a switch to multiple ESS ports is considered a multipath Fibre Channel connection, but is not sufficient (SDD will not provide failover capabilities even if you install it in this case).

To install SDD, type:

```
rpm -iv IBMsdd-N.N.N.N -1.i686.redhat.rpm
rpm -iv IBMsdd-N.N.N.N -1.i686.suse.rpm
```

(where N.N.N.N represents the current version release modification level number; N.N.N.N = 1.3.0.1 for example).

The installation copies the driver and its utilities in the/opt/IBMssd directory. Table 5-3 lists all components installed and their location.

Table 5-3 SDD components for a Linux host system

File	Directory	Description
sdd-mod.o-xxxy	/opt/IBMsdd	SDD device driver file (where XXX stands for the kernel level of your host system and y represents smp or up
vpath.conf	/etc	
executables	/opt/IBMsdd/bin /usr/sbin	SDD configuration and status tools Symbolic links to the SDD utilities
sdd.rcscript	/etc/init.d/sdd /usr/sbin/sdd	Symbolic link for the SDD system startup option Symbolic link for the SDD manual start or restart option

Configuring SDD

You can manually or automatically load and configure SDD on your Linux host system. Manual configuration requires that you use a set of SDD-specific commands, while automatic configuration requires a system restart.

Table 5-4 CLI for SDD

Command	Description
cfgvpath	Configures vpath devices
cfgvpath query	Displays all sd devices
lsvpcfg	Displays the current devices that are configured and their corresponding paths
rmvpath	Removes one or all vpath devices
sdd start	Loads the SDD driver and automatically configures disk devices for multipath access.
sdd stop	Unloads the SDD driver (requires no vpath devices currently in use)

Command	Description
sdd restart	Unloads the SDD driver (requires no vpath devices currently in use), and then loads the SDD driver and automatically configures disk devices for multipath access.

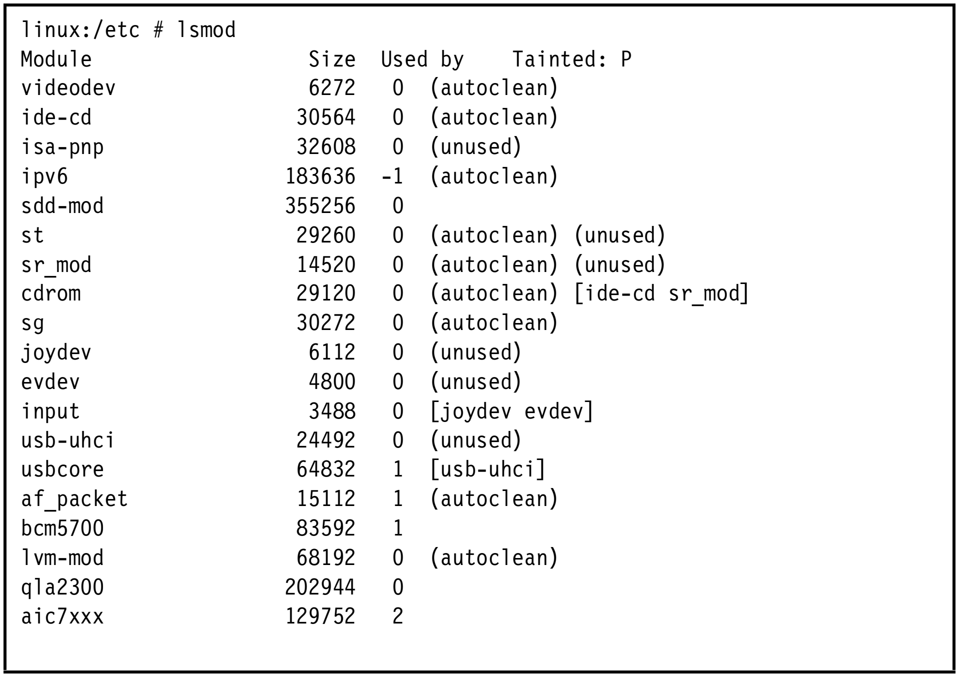
Using multiple commands for SDD configuration

To load and configure the SDD on your system, manually change to the directory `/opt/IBMsdd` and enter the command **`insmod`** ***`modulename`*** (in our case, and the module name was *sdd-mod.o-2.4.19.SuSE-152.k_smp-2.4.19-233* for SuSE SLES 8).

If you do not know your exact kernel version, use the command **`uname -a`** to list it.

To verify whether the SDD sdd-mod driver is loaded type **`cat /proc/modules`**.

To verify that the driver is loaded correctly please enter **`lsmod`**. Figure 5-20 shows an example of the expected output.



```
linux:/etc # lsmod
Module                  Size  Used by    Tainted: P
videodev                6272   0  (autoclean)
ide-cd                 30564   0  (autoclean)
isa-pnp                32608   0  (unused)
ipv6                  183636  -1  (autoclean)
sdd-mod               355256   0
st                     29260   0  (autoclean) (unused)
sr_mod                 14520   0  (autoclean) (unused)
cdrom                  29120   0  (autoclean) [ide-cd sr_mod]
sg                     30272   0  (autoclean)
joydev                  6112   0  (unused)
evdev                   4800   0  (unused)
input                   3488   0  [joydev evdev]
usb-uhci               24492   0  (unused)
usbcore                64832   1  [usb-uhci]
af_packet              15112   1  (autoclean)
bcm5700                83592   1
lvm-mod                68192   0  (autoclean)
qla2300               202944   0
aic7xxx               129752   2
```

Figure 5-20 Loaded SDD module

Change to the /opt/IBMsdd/bin directory and enter **cfgvpath query** as shown in Figure 5-21.

```
linux:/opt/IBMsdd/bin # cfgvpath query
/dev/sda  (  8,   0) host=1  ch=0  id=13 lun=0   vid=IBM-ESXS pid=ST318452LC
serial=xxxxxxxxxxxx                      ctlr_flag=0 ctlr_nbr=0 df_ctlr=0 X
/dev/sdb  (  8,  16) host=2  ch=0  id=0  lun=0   vid=IBM      pid=2105800
serial=70024678                          ctlr_flag=0 ctlr_nbr=0 df_ctlr=0
/dev/sdc  (  8,  32) host=3  ch=0  id=0  lun=0   vid=IBM      pid=2105800
serial=70024678                          ctlr_flag=0 ctlr_nbr=0 df_ctlr=0
```

Figure 5-21   Output of cfgvpath

The sample output shows the name and serial number of the SCSI disk device, its connection information, and its product identification. A capital letter X at the end of a line indicates that SDD currently does not support the device.

Type the **cfgvpath** as shown in Figure 5-22 to configure the SDD vpath devices. This merges the SCSI devices sdb and sdc into one virtual device, vpatha (actually, the effect is to make sdb and sdc dress the same target).

```
linux:/opt/IBMsdd/bin # cfgvpath
crw-r--r--    1 root     root     253,   0 Apr  4 06:03 /dev/IBMsdd
Making device file /dev/vpatha 247 0
Added vpatha ...
linux:/opt/IBMsdd/bin # ls /dev/vpath*
/dev/vpatha    /dev/vpatha11  /dev/vpatha14  /dev/vpatha3  /dev/vpatha6
/dev/vpatha9
/dev/vpatha1   /dev/vpatha12  /dev/vpatha15  /dev/vpatha4  /dev/vpatha7
/dev/vpatha10  /dev/vpatha13  /dev/vpatha2   /dev/vpatha5  /dev/vpatha8
linux:/opt/IBMsdd/bin # cat /etc/vpath.conf
vpatha 70024678
```

Figure 5-22   Configure SDD vpath devices

The configuration information is saved by default in the /etc/vpath.conf file to maintain vpath name persistence in subsequent driver loads and configurations.

You can remove an SDD vpath device by using the **rmvpath xxx** command, where **xxx** represents the name of the vpath device that is selected for removal.Type **cfgvpath ?** or **rmvpath ?** for more information about the **cfgvpath** or **rmvpath** commands.

To verify the vpath configuration enter **lsvpcfg** or **datapath query device**. If you successfully configured SDD vpath devices, output similar to the following is displayed by lsvpcfg:

```
000 vpath0 (247,0)70024678 =/dev/sdb /dev/sdl /dev/sdcg /dev/sdcp.
```

In the following sections we describe three different modes of operation:

Using a single command for SDD configuration

You can also manually load and configure SDD by issuing the **sdd start** command. Successful execution of the **sdd start** command performs all the tasks described in "Using multiple commands for SDD configuration" on page 133.

Use the **sdd stop** command to unconfigure and unload the SDD driver. Use the **sdd restart** command to unconfigure, unload, and then restart the SDD configuration process.

Configuring SDD at system startup

You can set up SDD to automatically load and configure when your Linux system boots. SDD provides the startup script **sdd.rcscript** file in the /opt/IBMsdd/bin directory and creates a symbolic link to /etc/init.d/sdd.

The following steps are necessary to launch SDD at your Red Hat system startup:

1. Log on to your Linux host system as the root user.
2. Type **chkconfig --level X sdd** on to enable run level X at startup (where X represents the system run level).
3. Type **chkconfig --list sdd** to verify that the system startup option is enabled for SDD configuration.
4. Restart your host system so that SDD is loaded and configured.
5. If necessary, you can disable the startup option by typing **chkconfig --level X sdd** off.

The following steps are necessary to launch SDD at your SuSE system startup:

1. Log on to your Linux host system as the root user.
2. Type **insserv sdd**
3. Restart your host system so that SDD is loaded and configured.
4. If necessary, you can disable the startup option by typing **insserv -r sdd**

In order for SDD to automatically load and configure, the HBA driver must already be loaded. This can be assured at start time by adding the appropriate driver(s) to the kernel's initial ramdisk (see 5.8 "Configuring FAStT" on page 137 for details to set up the initial ramdisk).

Partitioning SDD vpath devices

Disk partitions are known as logical devices. SDD for Linux allows configuration of whole devices only. The SDD naming scheme for disks and disk partitions follows the standard Linux disk naming convention. The following description illustrates the naming scheme for SCSI disks and disk partitions:

- The first two letters indicate the SCSI device.
- The next letter (or two letters), a-z, specifies the unique device name.
- A number following the device name denotes the partition number. For example, /dev/sda is the whole device, while /dev/sda1 is a logical device representing the first partition of the whole device /dev/sda. Each device and partition has its own major and minor number.

Similarly then, a specific device file /dev/vpathX is created for each supported multipath SCSI disk device (where `X` represents the unique device name; as with sd devices, `X` may be one or two letters).

Device files /dev/vpathXY are also created for each partition of the multipath device (where `Y` represents the corresponding partition number). When a file system or user application wants to use the logical device, it should refer to /dev/vpathXY (for example, /dev/vpatha1 or /dev/vpathbc7) as its multipath logical device. All I/O management, statistics, and failover processes of the logical device follow those of the whole device.

The output in Figure 5-22 on page 134 demonstrates how the partitions are named.

Note:

- SDD does not support the use of root (/), /var, /usr, /opt, /tmp and swap partitions.
- For supported file systems, use the standard UNIX fdisk command to partition vpath devices (e.g. fdisk /dev/vpatha).

To partition the new device please refer to Appendix A, "Storage from the OS view" on page 253. Please keep in mind that you have to start fdisk with the proper vpath as parameter (e.g. **`fdisk /dev/vpatha)`**.

5.7 BladeCenter Fibre Switch Module configuration

The BladeCenter SAN Utility application is used to access and configure the BladeCenter Fibre Channel switch modules. The SAN Utility can be installed on a BladeCenter HS20 Blade server or an external network management workstation configured with a supported version of Linux.

Before you configure your switch module, be sure that the management modules are properly configured. In order to access your switch module from an external environment, you may need to enable certain features, such as external ports and external management over all ports. See the applicable *BladeCenter Unit Installation and User's Guide* publications on the IBM eServer BladeCenter Documentation CD.

Installing the BladeCenter SAN Utility is detailed in Chapter 8. "BladeCenter SAN Utility" on page 177.

5.8 Configuring FAStT

It is necessary to set up the FAStT and configure LUNs to be accessed by the Linux system. The configuration of the FAStT is done using the Storage Manager tool and the FAStT MSJ (Management Suite for Java). Please refer to Chapter 9. "FAStT Storage Manager" on page 187, and Chapter 10. "FAStT MSJ (Management Suite for Java)" on page 211 on how to install and use these.

5.9 Configuring ESS

It is necessary to set up the ESS and configure LUNs to be accessed by the Linux System. In addition, as mentioned earlier, you must configure your ESS for multiport access for each LUN before you install SDD. This is done using the Storage Specialist. Please refer to Chapter 7., "Configuring ESS for Linux" on page 155.

6

Red Hat Cluster Manager

This chapter describes the preparation, installation, and customization tasks required for the Red Hat High Availability Cluster Manager.

For details, refer to the *Red Hat Cluster Manager Installation and Administration Guide*, which can be found at:

```
http://www.redhat.com/docs/manuals/enterprise/
```

Cluster Manager is supplied with Red Hat Enterprise Linux AS. It allows you to set up a high availability cluster containing two nodes and using shared disk storage. Future versions will support multiple nodes.

> **Restriction:** IBM does not currently support high availability clustering products under Linux with xSeries servers. These applications are supported by the vendor.

6.1 Preparing to install Cluster Manager

High Availability clusters should ideally have no single point-of-failure. This reliability provides data integrity and application availability in the event of a failure. Redundant hardware with shared storage along with cluster and application failover mechanisms mean a high availability cluster can meet the needs of the enterprise market.

Figure 6-1 shows the setup done for our experiment.

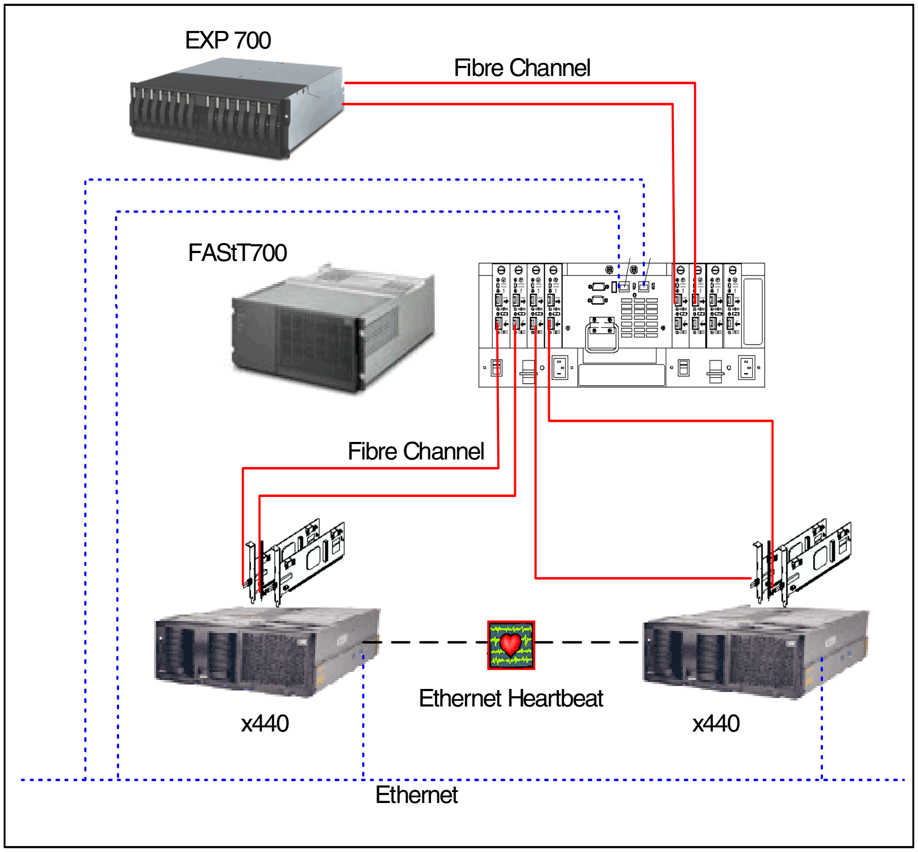


Figure 6-1 Configuration of HA cluster

6.1.1 Quorum partitions

It is important to note that the Quorum partitions require raw devices. Cluster Manager requires a Quorum partition and backup Quorum partition to store cluster state and configuration information. These should ideally be created on a logical drive on a RAID1 array. Create two partition no smaller than 10 MB but no larger than 20 MB, there is no need for it to be larger than this. Do not create file systems on the Quorum partitions, as these need to remain raw.

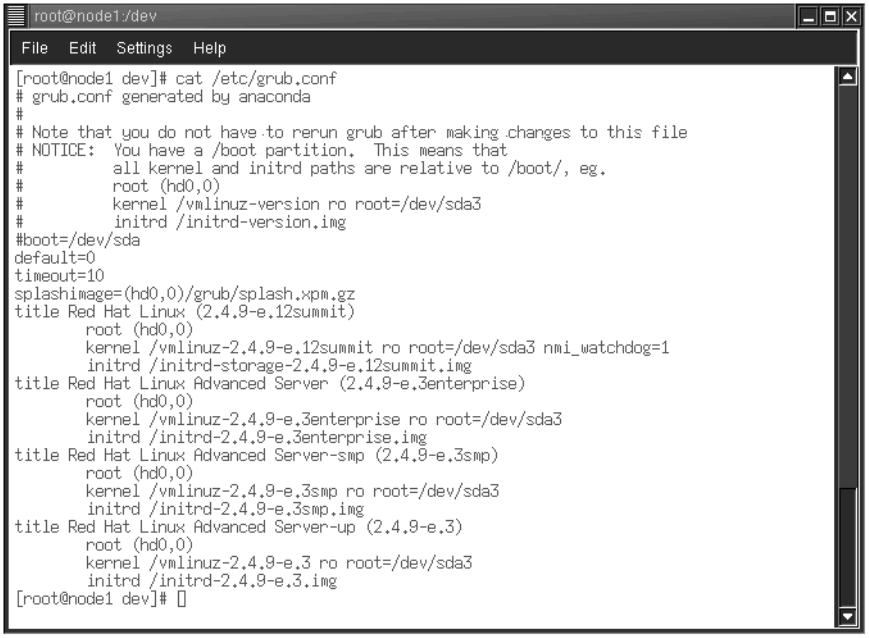
6.1.2 Software watchdog

A software watchdog is used to kill a hung node if no hardware devices such as power switches are being used for STONITH (shoot the other node in the head). The cluster quorum daemon (cluquorumd) will periodically reset the timer interval. If the daemon fails to reset the timer, the failed cluster member will reboot itself. It is necessary to create the device special for the watchdog timer. Carry out the following on both nodes of the cluster:

```
cd /dev
./MAKEDEV watchdog
```

6.1.3 NMI watchdog

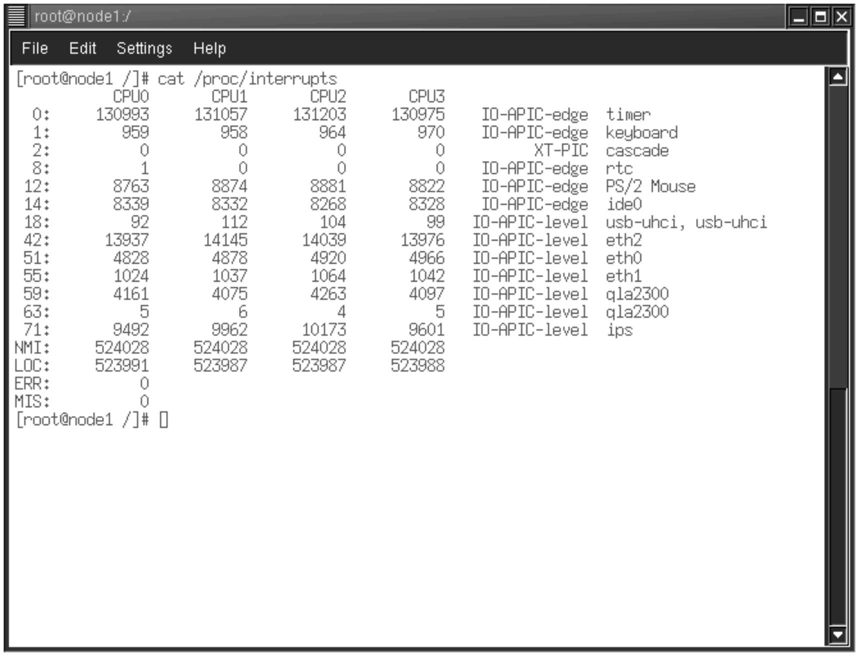
When using the software watchdog, it is recommended that you also use the NMI watchdog timer. The NMI watchdog timer may not work on some legacy systems. Edit `/etc/grub.conf` on both systems and add `nmi_watchdog=1` to the end of the relevant kernel line (Figure 6-2).



```
[root@node1 dev]# cat /etc/grub.conf
# grub.conf generated by anaconda
#
# Note that you do not have to rerun grub after making changes to this file
# NOTICE:  You have a /boot partition.  This means that
#          all kernel and initrd paths are relative to /boot/, eg.
#          root (hd0,0)
#          kernel /vmlinuz-version ro root=/dev/sda3
#          initrd /initrd-version.img
#boot=/dev/sda
default=0
timeout=10
splashimage=(hd0,0)/grub/splash.xpm.gz
title Red Hat Linux (2.4.9-e.12summit)
        root (hd0,0)
        kernel /vmlinuz-2.4.9-e.12summit ro root=/dev/sda3 nmi_watchdog=1
        initrd /initrd-storage-2.4.9-e.12summit.img
title Red Hat Linux Advanced Server (2.4.9-e.3enterprise)
        root (hd0,0)
        kernel /vmlinuz-2.4.9-e.3enterprise ro root=/dev/sda3
        initrd /initrd-2.4.9-e.3enterprise.img
title Red Hat Linux Advanced Server-smp (2.4.9-e.3smp)
        root (hd0,0)
        kernel /vmlinuz-2.4.9-e.3smp ro root=/dev/sda3
        initrd /initrd-2.4.9-e.3smp.img
title Red Hat Linux Advanced Server-up (2.4.9-e.3)
        root (hd0,0)
        kernel /vmlinuz-2.4.9-e.3 ro root=/dev/sda3
        initrd /initrd-2.4.9-e.3.img
[root@node1 dev]# 
```

Figure 6-2 grub.conf

The system needs to be rebooted in order to check that the system supports the NMI watchdog. Once the server is back online check /proc/interrupts (Figure 6-3). If the NMI entry is not zero, then the system supports the NMI watchdog timer. If the entry is zero then try changing the append in grub.conf to nmi_watchdog=2 and reboot. If it is still zero then the server does not support the NMI watchdog timer.



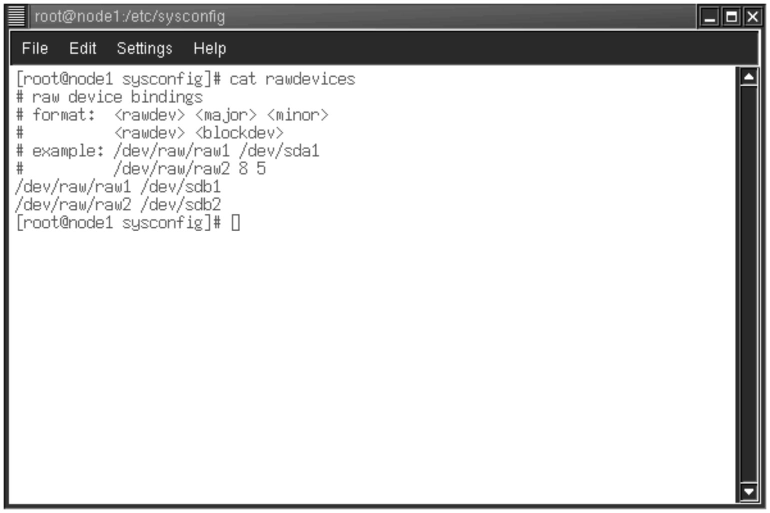
```
[root@node1 /]# cat /proc/interrupts
           CPU0       CPU1       CPU2       CPU3
  0:     130993     131057     131203     130975    IO-APIC-edge  timer
  1:        959        958        964        970    IO-APIC-edge  keyboard
  2:          0          0          0          0          XT-PIC  cascade
  8:          1          0          0          0    IO-APIC-edge  rtc
 12:       8763       8874       8881       8822    IO-APIC-edge  PS/2 Mouse
 14:       8339       8332       8268       8328    IO-APIC-edge  ide0
 18:         92        112        104         99   IO-APIC-level  usb-uhci, usb-uhci
 42:      13937      14145      14039      13976   IO-APIC-level  eth2
 51:       4828       4878       4920       4966   IO-APIC-level  eth0
 55:       1024       1037       1064       1042   IO-APIC-level  eth1
 59:       4161       4075       4263       4097   IO-APIC-level  qla2300
 63:          5          6          4          5   IO-APIC-level  qla2300
 71:       9492       9962      10173       9601   IO-APIC-level  ips
NMI:     524028     524028     524028     524028
LOC:     523991     523987     523987     523988
ERR:          0
MIS:          0
[root@node1 /]# []
```

Figure 6-3 /proc/interrupts

6.1.4 Rawdevices

The /etc/sysconfig/rawdevices file has to be edited on both systems for use by the primary and backup Quorum partitions. Add the following lines to the file where X = device letter (Figure 6-4):

```
/dev/raw/raw1 /dev/sdX1
/dev/raw/raw2 /dev/sdX2
```



```
[root@node1 sysconfig]# cat rawdevices
# raw device bindings
# format:  <rawdev> <major> <minor>
#          <rawdev> <blockdev>
# example: /dev/raw/raw1 /dev/sda1
#          /dev/raw/raw2 8 5
/dev/raw/raw1 /dev/sdb1
/dev/raw/raw2 /dev/sdb2
[root@node1 sysconfig]# []
```

Figure 6-4 /etc/sysconfig/rawdevices

Save the changes and restart the rawdevices service:

```
service rawdevices restart
```

6.1.5 Log file for Cluster Messages

The /etc/sysconf.log file can be edited to log events from the cluster to a different file away from the default log file. The cluster utilities and daemons log their messages to a syslog tag called local4.

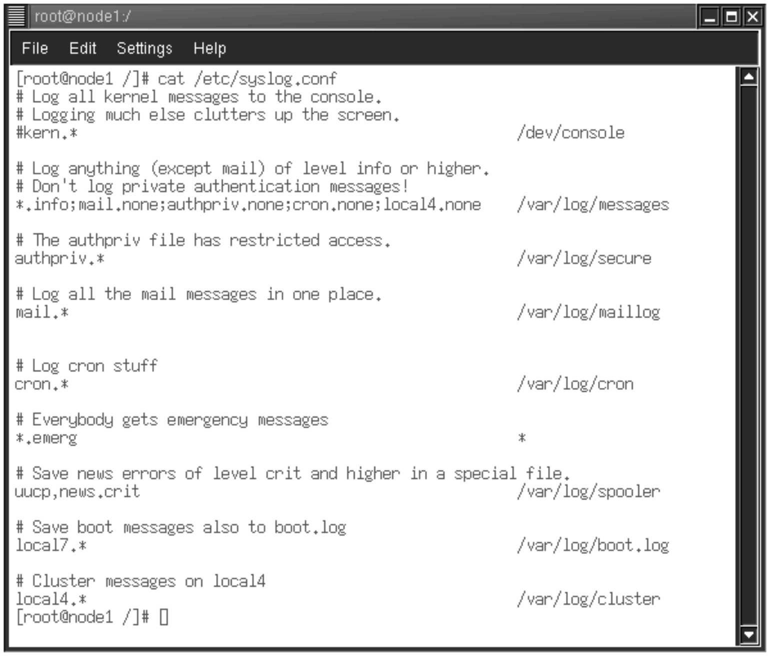
Edit /etc/syslog.conf on both systems and add local4.none to the following section to avoid events being duplicated in the cluster log and the default log:

```
# Log anything (except mail) of level info or higher.
# Don't log private authentication messages!
*.info;mail.none;authpriv.none;cron.none;local4.none    /var/log/messages
```

Add the following at the end of the file to log the cluster messages into a separate file:

```
# Cluster messages on local4
local4.*                                                /var/log/cluster
```

The /etc/syslog.conf file should look similar to Figure 6-5.



```
[root@node1 /]# cat /etc/syslog.conf
# Log all kernel messages to the console.
# Logging much else clutters up the screen.
#kern.*                                                 /dev/console

# Log anything (except mail) of level info or higher.
# Don't log private authentication messages!
*.info;mail.none;authpriv.none;cron.none;local4.none    /var/log/messages

# The authpriv file has restricted access.
authpriv.*                                              /var/log/secure

# Log all the mail messages in one place.
mail.*                                                  /var/log/maillog


# Log cron stuff
cron.*                                                  /var/log/cron

# Everybody gets emergency messages
*.emerg                                                 *

# Save news errors of level crit and higher in a special file.
uucp,news.crit                                          /var/log/spooler

# Save boot messages also to boot.log
local7.*                                                /var/log/boot.log

# Cluster messages on local4
local4.*                                                /var/log/cluster
[root@node1 /]# ▯
```

Figure 6-5 /etc/syslog.conf

Restart the syslog service:

```
service syslog restart
```

6.2 Installing Cluster Manager

This section takes you through the installation and basic setup of Cluster Manager.

6.2.1 Installation

It is possible that Cluster Manager was installed during the installation of Red Hat Enterprise Linux AS. Uninstall this version on both systems and obtain the latest version of `clumanager`. Type the following to uninstall:

```
rpm -e clumanager
```

Type the following to install the latest version:

```
rpm -ivh clumanager-<version>.rpm
```

6.2.2 Basic setup of Cluster Manager

We can now configure and setup the cluster software.

Type the following on only one of the cluster nodes:

```
/sbin/cluconfig
```

The execution of the script takes place triggering a number of prompts for information such as an IP alias for the cluster, heartbeat details, Quorum partitions, any watchdogs used. In most instances, the default answers in `[square brackets]` can be accepted.

In our setup we use two Ethernet cards per system as heartbeats. A serial interface can also be used for heartbeat if available on the system. Where network heartbeats are used, `/etc/hosts`, should be updated to include the heartbeat IP Addresses and host names.

The following is an example of the script and the answers we entered.

Example 6-1 Cluster setup

```
root@node1 rhcm]# /sbin/cluconfig

Red Hat Cluster Manager Configuration Utility (running on node1)

Enter cluster name [Red Hat Cluster Manager]: SG246261-Cluster
Enter IP address for cluster alias [NONE]: 192.168.1.30
--------------------------------
Information for Cluster Member 0
--------------------------------
Enter name of cluster member [node1]:
Looking for host node1 (may take a few seconds)...

Enter number of heartbeat channels (minimum = 1) [1]: 2
Information about Channel 0
Channel type: net or serial [net]:
Enter hostname of the cluster member on heartbeat channel 0 [node1]: hb1node1
Looking for host hb1node1 (may take a few seconds)...
Information about Channel 1
Channel type: net or serial [net]:
Enter hostname of the cluster member on heartbeat channel 1: hb2node1
Looking for host hb2node1 (may take a few seconds)...

Information about Quorum Partitions
Enter Primary Quorum Partition [/dev/raw/raw1]:
Enter Shadow Quorum Partition [/dev/raw/raw2]:
```

```
Information About the Power Switch That Power Cycles Member 'node1'
Choose one of the following power switches:
  o NONE
  o RPS10
  o BAYTECH
  o APCSERIAL
  o APCMASTER
  o WTI_NPS
  o SW_WATCHDOG
Power switch [NONE]: SW_WATCHDOG

--------------------------------
Information for Cluster Member 1
--------------------------------
Enter name of cluster member: node2
Looking for host node2 (may take a few seconds)...

Information about Channel 0
Enter hostname of the cluster member on heartbeat channel 0 [node2]: hb1node2
Looking for host hb1node2 (may take a few seconds)...
Information about Channel 1
Enter hostname of the cluster member on heartbeat channel 1: hb2node2
Looking for host hb2node2 (may take a few seconds)...

Information about Quorum Partitions
Enter Primary Quorum Partition [/dev/raw/raw1]:
Enter Shadow Quorum Partition [/dev/raw/raw2]:

Information About the Power Switch That Power Cycles Member 'node2'
Choose one of the following power switches:
  o NONE
  o RPS10
  o BAYTECH
  o APCSERIAL
  o APCMASTER
  o WTI_NPS
  o SW_WATCHDOG
Power switch [sw_watchdog]:

Cluster name: SG246261-Cluster
Cluster alias IP address: 192.168.1.30

--------------------
Member 0 Information
--------------------
Name: node1
Primary quorum partition: /dev/raw/raw1
Shadow quorum partition: /dev/raw/raw2
Heartbeat channels: 2
```

```
Channel type: net, Name: hb1node1
Channel type: net, Name: hb2node1
Power switch IP address or hostname: node1
Identifier on power controller for member node1: unused
--------------------
Member 1 Information
--------------------
Name: node2
Primary quorum partition: /dev/raw/raw1
Shadow quorum partition: /dev/raw/raw2
Heartbeat channels: 2
Channel type: net, Name: hb1node2
Channel type: net, Name: hb2node2
Power switch IP address or hostname: node2
Identifier on power controller for member node2: unused

--------------------------
Power Switch 0 Information
--------------------------
Power switch IP address or hostname: node1
Type: sw_watchdog
Login or port: unused
Password: unused
--------------------------
Power Switch 1 Information
--------------------------
Power switch IP address or hostname: node2
Type: sw_watchdog
Login or port: unused
Password: unused

Save the cluster member information? yes/no [yes]:
Writing to configuration file...done
Configuration information has been saved to /etc/cluster.conf.
----------------------------
Setting up Quorum Partitions
----------------------------
Running cludiskutil -I to initialize the quorum partitions: done
Saving configuration information to quorum partitions: done
Do you wish to enable monitoring, both locally and remotely, via the Cluster
GUI
? yes/no [yes]:

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

Configuration on this member is complete.

To configure the next member, invoke the following command on that system:
```

```
# /sbin/cluconfig --init=/dev/raw/raw1

Refer to the Red Hat Cluster Manager Installation and Administration Guide
for details.

[root@node1 rhcm]#
```

The basic setup is complete on the first node. Type the following on the second node:

```
/sbin/cluconfig --init=/dev/raw/raw1
```

The script looks for the required information on the Quorum partitions and pulls it in. You are then be prompted if you wish to save the configuration. Type `yes` and press Enter.

The cluster service will start automatically when the system is booted. To start the service manually now type the following:

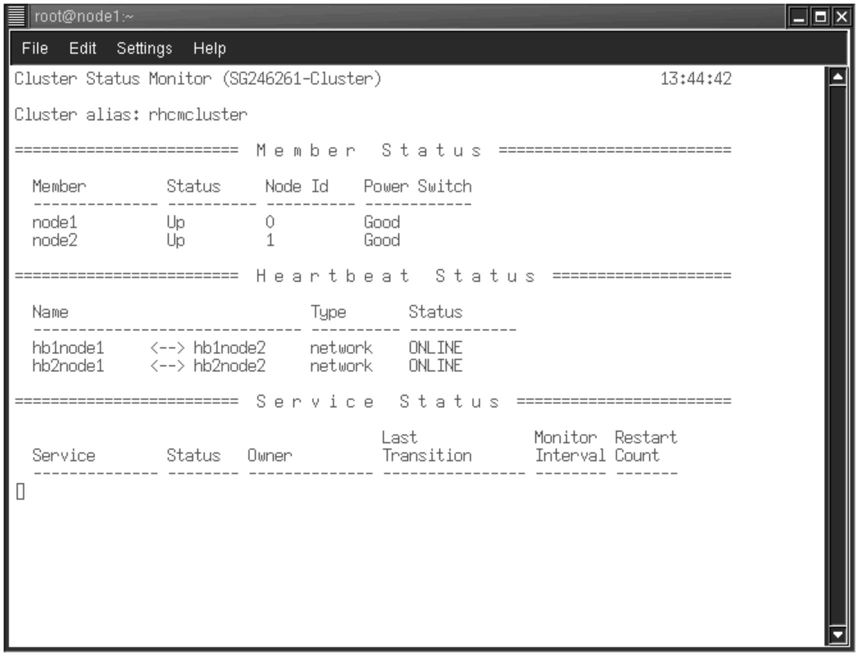
```
service cluster start
```

Type the following to check that the daemons are running:

```
service cluster status
```

Type the following to view the cluster node, heartbeat, and service status, where `-i` is the update interval in seconds (Figure 6-6).

```
clustat -i 2
```



```
Cluster Status Monitor (SG246261-Cluster)                        13:44:42

Cluster alias: rhcmcluster

========================  M e m b e r   S t a t u s  ========================

  Member         Status     Node Id    Power Switch
  -------------- ---------- ---------- ------------
  node1          Up         0          Good
  node2          Up         1          Good

========================  H e a r t b e a t   S t a t u s  ===================

  Name                           Type       Status
  ------------------------------ ---------- ------------
  hb1node1     <--> hb1node2     network    ONLINE
  hb2node1     <--> hb2node2     network    ONLINE

========================  S e r v i c e   S t a t u s  =======================

                                         Last             Monitor  Restart
  Service        Status   Owner          Transition       Interval Count
  -------------- -------- -------------- ---------------- -------- -------
```

Figure 6-6 Cluster status

Use `cluadmin` to add, delete, edit, and monitor the cluster and services. See the *Red Hat Cluster Manager Installation and Administration Guide* for more details.

6.2.3 Setting up an NFS clustered share

This section goes through the steps of how to set up an NFS clustered share. This share remains available even if the owning server dies as the other node takes ownership of the share. See the *Red Hat Cluster Manager Installation and Administration Guide* for more details of setting up a service and which services are available. Red Hat Cluster Manager does not require *cluster-aware* software. A start-stop script can be created for almost any application you choose to make highly available. An example start-stop script is provided by Red Hat.

NFS and portmap should be started on startup of the system OS. Type the following on both nodes to enable this:

```
chkconfig --level 345 nfs on
chkconfig --level 345 portmap on
```

Start the `cluadmin` command line utility on one node only:

```
/sbin/cluadmin
```

You will drop into the `cluadmin` shell. Type the following to add a service:

```
service add
```

You will receive various prompts (Example 6-2) for answers from the script such as service name, preferred node, IP Address, disk device, mount point, exports, etc. You can assign a preferred node for the service to start on. You can also decide if you would like to relocate the service back to the preferred node in the event of a failover where the faulty node has come back online.

Services can be shared between the systems rather than having an active-passive setup where the passive system just sits there waiting for the other to fail. In an active-active setup you should ensure that there are enough system resources for a single system to take over in the event of a failure.

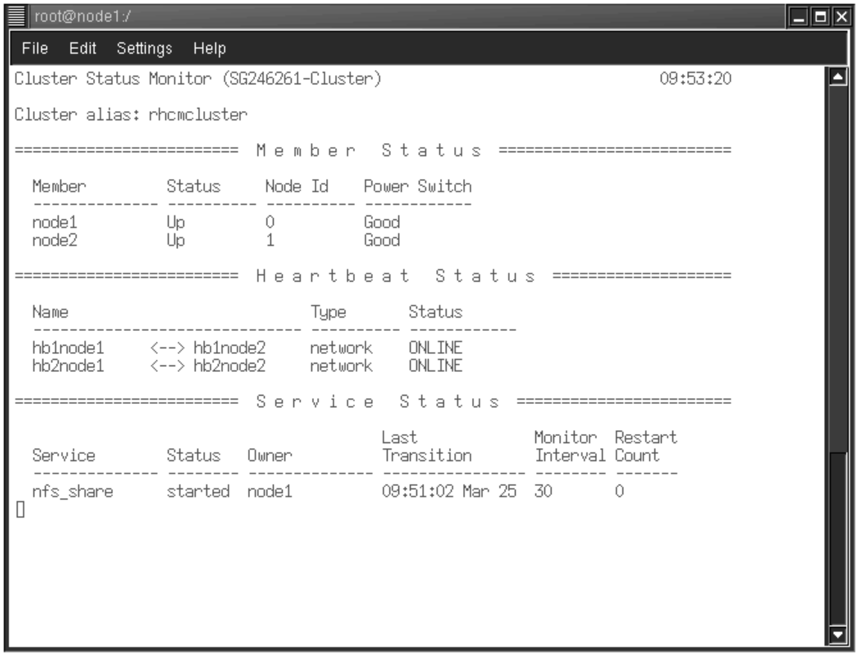
The following is an example of the script and the answers we entered.

Example 6-2 Adding an NFS service

```
root@node2 root]# /sbin/cluadmin
Tue Mar 25 09:30:27 GMT 2003

You can obtain help by entering help and one of the following commands:

cluster     service        clear
help        apropos        exit
version     quit
cluadmin> service add

  The user interface will prompt you for information about the service.
   Not all information is required for all services.

  Enter a question mark (?) at a prompt to obtain help.

  Enter a colon (:) and a single-character command at a prompt to do
  one of the following:

  c - Cancel and return to the top-level cluadmin command
  r - Restart to the initial prompt while keeping previous responses
  p - Proceed with the next prompt

Currently defined services:

Service name: nfs_share
Preferred member [None]: node1
Relocate when the preferred member joins the cluster (yes/no/?) [no]:
User script (e.g., /usr/foo/script or None) [None]:
Status check interval [0]: 30
```

```
Do you want to add an IP address to the service (yes/no/?) [no]: yes

   IP Address Information

IP address: 192.168.1.40
Netmask (e.g. 255.255.255.0 or None) [None]:
Broadcast (e.g. X.Y.Z.255 or None) [None]:
Do you want to (a)dd, (m)odify, (d)elete or (s)how an IP address, or are you
(f)inished adding IP addresses [f]:
Do you want to add a disk device to the service (yes/no/?) [no]: yes

Disk Device Information

Device special file (e.g., /dev/sdb4): /dev/sdc1
Filesystem type (e.g., ext2, or ext3): ext3
Mount point (e.g., /usr/mnt/service1) [None]: /mnt/nfs
Mount options (e.g., rw,nosuid,sync): rw,nosuid,sync
Forced unmount support (yes/no/?) [yes]:
Would you like to allow NFS access to this filesystem (yes/no/?)  [no]: yes

You will now be prompted for the NFS export configuration:

Export directory name [/mnt/nfs]:

Authorized NFS clients

Export client name [*]:
Export client options [None]: rw
Do you want to (a)dd, (m)odify, (d)elete or (s)how NFS CLIENTS, or are you
(f)inished adding CLIENTS [f]:
Do you want to (a)dd, (m)odify, (d)elete or (s)how NFS EXPORTS, or are you
(f)in
ished adding EXPORTS [f]:
Would you like to share to Windows clients (yes/no/?)  [no]:
Do you want to (a)dd, (m)odify, (d)elete or (s)how DEVICES, or are you
(f)inishe
d adding DEVICES [f]:
name: nfs_share
preferred node: node1
relocate: no
user script: None
monitor interval: 30
IP address 0: 192.168.1.40
  netmask 0: None
  broadcast 0: None
device 0: /dev/sdc1
  mount point, device 0: /mnt/nfs
  mount fstype, device 0: ext3
  mount options, device 0: rw,nosuid,sync
  force unmount, device 0: yes
  samba share, device 0: None
NFS export 0: /mnt/nfs
  Client 0: *, rw
Add nfs_share service as shown? (yes/no/?) yes

  0) node1   (preferred)
  1) node2
  c) cancel

Choose member to start service on: 0
Added nfs_share.
cluadmin>
```

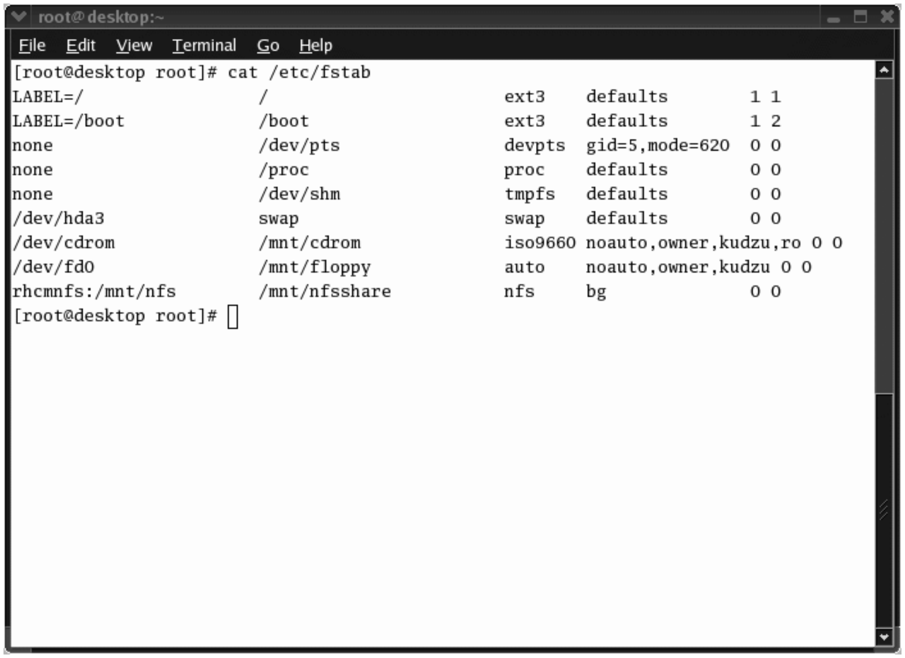
Once complete, the service should automatically start. If you run **`clustat -i 2`** you should see a screen similar to Figure 6-7 showing the service on the node you assigned it to start on.



```
Cluster Status Monitor (SG246261-Cluster)                          09:53:20

Cluster alias: rhcmcluster

========================  M e m b e r   S t a t u s  ========================

  Member         Status     Node Id    Power Switch
  -------------- ---------- ---------- ------------
  node1          Up         0          Good
  node2          Up         1          Good

========================  H e a r t b e a t   S t a t u s  ===================

  Name                           Type       Status
  ------------------------------ ---------- ------------
  hb1node1     <--> hb1node2     network    ONLINE
  hb2node1     <--> hb2node2     network    ONLINE

========================  S e r v i c e   S t a t u s  =======================

                                         Last             Monitor  Restart
  Service        Status   Owner          Transition       Interval Count
  -------------- -------- -------------- ---------------- -------- -------
  nfs_share      started  node1          09:51:02 Mar 25  30       0
```

Figure 6-7 Clustered NFS service added

To now use this service from a client system, you should create a folder for the NFS share to mount into. Add an entry into `/etc/hosts` if you added an IP address for the service. Edit /etc/fstab to mount the NFS share on startup. The entry should be similar to the last line of the following example (Figure 6-8).



```
[root@desktop root]# cat /etc/fstab
LABEL=/                 /                       ext3    defaults        1 1
LABEL=/boot             /boot                   ext3    defaults        1 2
none                    /dev/pts                devpts  gid=5,mode=620  0 0
none                    /proc                   proc    defaults        0 0
none                    /dev/shm                tmpfs   defaults        0 0
/dev/hda3               swap                    swap    defaults        0 0
/dev/cdrom              /mnt/cdrom              iso9660 noauto,owner,kudzu,ro 0 0
/dev/fd0                /mnt/floppy             auto    noauto,owner,kudzu 0 0
rhcmnfs:/mnt/nfs        /mnt/nfsshare           nfs     bg              0 0
[root@desktop root]#
```

Figure 6-8 Example of fstab

To mount the NFS share now without restarting the client system, type the following:

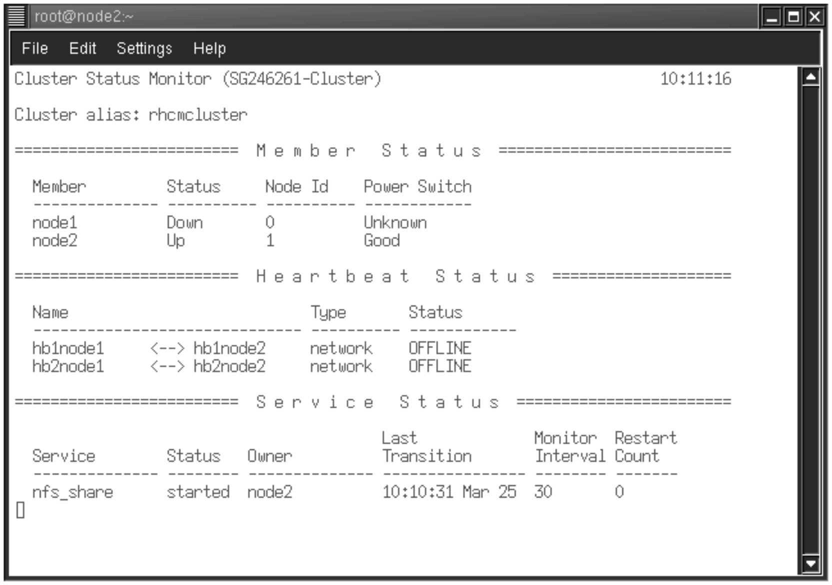
```
mount -a
```

To test the cluster, cause the node which owns the server to hang, crash, or reboot.

The software watchdog will kill the hung server and the healthy node will take over the NFS service. Meanwhile, there should be no, or very little, interruption to the NFS service. You could ping the IP address of the NFS service during failover to verify this.

Another good demonstration of this failover using NFS shares is to place MP3 files on the share. Use the client system to stream the MP3 files and play the music from another system. During failover there should be little or no interruption to the music or the streaming.

Figure 6-9 shows the clustat output during failover.



```
Cluster Status Monitor (SG246261-Cluster)                            10:11:16

Cluster alias: rhcmcluster

========================= M e m b e r   S t a t u s ==========================

  Member         Status     Node Id    Power Switch
  -------------- ---------- ---------- ------------
  node1          Down       0          Unknown
  node2          Up         1          Good

========================= H e a r t b e a t   S t a t u s ====================

  Name                           Type       Status
  ------------------------------ ---------- ------------
  hb1node1     <--> hb1node2     network    OFFLINE
  hb2node1     <--> hb2node2     network    OFFLINE

========================= S e r v i c e   S t a t u s ========================

                                         Last             Monitor  Restart
  Service        Status   Owner          Transition       Interval Count
  -------------- -------- -------------- ---------------- -------- -------
  nfs_share      started  node2          10:10:31 Mar 25  30       0
```

Figure 6-9 Clustat during failover

You can cleanly relocate the service using the following command:

```
cluadmin -- service relocate nfs_share
```

7

Configuring ESS for Linux

This chapter describes functions of the IBM TotalStorage Enterprise Storage Server Specialist (ESS Specialist). The information presented is an extract and summary from the ESS user guide with focus on how to configure fixed block storage in the ESS, as required by Linux systems.

This chapter is intended for Linux users not familiar with the function and operations of the ESS Specialist.

Remember that Linux for zSeries requires S/390 storage (CKD) to IPL (boot) from. For detailed information on how to configure the ESS for CKD hosts as well as open systems environments, please refer to the redbook *IBM TotalStorage Enterprise Storage Server: Implementing the ESS in Your Environment*, SG24-5420, and to *IBM TotalStorage Enterprise Storage Server Web Interface User's Guide*, SC26-7448.

The settings and procedures described in this chapter apply to the ESS Model 800.

7.1 IBM TotalStorage ESS Specialist

The ESS includes the ESS Specialist, which is a network enabled management tool that allows the storage administrator to monitor and manage storage from the IBM TotalStorage Enterprise Storage Server Master Console (ESS Master Console), or from a remote workstation using a Web browser.

By using a secure Internet connection (LAN with a Web browser), such as Netscape Navigator, or Microsoft Internet Explorer, your storage administrator can coordinate the consolidation effort and easily integrate storage capacity into the ESS.

The ESS Specialist provides you with the ability to do the following:

- Monitor error logs: If a problem occurs, a description of the problem including the failed component, the problem severity, and who is to be automatically notified is described.
- View the ESS status: Logical schematic of the ESS environment including the host attached ports, controller and cache storage, device adapters, devices and host icons may be checked.
- View and update the configuration: A color schemed view of the storage, including the amount of space allocated and assigned to one or more hosts, space allocated and not yet assigned, and space not allocated to logical volumes may be viewed.
- Add host systems or delete host systems
- Configure host ports
- Add volumes, remove volumes, and reassign volumes between different servers. Volumes can be reassigned between hosts as follows:
  - Removing volumes (or unassigning volumes from hosts). Volumes can be removed by removing all logically attached host connections to the logical volume.
  - Adding volumes. Volumes can be added from subsystem capacity that has never been defined or after an array has been reinitialized.
  - Reclaiming previously defined logical volumes
- View communication resource settings, such as TCP/IP configuration and users
- View cluster Licensed Internal Code (LIC) levels. You can view the active level, next level yet to be activated, and the previous level.
- Select one of the following authorization levels for each user:
  - Viewer. A viewer can view the current configuration and status information.

  - Operator. An operator can perform view and operation functions, such as changing the remote service and PE password.
  - Configurator. A configurator can view the current configuration and status information and can make changes to the configuration.
  - Administrator. An administrator can define new user IDs, delete old IDs, and assign, change, or revoke passwords and levels of authorization.
- Web support for ESS Copy Services (PPRC and FlashCopy)

7.2 Using the ESS Specialist

The ESS Specialist interface (and the ESS Copy Services interface) consist of a set of Java applets, which are programs that are dynamically loaded by the browser, and which execute within your browser. When you request a change to the configuration, the Java applets communicate with the microcode running on the ESS clusters to retrieve the current configuration data, submit the requested configuration change, and display the outcome of the request.

You must use a browser that contains the proper Java Virtual Machine (JVM) implementation to support these applets. The browser software provided by different companies, and even different versions of the same browser, vary widely with respect to their JVM support. Consequently, not all browsers are capable of supporting the ESS Specialist or ESS Copy Services.

The ESS Web interfaces support both the Netscape Navigator and the Microsoft Internet Explorer (MSIE) versions listed in Table 7-1.

Table 7-1 Web browsers supported by ESS Web interfaces

Netscape level (See Note 1)	**MSIE level (See Notes 2, 3, 4)**
Netscape 4.04 with JDK 1.1 fixpack	MSIE 4.x with Microsoft Java Virtual Machine (JVM) 4.0 or 5.0
Netscape 4.05 with JDK 1.1 fixpack	MSIE 5.x with Microsoft JVM 4.0 or 5.0
Netscape 4.06 (no fixpack required)	
Netscape 4.5x (no fixpack required)	
Netscape 4.7x (no fixpack required)	

<table>
<tr><th>Netscape level (See Note 1)</th><th>MSIE level (See Notes 2, 3, 4)</th></tr>
<tr><td colspan="2"><b>Notes:</b>
1. The ESS Web interfaces do not support Netscape above version 4.7.x
2. If your ESS is running with ESS LIC earlier than level 1.3.0 or SC01206, the performance of the ESS Web interfaces on MSIE 5.0 with JVM 5.0 is slower than with Netscape. It is recommended that you use Netscape as the browser or move to LIC level 1.3.0 or higher
3. MSIE 5.0 with JVM 4.0 is supported with all levels of ESS code. However, it is not recommend that you change JVM 5.x to JVM 4.0 on the ESSNet machine in order to improve performance. It is not trivial to change the JVM to a lower level.
4. The ESS Master Console running Linux does not support the MSIE browser.</td></tr>
</table>

Using the ESS Specialist client from Linux

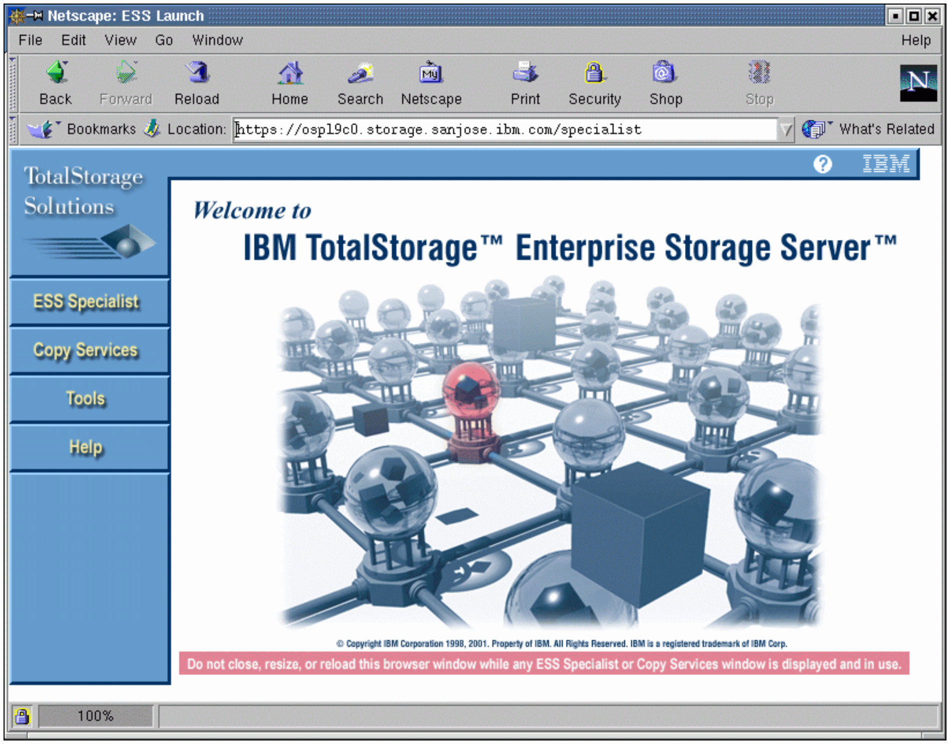
In order to start and successfully work with the ESS Specialist client from a Linux workstation, you have to install a supported version of the Netscape Web browser. We used Netscape Version 4.79, which is available from:

```
http://wp.netscape.com/download/archive/client_archive47x.html
```

> **Note:** In preparation of this redbook, we have tried several different combinations of current Web browsers, plug-ins, and Java virtual machines (e.g. Mozilla, Konqueror, Netscape, IBM, and Sun JDKs). We were not able to get any combination of the more recent software to work properly. Our recommendation is to just use the older Netscape version that includes a supported JVM, which worked without any problems.

First logon

When you open a Web browser and log on to your ESS Specialist for the first time, you will get prompted to accept the site certificate from the ESS. After you accepted the certificate you get a welcome screen, as shown in Figure 7-1, from where you can start the actual ESS Specialist Java client application.



https://ospl9c0.storage.sanjose.ibm.com/specialist

Welcome to

IBM TotalStorage™ Enterprise Storage Server™

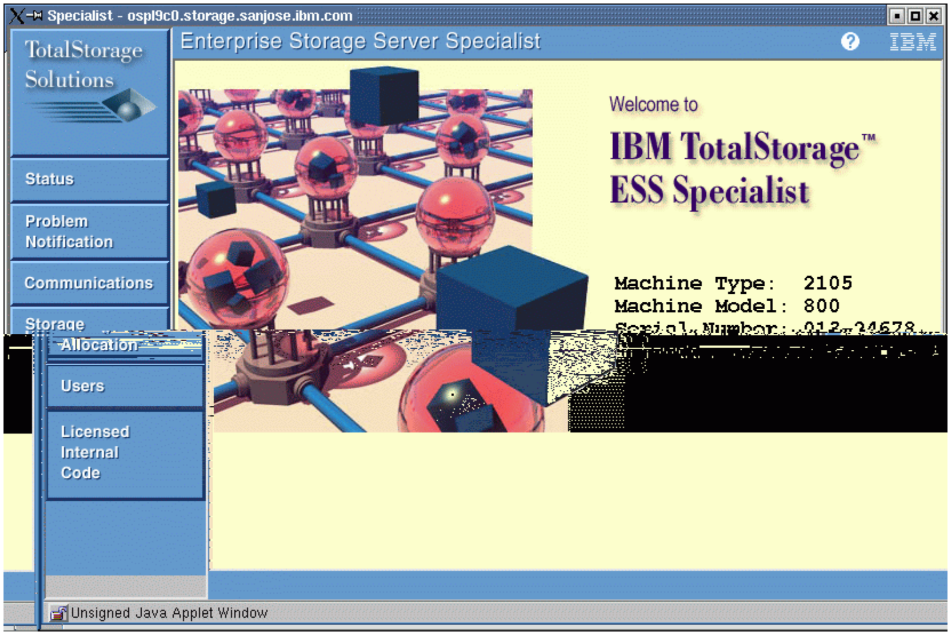
© Copyright IBM Corporation 1998, 2001. Property of IBM. All Rights Reserved. IBM is a registered trademark of IBM Corp.

Do not close, resize, or reload this browser window while any ESS Specialist or Copy Services window is displayed and in use.

Figure 7-1 ESS welcome screen

When you click the **ESS Specialist** menu link, you are prompted for a user name and a password.[1] Depending on the performance of your client system, it might take awhile for the ESS Specialist applet to execute and display the main screen, as shown in Figure 7-2.

[1] To be able to make changes to the ESS configuration, you need a user of the class Configurator on the ESS master console, which you have to obtain from the administrator or which you create during the ESS setup and installation.


Specialist - ospl9c0.storage.sanjose.ibm.com
TotalStorage Solutions
Enterprise Storage Server Specialist
IBM
Status
Problem Notification
Communications
Storage Allocation
Users
Licensed Internal Code
Welcome to
IBM TotalStorage™ ESS Specialist
Unsigned Java Applet Window

Machine Type: 2105
Machine Model: 800
Serial Number: 012-24678

Figure 7-2 ESS Specialist main screen

The navigation frame of ESS Specialist contains six buttons that provide access to the major categories of information and tasks that you can perform. The buttons are:

Status The Status button accesses the operational status of your ESS, in graphical and log views.

Problem Notification The Problem Notification button accesses panels that enable you to set up ways to have the ESS automatically alert you and those you designate of operational problems.

Communications The Communications button accesses panels that enable you to set up remote technical support.

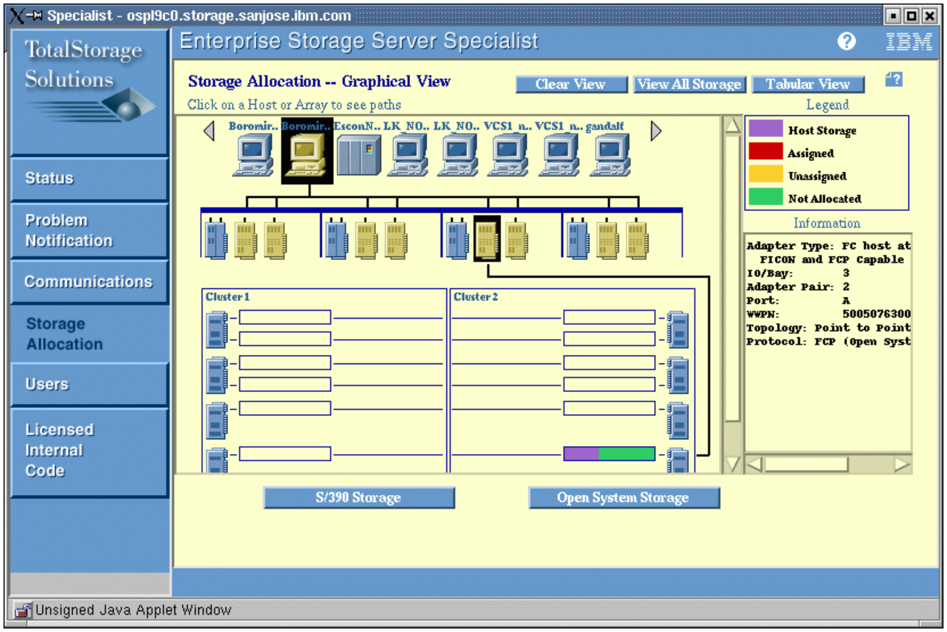
Storage allocation The Storage Allocation button accesses panels that enable you to set up and view the storage segmentation in your ESS.

Users The Users button accesses panels that enable you to give ESS users various levels of access to the ESS.

Licensed Internal Code The Licensed Internal Code button accesses panels that display the licensed internal code (LIC) in your ESS, the cluster on which the LIC is installed, and the licensed feature codes that you have purchased.

7.2.1 ESS attachment configurations

ESS Specialist displays the Storage Allocation panel after you click **Storage Allocation** from the navigation frame, which is shown in Figure 7-3. The Storage Allocation panel contains a logical view of the components of the ESS storage facility. It shows the server's logical configuration of host systems, host ports, storage adapters, arrays, and volume assignments. The Storage Allocation panel provides access to a graphical view and a tabular view of ESS components. The graphical view depicts the hardware components of the ESS with interconnecting lines to indicate a particular assignment or logical connection. The tabular view presents the same information in table format.



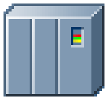
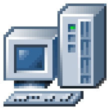
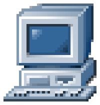
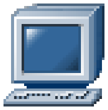
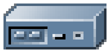
```
Adapter Type: FC host at
  FICON and FCP Capable
IO/Bay:          3
Adapter Pair:  2
Port:            A
WWPN:            5005076300
Topology: Point to Point
Protocol: FCP (Open Syst
```

Figure 7-3 Storage Allocation panel

The screen contains the following elements:

- The icons in the top row of the Storage Allocation-Graphical View panel represent the host systems that are attached to the ESS. Table 7-2 shows the types of host icons.

Table 7-2 Host icons and their descriptions

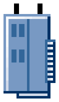
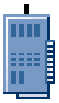
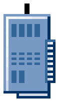
Host icon	Description
	A rectangular icon divided by two vertical lines represents a S/390 or zSeries host. It can also represent a Linux partition on that host. See Note 1.
	An icon of a display screen next to a box represents a AS/400® or iSeries host, except for Linux partitions within the host (UNIX icons). See Note 2.
	An icon of a display screen on top of a thin box represents a Microsoft Windows, Linux (x86), or Novell host. See Note 2.
	An icon of a display screen and a keyboard represents a UNIX host, such as Sun Solaris, IBM RS/6000, HP 9000, and Linux partitions in iSeries and pSeries hosts. See Note 2.
	A low, wide rectangular icon represents a Cisco iSCSI gateway (the SN 5420). See Note 2.

Notes:
1. When the icon has a FiconNet or EsconNet label, it is a pseudo-host representing all attached S/390 and zSeries hosts that are not yet defined to the ESS, attached through FICON or ESCON protocol, respectively.
2. When the icon has an Anonymous label, it represents open systems connected through the Fibre Channel protocol.

When one of the four types of icons represents a host not explicitly defined to the ESS, but which has access to one or more volumes configured on the ESS, the icon is called a “pseudo-host.” ESS Specialist adds a pseudo-host whenever you configure the LUN access mode for a Fibre Channel port to access-any mode.

- The icons in the second row represent the host adapter cards installed in the ESS. The lines between them indicate the four host adapter bays in the ESS. Each of the bays has a capacity of four host adapter cards. Depending on how many host adapters you purchased, the host bays may not be fully populated. Each host adapter card can be one of three types shown in Table 7-3, and cards of all types can be mixed in any order in any of the bays.

Table 7-3 Host adapter types

Adapter icon	Description
	ESCON adapter card (2 ports per card).
	Fibre Channel adapter cards (1 port per adapter, using FICON or FCP protocols)
	SCSI adapter card (2 ports per adapter).
Note that the ESCON and SCSI adapters contain two ports per adapter. The icons for these adapters are divided in half so that clicking on the left side of the adapter selects port A and clicking on the right side of the adapter selects port B. Fibre-channel adapters have only one port.	

- The graphical area below the row of host adapter icons displays the relationship of storage areas in the two ESS clusters. It is split into two columns each representing a cluster, and each row in each column has an icon at the outside of the row that represents a device adapter card. The lines between the device adapter icons represent the connections between the disk groups supported by the device adapter cards.

  There are normally four pairs of device adapter cards installed in an ESS, four cards in each cluster. The ESS uses SSA device adapters, each of which supports two SSA loops.

  The disk groups are the rectangles on the SSA loops. There are normally 16 disk groups installed in an ESS, and two disk groups on each SSA loop. The minimum ESS configuration is four disk groups, the maximum configuration is

48 disk groups (that is, six disk groups per SSA loop). The disk groups are usually split evenly among the two clusters.

The individual disk groups are drawn as small rectangles on the SSA loops and between the SSA adapter pairs of the clusters. When you select a host and associated port, each associated disk group is filled with a color-coded bar graph representing the space utilization within that group. The color coding for the bar graph is described in the legend box on the right side of the panel, has the following meaning:

Violet **Host Storage:** Violet represents the storage space occupied by volumes that are assigned to the currently selected host system.

Red **Assigned:** Red represents the storage space occupied by volumes that are assigned to one or more host systems, but not to the currently selected host system.

Yellow **Unassigned:** Yellow represents the storage space occupied by volumes that are allocated, but are not currently assigned to any host systems. This is usually a temporary condition, since unassigned space is essentially wasted space.

Green **Not allocated:** Green represents free storage space, that is, storage where no volumes have yet been allocated.

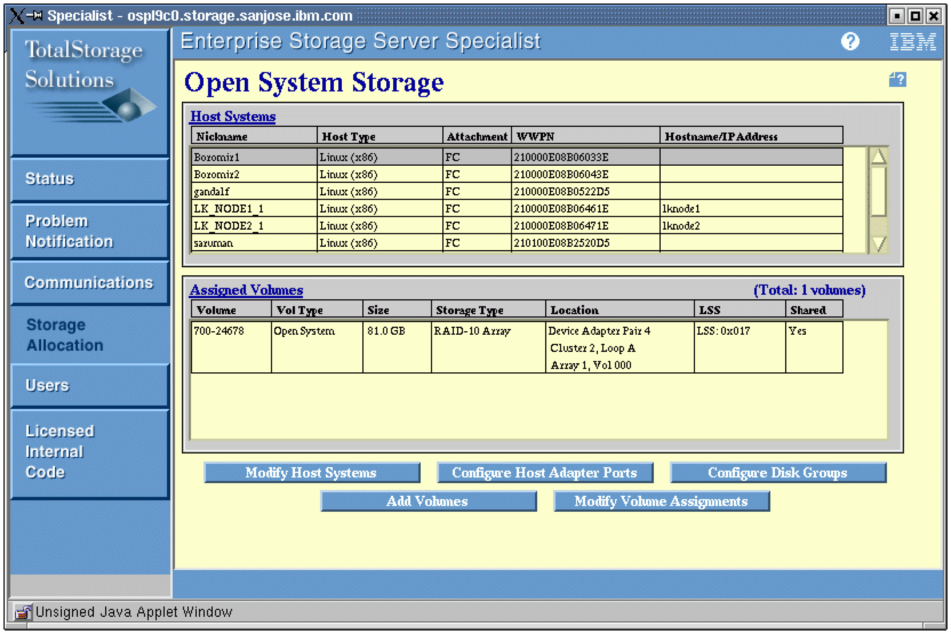
- At the bottom of the Storage Allocation-Graphical View panel are two buttons:
  - **S/390 Storage** - This button accesses the S/390 Storage panel, which you use for configuring storage for data in count-key-data (CKD) format attached to S/390 and zSeries (mainframe) host systems.
  - **Open Systems Storage** - This button accesses the Open Systems Storage panel, which you use for configuring storage for data in fixed-block (FB) format attached to UNIX-based, Intel-based and AS/400 host systems.

For this book we only use the Open Systems Storage function, since we concentrate on attachment of Fibre Channel storage for all the Linux hardware platforms. Even though the Linux for zSeries uses S/390 storage (CKD) as well to IPL (boot) from, the necessary steps to configure fixed block storage on the ESS are the same as for the other platforms. The explanation of how to configure S/390 storage is beyond the scope of this book. Please refer to the *IBM TotalStorage Enterprise Storage Server Web Interface User's Guide*, SC26-7448, for information on how to configure S/390 storage.

7.2.2 Storage allocation procedure

Use the Open System Storage panel to define and modify the definitions for Fibre Channel and SCSI-attached Linux host systems, and to assign physical storage units for them.

1. If you are not already on the Storage Allocation panel, **click** the corresponding button in the navigation frame of ESS Specialist. The Storage Allocation-Graphical View panel opens.
2. At the bottom of the panel, **click** Open System Storage. The Open System Storage panel opens as shown in Figure 7-4.



Open System Storage

Host Systems

Nickname	Host Type	Attachment	WWPN	Hostname/IP Address
Boromir1	Linux (x86)	FC	210000E08B06033E	
Boromir2	Linux (x86)	FC	210000E08B06043E	
gandalf	Linux (x86)	FC	210000E08B0522D5	
LK_NODE1_1	Linux (x86)	FC	210000E08B06461E	lknode1
LK_NODE2_1	Linux (x86)	FC	210000E08B06471E	lknode2
saruman	Linux (x86)	FC	210100E08B2520D5	

Assigned Volumes (Total: 1 volumes)

Volume	Vol Type	Size	Storage Type	Location	LSS	Shared
700-24678	Open System	81.0 GB	RAID-10 Array	Device Adapter Pair 4 Cluster 2, Loop A Array 1, Vol 000	LSS: 0x017	Yes

Figure 7-4 Open System Storage panel

3. Inspect the list of defined open-systems hosts and storage allocations in order to determine the remaining storage configuration tasks that you need to perform.
4. You generally perform configuration actions in the following order by clicking the associated buttons at the bottom of the Open System Storage panel:
   a. Click **Modify Host Systems** to open the Modify Host Systems panel and to define new hosts to ESS Specialist.
   b. Click **Configure Host Adapter Ports** to open the Configure Host Adapter Ports panel and identify to ESS Specialist the characteristics of the new

host adapter ports. For SCSI ports, you also use this panel to associate a port with a host system.

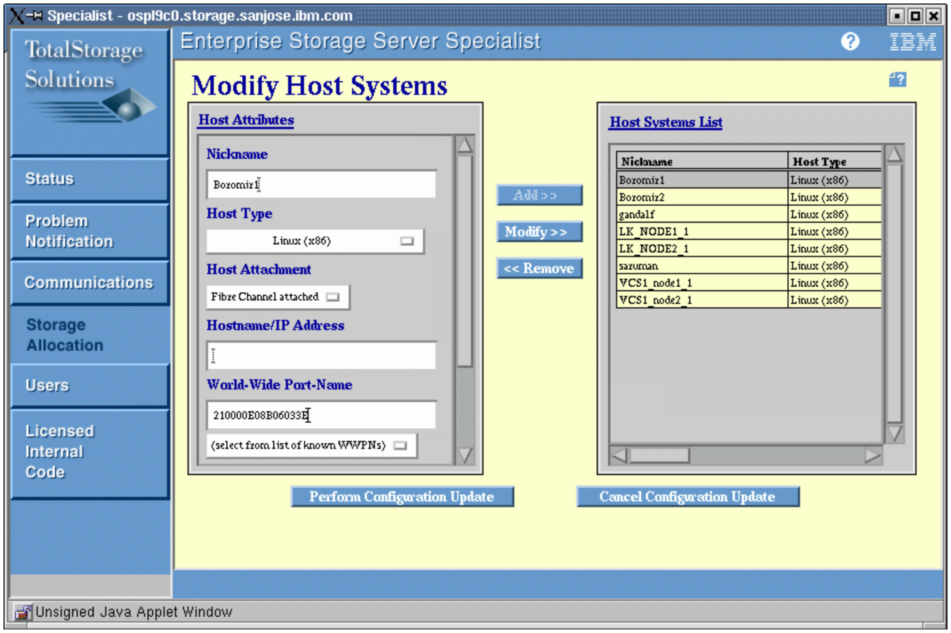
c. Click **Configure Disk Groups** to open the Fixed Block Storage panel and select a storage type and track format for a selected disk group.

d. Click **Add Volumes** to open the Add Volumes (1 of 2) panel and define numbers and sizes of volumes for selected arrays.

e. Click **Modify Volume Assignments** to open the Modify Volume Assignments panel and modify volume assignments, such as creating shared access to a volume, or removing an assignment.

7.2.3 Defining Linux hosts

On the Open System Storage panel click **Modify Host Systems**. The Modify Host Systems panel opens as shown in Figure 7-5.


Specialist - ospl9c0.storage.sanjose.ibm.com
TotalStorage Solutions
Status
Problem Notification
Communications
Storage Allocation
Users
Licensed Internal Code
Enterprise Storage Server Specialist
IBM
Add >>
Modify >>
<< Remove
Perform Configuration Update
Cancel Configuration Update
Unsigned Java Applet Window

Modify Host Systems

Host Attributes

Nickname: Boromir1

Host Type: Linux (x86)

Host Attachment: Fibre Channel attached

Hostname/IP Address:

World-Wide Port-Name: 210000E08B06033E

(select from list of known WWPNs)

Host Systems List

Nickname	Host Type
Boromir1	Linux (x86)
Boromir2	Linux (x86)
gandalf	Linux (x86)
LK_NODE1_1	Linux (x86)
LK_NODE2_1	Linux (x86)
saruman	Linux (x86)
VCS1_node1_1	Linux (x86)
VCS1_node2_1	Linux (x86)

Figure 7-5 Modify Host Systems panel

Enter the characteristics of the new host in the Host Attributes sub-panel:

1. In the Nickname field, type a name that uniquely identifies the associated host or host attachment within the ESS. The host nickname must not exceed 29 characters. ESS Specialist ignores the leading and trailing blanks that you entered.

2. Select a **host type** from the list in the Host Type field. There are three types of Linux hosts that you can choose from:
   - Linux (x86): PC servers with the Linux operating system with v2.4 kernel. Although you can configure 256 LUNs on the ESS, a Linux host can only support 128 LUNs.
   - Linux (iSeries/pSeries)
   - Linux (zSeries)
3. From the Host Attachment list you would usually select the type of interface used to connect the host to the ESS, but since the ESS supports Linux hosts only through the FCP protocol, **Fibre-channel attached** is the only option available in the Host Attachment list.
4. In the Hostname/IP Address field, if the host is connected to the IP network, you have the option of entering the host name or the dotted decimal IP address. You should enter information in this field if you are using the IBM TotalStorage Expert software package.
5. The Worldwide Port Name field is enabled if you select **Fibre-channel attached** from the Host Attachment list. If the host to be identified is Fibre-Channel attached, select the appropriate **WWPN** from the list in the Worldwide Port Name field, or type the WWPN in manually. The WWPN list contains the WWPNs of all host Fibre Channel adapters that are currently connected (logged in) but not yet defined to the ESS. The ESS discovers the WWPN from your host Fibre Channel adapter when you connect your host system to the ESS. If the connection is made through a single ESS Fibre Channel port, that port identifier is listed in parentheses following the WWPN. If the connection is made through multiple ESS Fibre Channel ports, the ports are not indicated in the list.
6. Scroll down in the Host Attributes sub-panel to display **Fibre Channel Ports** list in this field. The highlighted entries indicate the ports that this particular host system can use to access its assigned volumes in the ESS. If the first entry, `All installed ports`, is highlighted (which is the default) then this host system can use all Fibre Channel ports in the ESS to access its assigned volumes. Alternatively, to limit the ports through which this host can access the ESS, select one or more individual ports in the list.

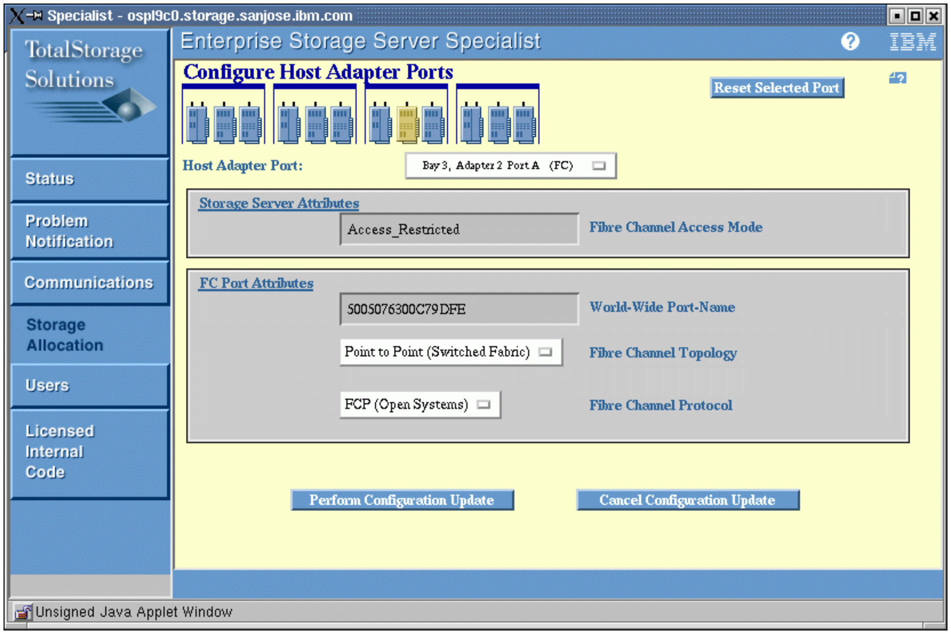
After you have filled in the complete information click **Add**. The Host Systems List, which is on the right side of the Modify Host Systems panel, displays the data that you entered about the new host, along with the attributes of all currently defined host systems.

Next click **Perform Configuration Update** at the bottom of the panel to apply the configuration changes to the ESS. A progress bar indicates the progress of the configuration action as the ESS processes it. Alternatively, click **Cancel**

Configuration Update to cancel the information that you entered. Clicking either button returns you to the Open System Storage panel, which displays the characteristics as you entered them of the new host or hosts that you defined.

7.2.4 Configure the host adapters

At the bottom of the Open System Storage panel, click **Configure Host Adapter Ports**. The Configure Host Adapter Ports panel opens. Click an **icon** for a Fibre Channel adapter (the icon has one port; see Table 7-3). Alternatively, select the **adapter** from the Host Adapter Ports list that is below the icon row. Figure 7-6 shows an example of the fields that are enabled on the Configure Host Adapter Ports panel when you click a Fibre Channel adapter.


Specialist - ospl9c0.storage.sanjose.ibm.com
TotalStorage Solutions
Enterprise Storage Server Specialist
IBM
Status
Problem Notification
Communications
Storage Allocation
Users
Licensed Internal Code
Reset Selected Port
Perform Configuration Update
Cancel Configuration Update
Unsigned Java Applet Window

Configure Host Adapter Ports

Host Adapter Port: Bay 3, Adapter 2 Port A (FC)

Storage Server Attributes

Fibre Channel Access Mode: Access_Restricted

FC Port Attributes

World-Wide Port-Name: 5005076300C79DFE

Fibre Channel Topology: Point to Point (Switched Fabric)

Fibre Channel Protocol: FCP (Open Systems)

Figure 7-6 Configure Host Adapter Ports panel

1. The Fibre Channel Access Mode field in the Storage Server Attributes box shows the current Fibre Channel access mode for the ESS, either:

   – Access-any
   – Access-restricted

   > **Note:** If you want to change the Fibre Channel access mode, ask your Service Support Representative to perform this change during a service action or during installation of a new host adapter.

2. Select one of the following attributes from the Fibre-Channel Topology list in the FC Port Attributes box:
   - **Point to Point (Switched Fabric)**. For unconfigured ports, point-to-point topology is the only choice if you access the Configure Host Adapter Ports panel from the S/390 Storage panel.
   - **Arbitrated Loop (Direct Connect)**

   > **Note:** If you want to connect a Linux for zSeries system, only a switched environment is supported. Thus, you have to select **Point to Point** (Switched Fabric).

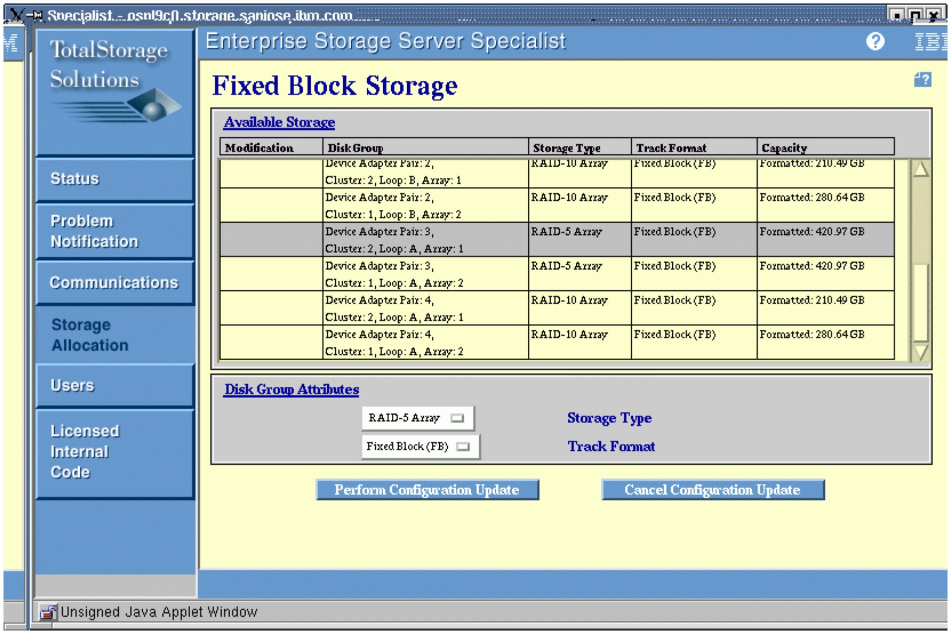
   If the port is already configured and you want to change its settings, you must first remove the configuration for the port by selecting **Undefined** from the list and then clicking **Perform Configuration Update**.

You can click **Reset Selected Port** at any time to cancel any pending configuration changes made to the currently selected port. To configure another Fibre Channel port, select the **port** from the Host Adapter Port list or from the port graphic, then repeat the configuration steps above. When all entries are complete for the Fibre Channel port, click **Perform Configuration Update** to apply, or **Cancel Configuration Update** to cancel the configuration step for all modified ports.

7.2.5 Configure RAID arrays

Before you can allocate storage space as logical volumes for use by specific host systems, you must define disk groups on the Fixed Block Storage panel. Defining a disk group means that you make two formatting selections for the selected disk group: storage type and track format. If you are allocating storage for an open-systems or Linux host (in other words, for a system other than one using the count-key-data (CKD) data format), use the Fixed Block Storage panel.

At the bottom of the Open System Storage panel, click **Configure Disk Groups**. The Fixed Block Storage panel opens as shown in Figure 7-7.



Specialist - ospl9cl0.storage.sanjose.ibm.com
TotalStorage Solutions
Status
Problem Notification
Communications
Storage Allocation
Users
Licensed Internal Code
Enterprise Storage Server Specialist
Unsigned Java Applet Window

Fixed Block Storage

Available Storage

Modification	Disk Group	Storage Type	Track Format	Capacity
	Device Adapter Pair: 2, Cluster: 2, Loop: B, Array: 1	RAID-10 Array	Fixed Block (FB)	Formatted: 210.49 GB
	Device Adapter Pair: 2, Cluster: 1, Loop: B, Array: 2	RAID-10 Array	Fixed Block (FB)	Formatted: 280.64 GB
	Device Adapter Pair: 3, Cluster: 2, Loop: A, Array: 1	RAID-5 Array	Fixed Block (FB)	Formatted: 420.97 GB
	Device Adapter Pair: 3, Cluster: 1, Loop: A, Array: 2	RAID-5 Array	Fixed Block (FB)	Formatted: 420.97 GB
	Device Adapter Pair: 4, Cluster: 2, Loop: A, Array: 1	RAID-10 Array	Fixed Block (FB)	Formatted: 210.49 GB
	Device Adapter Pair: 4, Cluster: 1, Loop: A, Array: 2	RAID-10 Array	Fixed Block (FB)	Formatted: 280.64 GB

Disk Group Attributes

Storage Type: RAID-5 Array

Track Format: Fixed Block (FB)

Figure 7-7 Fixed Block Storage panel

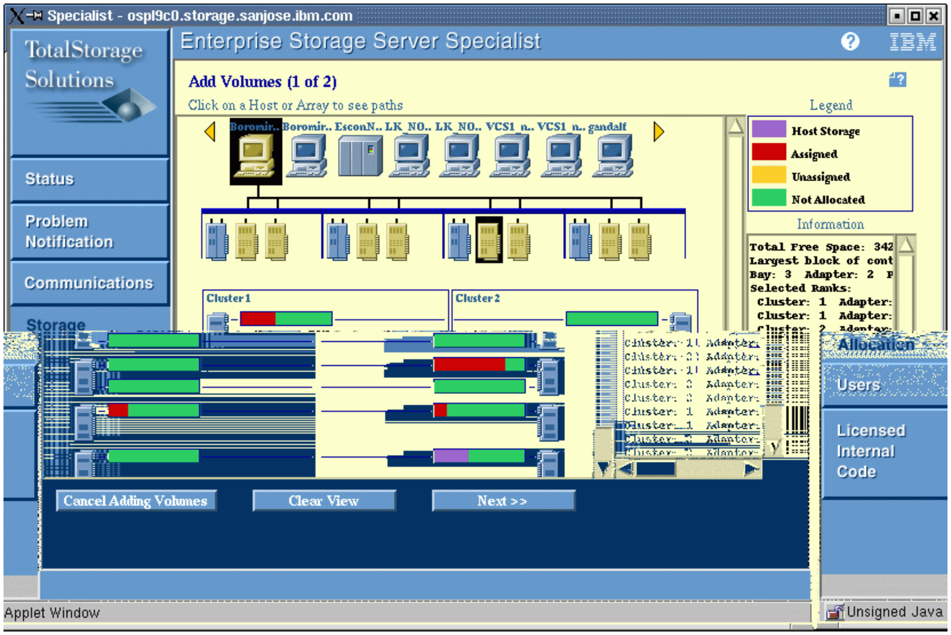
1. From the Available Storage table, select a **disk group**. If you select a disk group that has a value of Undefined in the Storage Type column, you can be sure that the disk group is not already being used. If you have a system that contains RAID-10 disk groups, you can gain capacity by converting them to RAID 5, or if you have a system containing RAID-5 disk groups, you can increase performance by converting them to RAID 10.
2. Select one of the following **storage types** from the Storage Type list:
   - RAID-5 Array
   - RAID-10 Array

   If you select RAID-10 Array or RAID-5 Array, the value in the Track Format field is automatically set to Fixed Block.

Repeat the sequence of steps on the Fixed Block Storage panel as needed to define or undefined RAID disk groups. Defined appears in the Modification column in the table to indicate which entries will be affected when you commit your changes. When you complete your desired changes, click **Perform Configuration Update** to apply the changes, or click **Cancel Configuration Update** to cancel the changes. In either case, the Open System Storage panel replaces the Fixed Block Storage panel in the browser window. Alternatively, you can click one of the buttons on the navigation pane to open another panel without committing the changes.

7.2.6 Allocating storage for Linux

Click **Add Volumes** at the bottom of the Open System Storage to panel to create fixed-block volumes on the ESS for use by your Linux hosts. This section divides the task of assigning open-systems volumes into two subsections, one for each panel. First the Add Volume (1 of 2) panel opens as shown in Figure 7-8.



```
Total Free Space: 342
Largest block of cont
Bay: 3  Adapter: 2  P
Selected Ranks:
 Cluster: 1  Adapter:
 Cluster: 1  Adapter:
```

Figure 7-8 Add Volumes panel (1 of 2)

1. In the top row of icons, click the **host icon** that represents the host system to which you want to assign the new volumes. The host adapter ports that are connected to this host system become highlighted. If the Linux host has any volumes already assigned, the disk groups that contain those volumes also become highlighted. You can select only one host system at a time for an add-volumes sequence. To refresh the view before selecting a different host system icon, click **Clear View** at the bottom of the panel.

   The host system that you select is highlighted in yellow and appears with a black background to indicate selection. The name of the host system appears in the Information box at the bottom right of the panel. A thin black line is drawn from this host system to one or more host ports in the second row that is attached to this system.

> **Note:** Before adding a LUN (a synonym for volume when speaking of open-systems volumes), or in any way modifying the LUN assignments for a host system port, ensure that the affected LUNs are offline or not in use by any attached host system. After adding a LUN, it might be necessary to restart the attached host systems in order to establish connectivity to the added LUN.

2. Select a **highlighted host adapter port**. The background of the port you select changes to black. Volumes are assigned to the selected host and are visible on all the ESS Fibre Channel ports that were defined for that host. You need to select the highlighted ports associated with the attached host to limit the host's access to those specific ports. A Fibre Channel attached open-systems host shows a connection to all of its configured Fibre Channel adapters.

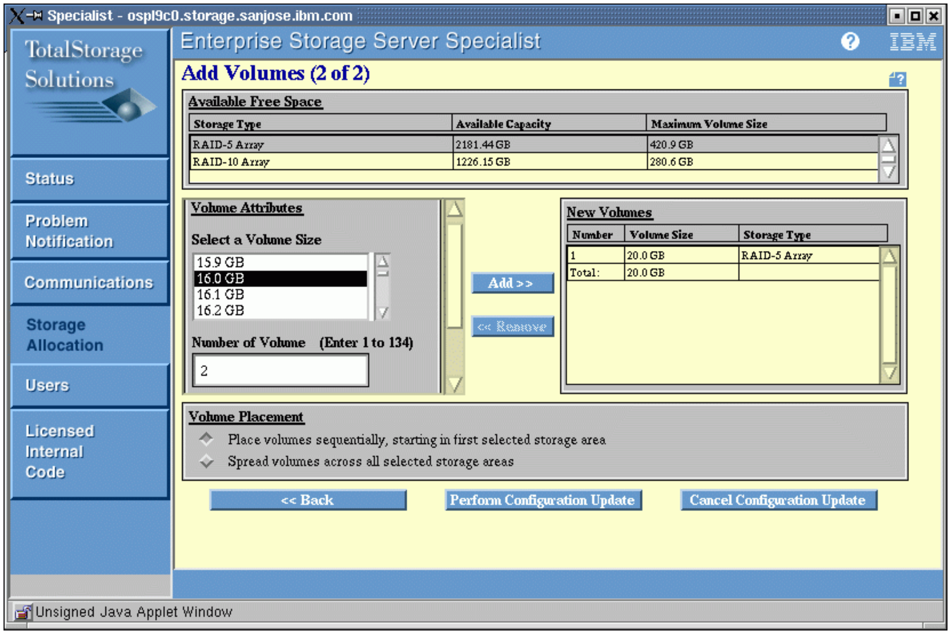
After you select a port, all fixed-block disk groups are selected by default and are highlighted. This provides a quick overview as to where the new volumes can be created. Optionally, select one or more fixed-block groups. The Information box on the right side of the panel summarizes the list of selected disk groups. If you select any groups, they will be the only ones used for the add-volume process. Alternatively, if you do not select any groups, the ESS will use all available disk groups for the add-volume process.

To proceed click **Next** and the Add Volumes panel (2 of 2) opens as shown in Figure 7-9:

1. In the Available Free Space table (at the top of the Add Volumes (2 of 2) panel), select the **table row** that corresponds to the type of storage that you want to allocate. You must select a type for which the available capacity is greater than 0.

   The information in the table is based on the capacities of the storage areas selected on the first Add Volumes panel. Each row in the table shows the available capacity for a specific storage type. Whether RAID 5 or RAID 10 is displayed in the table depends first on the storage types that are available for your ESS, and on what storage type you selected when you configured your fixed-block disk groups (see 7.2.5, "Configure RAID arrays" on page 169).

   The Available Capacity column indicates the total available space in the selected storage areas, and the Maximum Volume Size column displays the largest contiguous free space in the selected storage areas, by storage type. If you selected particular disk groups on the Add Volumes (1 of 2) panel, the defined storage types available on those selected disk groups are reflected in the Available Free Space table.


Specialist - ospl9c0.storage.sanjose.ibm.com
TotalStorage Solutions
Status
Problem Notification
Communications
Storage Allocation
Users
Licensed Internal Code
Enterprise Storage Server Specialist
IBM
Add >>
<< Remove
<< Back
Perform Configuration Update
Cancel Configuration Update
Unsigned Java Applet Window

Add Volumes (2 of 2)

Available Free Space

Storage Type	Available Capacity	Maximum Volume Size
RAID-5 Array	2181.44 GB	420.9 GB
RAID-10 Array	1226.15 GB	280.6 GB

Volume Attributes

Select a Volume Size: 15.9 GB, 16.0 GB (selected), 16.1 GB, 16.2 GB

Number of Volume (Enter 1 to 134): 2

New Volumes

Number	Volume Size	Storage Type
1	20.0 GB	RAID-5 Array
Total:	20.0 GB	

Volume Placement

○ Place volumes sequentially, starting in first selected storage area
○ Spread volumes across all selected storage areas

Figure 7-9 Add Volumes panel (2 of 2)

2. In the Volume Attributes section:

   a. Select a **Volume Size** from the so named list. The Available Free Space table at the top of the panel displays the maximum volume size that can be allocated in the selected storage areas. Size is specified in gigabytes.

   b. In the Number of Volumes field, type the number of volumes that you want to add.

   c. If you have not already used the Available Free Space table to select the type of **disk group** in which to place the new volumes; you can scroll down in the Volume Attributes section and select the **type** from the Storage Type list.

3. When you are creating multiple volumes and you have selected multiple disk groups, you can choose how the volumes should be placed. Near the bottom of the panel, select one of the following **Volume Placement** options:

   – **Place volumes sequentially, starting in first selected storage area** (the default): This choice creates volumes starting in the first storage area, and continues allocating them there until the capacity in that storage area is depleted. At this point, allocation continues in the second storage area, and so on.

– **Spread volumes across all selected storage areas**: This choice allocates the first volume in the first storage area, the second volume in the second area, the third volume in the third area, and so on. Wrapping occurs if necessary, so that after allocating a volume in the last area, the next volume is allocated in the first area. If a particular volume is too large for a particular area, that area is skipped.

4. To update the New Volumes table click **Add**. The Available Capacity and Maximum Volume Size columns in the Available Free Space table at the top of the panel are also updated. The New Volumes table lists all defined volumes that are pending creation. The last row in the table indicates the total amount of space, in gigabytes, that has been defined. To remove volumes from the list to be created, select one or more volumes from the New Volumes table, then click **Remove**. The last row in the table and the Available Capacity and Maximum Volume Size columns at the top of the panel are updated appropriately.

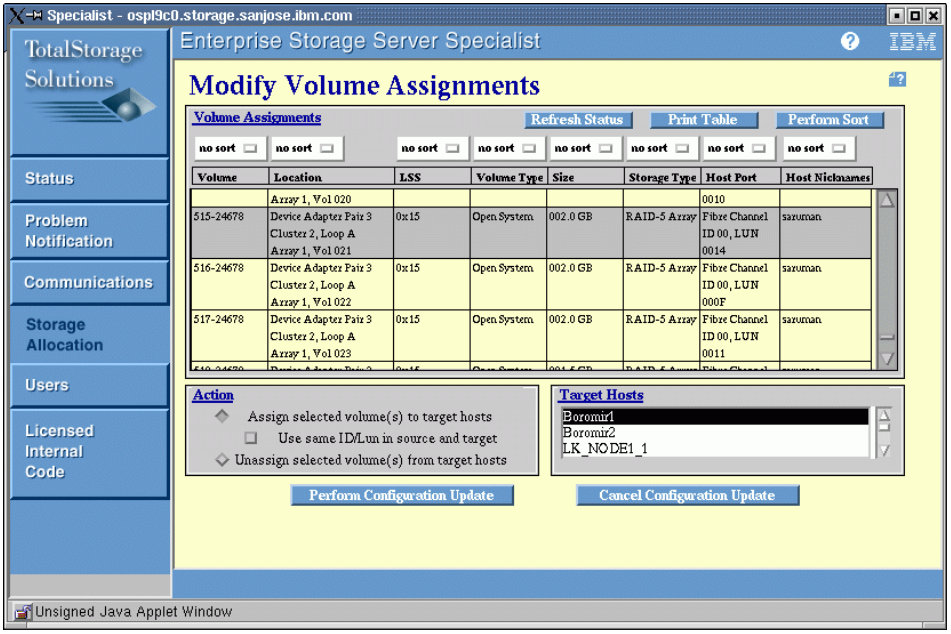
> **Note:** It is important that you carefully plan, create, and assign your LUNs. Once you perform the configuration update, the only way to physically delete the LUN is to reformat the whole array.

When you are finished adding or deleting volumes, click **Perform Configuration Update** to apply the changes to the fixed-block volume configuration or **Cancel Configuration Update** to cancel the changes. A progress bar indicates the progress of the configuration action as the ESS processes it. After the configuration changes are completed or cancelled, you are returned to the Open System Storage panel.

> **Attention:** When volumes are created and assigned to a port, the ESS initiates a formatting process that can exceed one hour for one or more volumes. A volume is not available to the host system while the ESS is formatting it. Do not attempt to use the newly created volumes until the ESS reports that it has completed the configuration. You can click **Refresh Status** on the Modify Volume Assignments panel to check the progress of the formatting process on a volume. If the formatting process is complete, no progress indication appears. Note also that the refresh function requires some time to run.

Modify the volume assignments

After using the Add Volumes panels to create a volume and assign it to a Linux host, you can use the Modify Volume Assignments panel to modify the assignment of the volume, including removing it from the host or assigning it to an additional host. Click **Modify Volume Assignments** in the Open System Storage panel for the new panel shown in Figure 7-10.



Modify Volume Assignments

Volume	Location	LSS	Volume Type	Size	Storage Type	Host Port	Host Nicknames
	Array 1, Vol 020					0010	
515-24678	Device Adapter Pair 3 Cluster 2, Loop A Array 1, Vol 021	0x15	Open System	002.0 GB	RAID-5 Array	Fibre Channel ID 00, LUN 0014	saruman
516-24678	Device Adapter Pair 3 Cluster 2, Loop A Array 1, Vol 022	0x15	Open System	002.0 GB	RAID-5 Array	Fibre Channel ID 00, LUN 000F	saruman
517-24678	Device Adapter Pair 3 Cluster 2, Loop A Array 1, Vol 023	0x15	Open System	002.0 GB	RAID-5 Array	Fibre Channel ID 00, LUN 0011	saruman

Action

◆ Assign selected volume(s) to target hosts

□ Use same ID/Lun in source and target

◇ Unassign selected volume(s) from target hosts

Target Hosts

Boromir1
Boromir2
LK_NODE1_1

Figure 7-10 Modify volume assignments

Click one or more **rows** in the table to select the volumes you want to assign. When you select a row, the two check boxes in the Action box below the table are enabled:

- In the Action box, select **Assign selected volume(s) to target hosts**.

  The optional **Use same ID/LUN in source** and target action becomes enabled. Select this if you want to keep the same volume ID and LUN for source and target hosts. The display changes to show in the Target Host box, hosts that are compatible for sharing the selected volume.

  – The **Target Host** box is populated with all the hosts that are connected by SCSI or Fibre Channel adapters to the ESS. Use the scroll bar to see the complete list. Select the host to which you want to assign the volume.

Click **Perform Configuration Update** to apply the modifications; or click **Cancel Configuration Update** to cancel the modifications to the volume assignments.

Perform the following steps to remove a volume assignment from a host:

- Open the Modify Volume Assignments panel, as described previously.
- In the **Volume Assignments** table, select the volume-host association to be removed. If you want to remove more than one association at one time, you can use the table to do so by selecting more than one row.

- Select **Unassign selected volume(s) from target host**.
- Select the **host** in the Target Host box. Depending on the characteristics of the associations between the volume and its associated hosts, the Host Nicknames field for a particular row might contain more than one host. In that case, the Target Host box would display all of the host names. Also, the Target Host box would display more than one host name if you selected more than one row in the Volume Assignments table. In the case where you have more than one host in the Target Host list, select each host that you want to remove from association with the selected volume.

Click **Perform Configuration Update** to apply the modifications, or click **Cancel Configuration Update** to cancel the modifications to the volume assignments. After the changes are made, ESS Specialist leaves you on the panel and refreshes the view, so it is easy to make changes incrementally.

Note: Removing a volume from a target host does not affect the volume definition, nor does it affect any of the data on the volume. The volume is not deleted; only the assignment to the target host is deleted.

8

BladeCenter SAN Utility

This chapter provides the details for installing the BladeCenter SAN Utility.

8.1 Installing the BladeCenter SAN Utility

The BladeCenter SAN Utility application is used to access and configure the BladeCenter Fibre Channel switch modules. The SAN Utility can be installed on a BladeCenter HS20 blade server or an external network management workstation configured with a supported version of Linux.

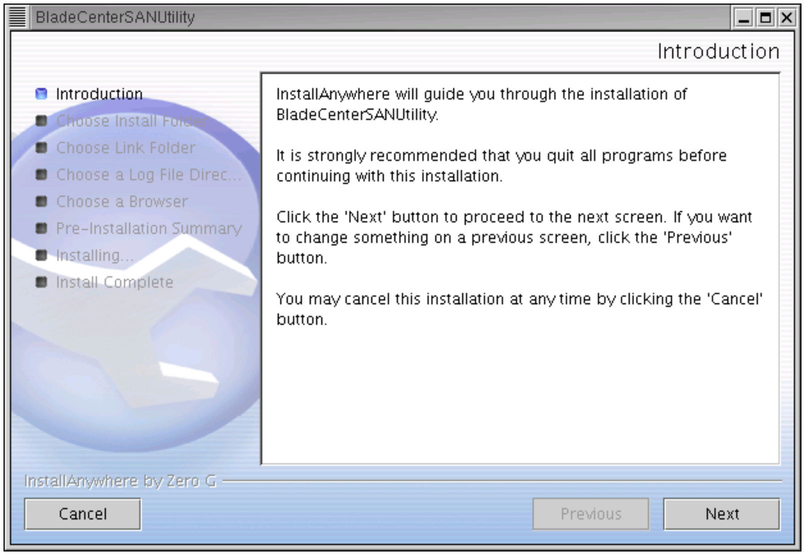
Before you configure your switch module, be sure that the management modules are properly configured. In order to access your switch module from an external environment, you may need to enable certain features, such as external ports and external management over all ports. See the applicable *BladeCenter Unit Installation and User's Guide* publications on the IBM eServer BladeCenter Documentation CD or at:

http://www.ibm.com/pc/support

Obtain the latest version of the BladeCenter SAN Utility. Install the SAN Utility using the following command:

```
sh Linux_<version>.bin
```

You should see a screen similar to Figure 8-1, read the details and click **Next**.



InstallAnywhere will guide you through the installation of BladeCenterSANUtility.

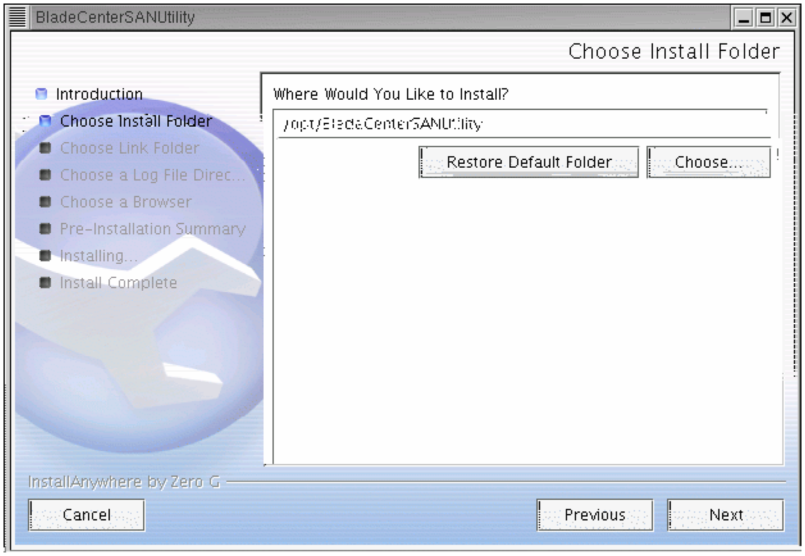
It is strongly recommended that you quit all programs before continuing with this installation.

Click the 'Next' button to proceed to the next screen. If you want to change something on a previous screen, click the 'Previous' button.

You may cancel this installation at any time by clicking the 'Cancel' button.

Figure 8-1 SAN Utility Installation Introduction screen

Choose the location for the installation. Stay with the default location or enter another location and click **Next** (Figure 8-2).



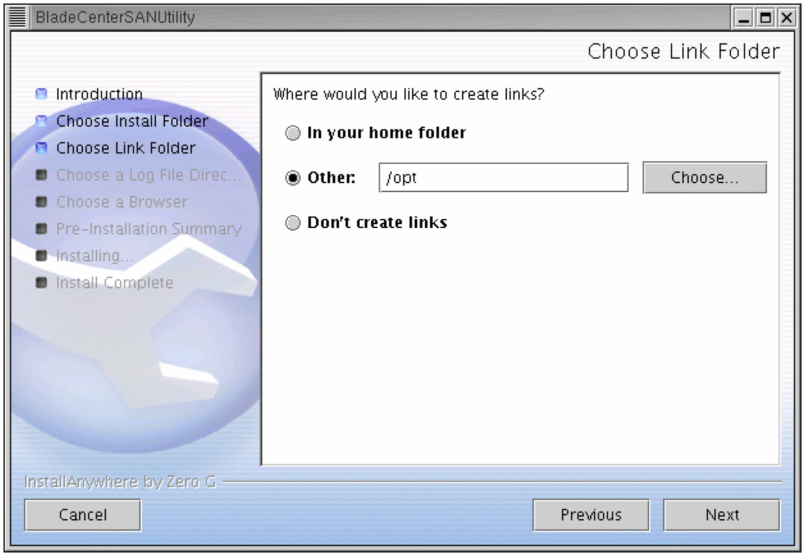
Where Would You Like to Install?

/opt/BladeCenterSANUtility

Figure 8-2 SAN Utility Installation Folder screen

You then get the following options regarding the creation of links (Figure 8-3):

- In your home folder
- Other: /opt is default
- Do not create links



Where would you like to create links?

○ In your home folder

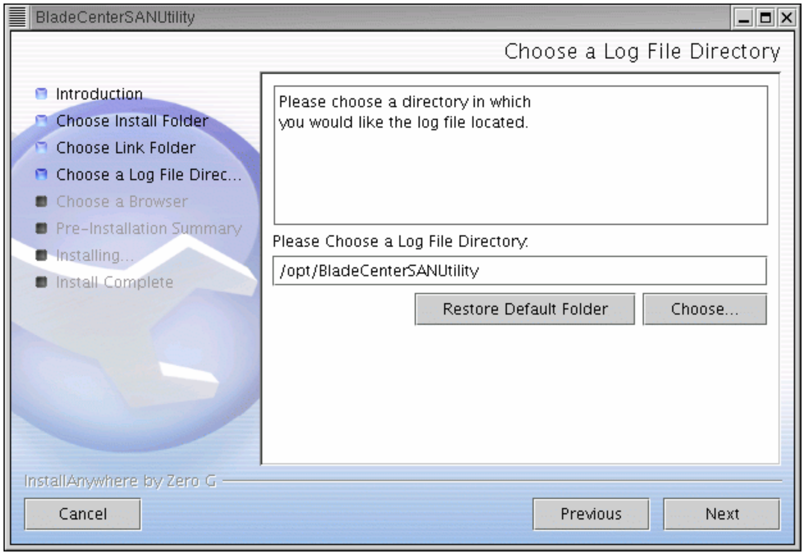
● Other: /opt

○ Don't create links

Figure 8-3 SAN Utility Create Links screen

Make your choice and click **Next**.

The next screen requires a location for log files. Stay with the default or enter a different location and click **Next** (Figure 8-4).



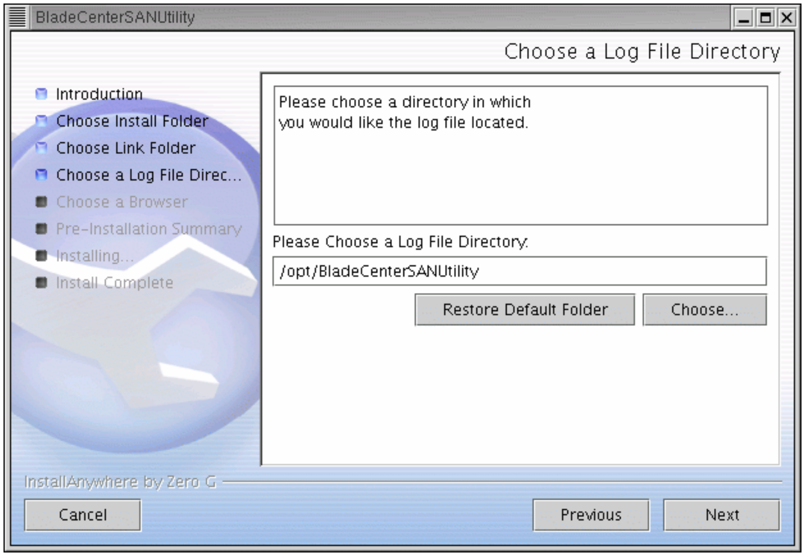
Please choose a directory in which you would like the log file located.

Please Choose a Log File Directory:

/opt/BladeCenterSANUtility

Figure 8-4 SAN Utility Log Location screen

Next, select the **browser** you would use for any online help and click **Next** (Figure 8-5).



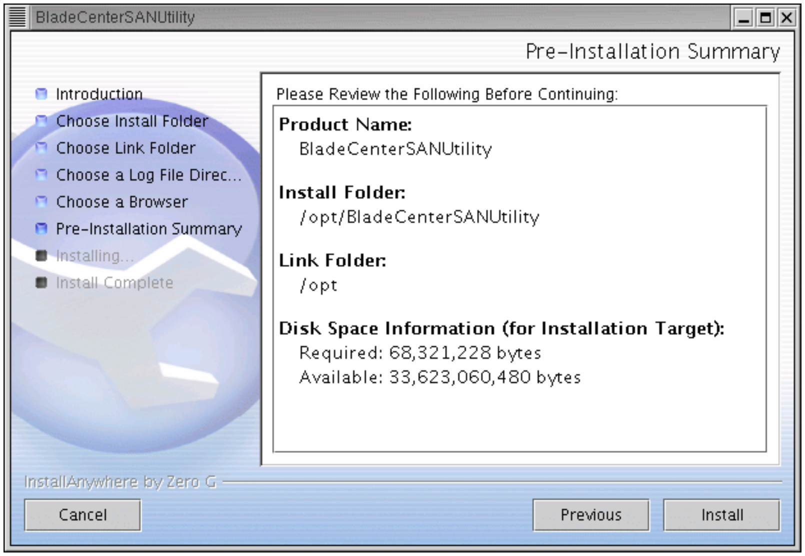
Please choose a directory in which you would like the log file located.

Please Choose a Log File Directory:

/opt/BladeCenterSANUtility

Figure 8-5 SAN Utility Browser Select screen

The next screen gives a pre-installation summary. This is the last chance to go back and change any settings before the software starts to copy files to disk. Make any changes if required and click **Install** (Figure 8-6).



Please Review the Following Before Continuing:

Product Name:
BladeCenterSANUtility

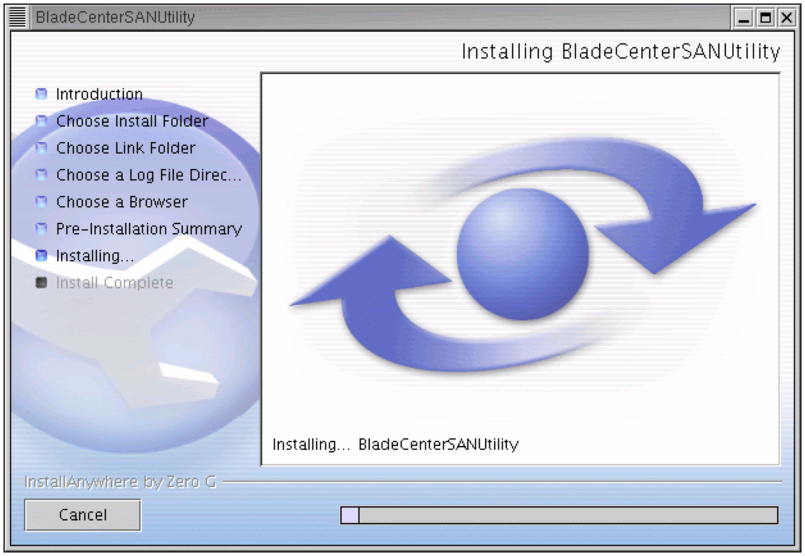
Install Folder:
/opt/BladeCenterSANUtility

Link Folder:
/opt

Disk Space Information (for Installation Target):
Required: 68,321,228 bytes
Available: 33,623,060,480 bytes

Figure 8-6 SAN Utility Pre-installation Summary screen

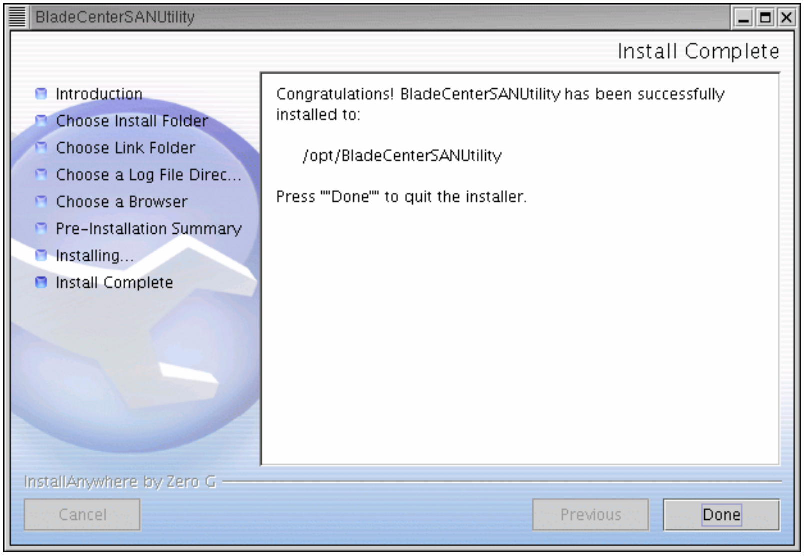
Installation begins (Figure 8-7).



Installing... BladeCenterSANUtility

Figure 8-7 SAN Utility Installation screen

Once installation is complete click **Done** (Figure 8-8).



Congratulations! BladeCenterSANUtility has been successfully installed to:

/opt/BladeCenterSANUtility

Press ""Done"" to quit the installer.

Figure 8-8 SAN Utility Install Complete screen

To run the SAN Utility type the following where `X` is the location where you installed the software:

```
/X/runBladeCenterSANUtility
```

You should see a screen similar to Figure 8-9.

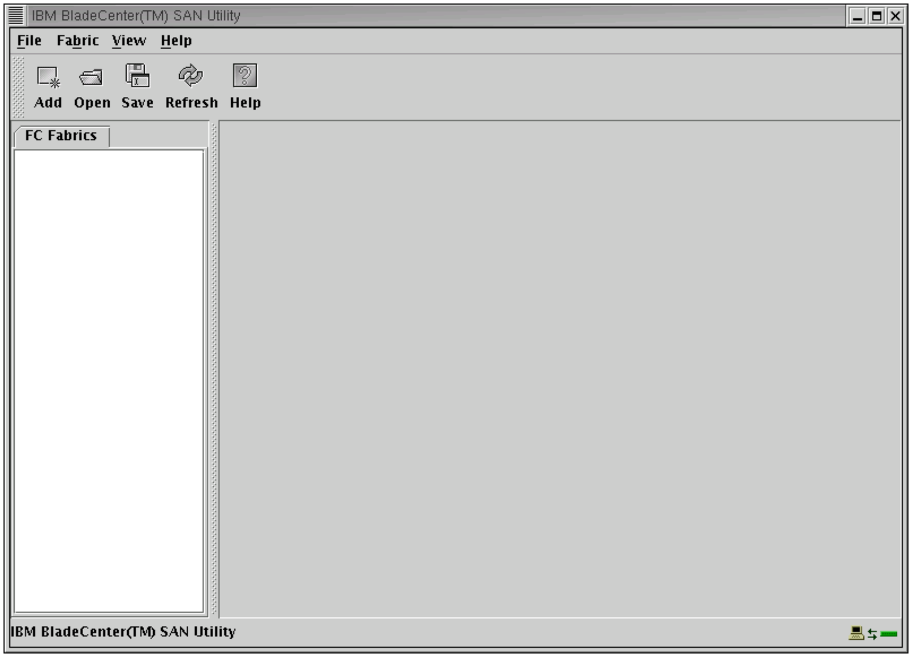
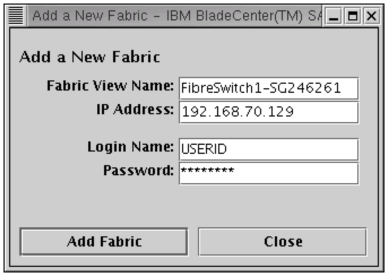

IBM BladeCenter(TM) SAN Utility
File
Fabric
View
Help
Add
Open
Save
Refresh
Help
FC Fabrics
IBM BladeCenter(TM) SAN Utility

Figure 8-9 SAN Utility screen

Click **Add** and enter the required details in the fields of the Add New Fabric screen (Figure 8-10).

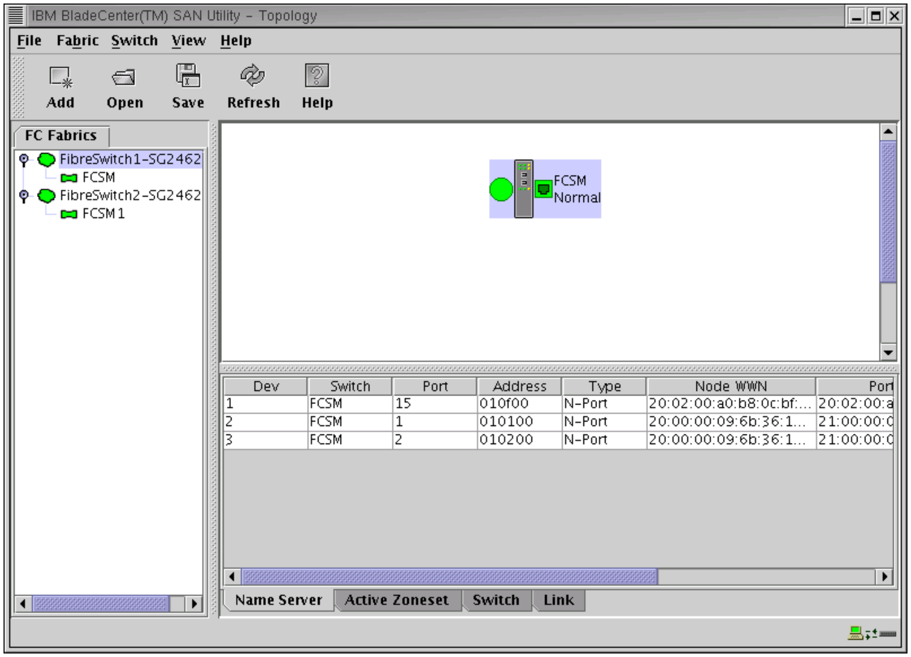
The default user ID is `USERID`, the default password is `PASSW0RD` (the sixth character is a zero, not the letter O). The user ID and password are case sensitive.



Fabric View Name: FibreSwitch1-SG246261
IP Address: 192.168.70.129
Login Name: USERID
Password: ********

Figure 8-10 Add New Fabric screen

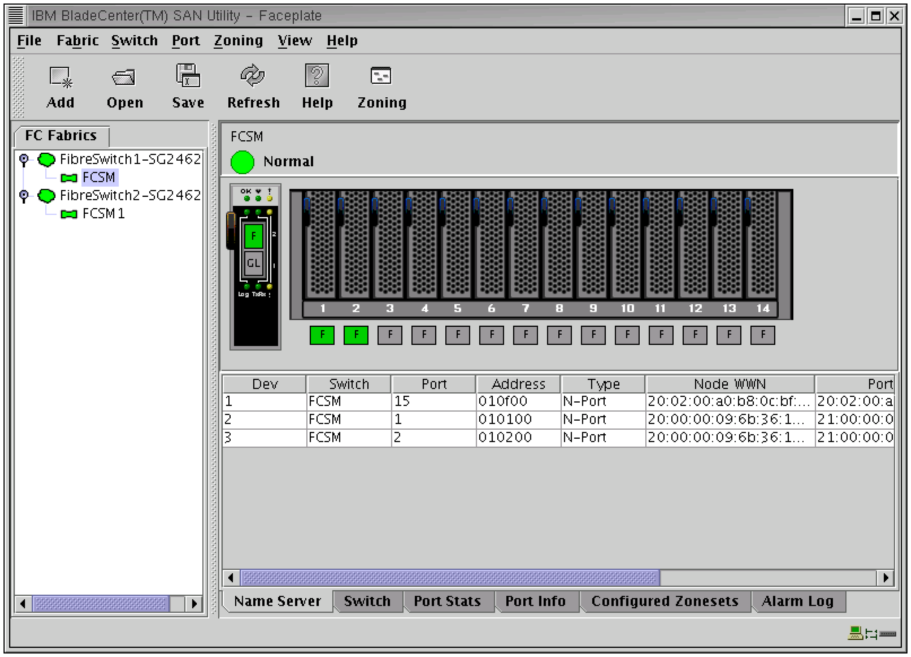
Highlighting the name of the fabric on the left hand column shows the `Topology` view (Figure 8-11).



Dev	Switch	Port	Address	Type	Node WWN	Port
1	FCSM	15	010f00	N-Port	20:02:00:a0:b8:0c:bf:...	20:02:00:a
2	FCSM	1	010100	N-Port	20:00:00:09:6b:36:1...	21:00:00:0
3	FCSM	2	010200	N-Port	20:00:00:09:6b:36:1...	21:00:00:0

Figure 8-11 SAN Utility Topology View screen

Highlighting the **Fibre Switch Module** on the left hand side shows the Faceplate view (Figure 8-12).



Dev	Switch	Port	Address	Type	Node WWN	Port
1	FCSM	15	010f00	N-Port	20:02:00:a0:b8:0c:bf:...	20:02:00:a
2	FCSM	1	010100	N-Port	20:00:00:09:6b:36:1...	21:00:00:0
3	FCSM	2	010200	N-Port	20:00:00:09:6b:36:1...	21:00:00:0

Figure 8-12 SAN Utility Faceplate view

Attention: If using two Fibre Switch Modules it is necessary to delete the factory default zoning in order for the switches to communicate with the storage.

The SAN Utility software can be used to change such things as port settings, zoning, alarms.

Once the Fibre Switch Modules are configured, you are ready to start preparing the storage.

9

FAStT Storage Manager

This chapter contains a description of the FAStT Storage Manager. FAStT software is used to define storage in the FAStT and to manage its availability to fibre-attached hosts.

This is not an exhaustive coverage as there are plenty of redbooks on the subject. While our emphasis here is on using the Storage Manager in the Linux environment, we also try to keep the content of this chapter as host-neutral as possible, with the host specific details being reserved to their respective chapters. We cover these topics:

- Storage Manager concepts
- FAStT Storage Manager client: Getting started
- Updating the FAStT firmware
- Setting up arrays, LUNs, and Storage Partitioning
- Configuring the LUN attachment

9.1 FAStT Storage Manager

The Fibre Array Storage Technology[1] (FAStT) requires a set of software and firmware to integrate with a host. The software includes a management component that provides a user interface to administer the storage system and additional programs such as an agent and utilities that may be required. Also included in the software may be a host-level driver that enables I/O path transfer within the host in the event of path failure in a multipath attachment. Additional software may also be needed to implement an appropriate environment on the management station, the host using storage provided by the FAStT, or both.

In the general case, The FAStT Storage Manager Software includes the Storage Management Client, the Storage Management Agent, the Storage Management Utilities, and the Redundant Disk Array Controller (RDAC). Not all installations require all components of the FAStT Storage Manager suite.

9.1.1 Storage Manager concepts

Management of the FAStT storage subsystem requires an instance of the FAStT Storage Manager client on a management station. Because the FAStT Storage Manager is a Java-based application, any suitable management station[2] may be used provided it can communicate with the FAStT storage subsystem in one of two ways:

in-band

The Storage Manager client communicates with an agent process over either a physical or loopback network connection. The agent process runs on one or more of the systems connected via Fibre Channel to the FAStT storage subsystem. The agent process communicates with the storage controllers over the Fibre Channel by means of a special interface in the controller known as an Access LUN. This method is also known as indirect control because of the intermediary agent process. No connection is required to the network ports on the FAStT controllers.

out-of-band

The Storage Manager client communicates over physical network connections to the FAStT controllers. The Access LUN is not used. This method is also known as direct control since no additional agent process is required, and no control communication takes place over the Fibre Channel. The default settings are 192.168.128.101 and 192.168.128.102. Linux communicates with the FAStT out-of-band, so if these IPs do not fit in with your network, then these will need to be changed.

[1] See also the redbook, *Fibre Array Storage Technology - A FAStT Introduction*, SG24-6246

[2] Supported systems include Linux, AIX, Windows, Solaris, Novell NetWare and HP-UX

Setting the IP and mask on a FAStT Controller

Connect a serial null modem cable between a serial port on the server (otherwise use a client system) and Port A on the FAStT. This allows the configuration of Controller A.

Use a terminal program such as `minicom` under Linux or HyperTerminal under Microsoft Windows.

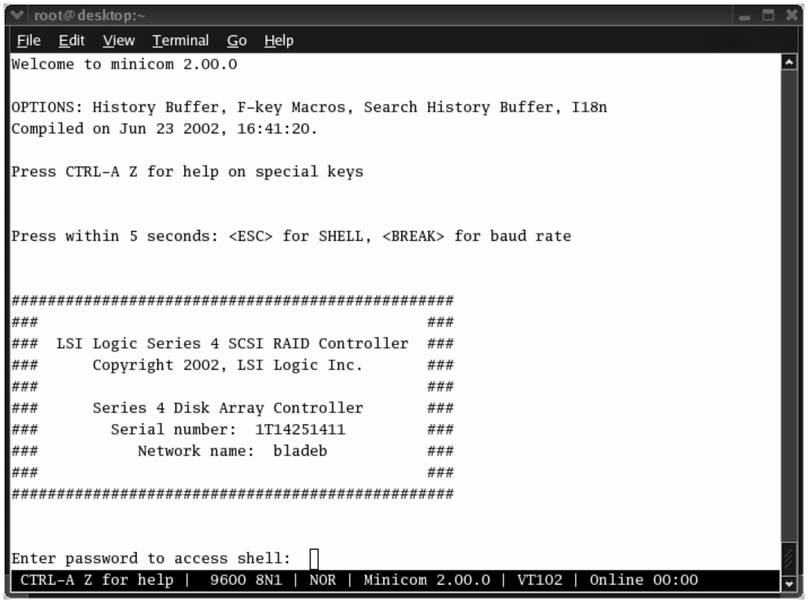
Use the following settings:

- Serial device (devv/ttyS1)
- Speed (9600)
- Data/parity/stop bits (8N1)
- Hardware Flow Control (Yes)
- Software Flow Control (No)

If using `minicom`, press Ctrl-A F to send break. If using HyperTerminal, press Ctrl-Break every 5 seconds until the ASCII characters become human readable.

Press Escape to enter the FAStT shell.

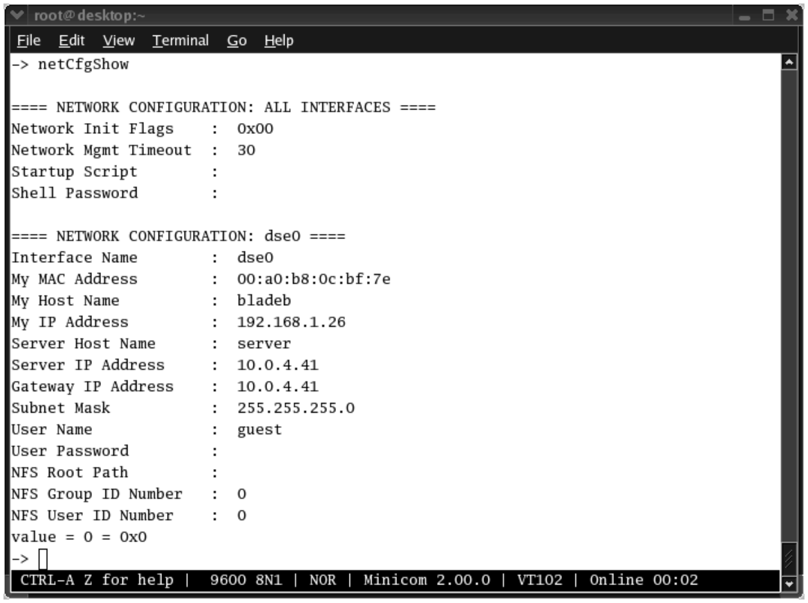
The shell prompts you for a password (Figure 9-1). The default is `infiniti`



```
Welcome to minicom 2.00.0

OPTIONS: History Buffer, F-key Macros, Search History Buffer, I18n
Compiled on Jun 23 2002, 16:41:20.

Press CTRL-A Z for help on special keys


Press within 5 seconds: <ESC> for SHELL, <BREAK> for baud rate


##################################################
###                                            ###
###  LSI Logic Series 4 SCSI RAID Controller   ###
###      Copyright 2002, LSI Logic Inc.        ###
###                                            ###
###      Series 4 Disk Array Controller        ###
###        Serial number:  1T14251411          ###
###           Network name:  bladeb            ###
###                                            ###
##################################################


Enter password to access shell:
 CTRL-A Z for help |  9600 8N1 | NOR | Minicom 2.00.0 | VT102 | Online 00:00
```

Figure 9-1 Password prompt for shell

Type `netCfgShow` to view the current settings (Figure 9-2).



```
-> netCfgShow

==== NETWORK CONFIGURATION: ALL INTERFACES ====
Network Init Flags    :  0x00
Network Mgmt Timeout  :  30
Startup Script        :
Shell Password        :

==== NETWORK CONFIGURATION: dse0 ====
Interface Name        :  dse0
My MAC Address        :  00:a0:b8:0c:bf:7e
My Host Name          :  bladeb
My IP Address         :  192.168.1.26
Server Host Name      :  server
Server IP Address     :  10.0.4.41
Gateway IP Address    :  10.0.4.41
Subnet Mask           :  255.255.255.0
User Name             :  guest
User Password         :
NFS Root Path         :
NFS Group ID Number   :  0
NFS User ID Number    :  0
value = 0 = 0x0
->
 CTRL-A Z for help |  9600 8N1 | NOR | Minicom 2.00.0 | VT102 | Online 00:02
```

Figure 9-2 netCfgShow screen

Type `netCfgSet` to change the settings. Change only `My IP Address` and `Subnet Mask`. Press Enter to bypass the other settings.

> **Important:** Do not change any other settings unless advised to do so by IBM Level 2 Technical Support. Any other alterations will place the FAStT in an unsupported configuration.

Once complete type `sysReboot` to restart the controller blade.

Wait for the reboot to complete and repeat this whole process again for Controller B by connecting the null modem cable to Port B, and reconnecting the terminal software.

Is there a Storage Manager agent for Linux?

The FAStT Storage Manager software does not include an agent process for all operating systems. For example, there is no agent that runs under Linux. A Linux system can run the Storage Manager client and manage a FAStT system using in-band control if there is a host that has a Fibre Channel connection to the FAStT, and runs the Storage Manager agent. In this case, the Storage Manager client on the Linux system would communicate with the agent on the non-Linux system. However, if the only hosts that are attached to the FAStT storage subsystem are Linux hosts, or if there are no attached hosts for which agent software is available that can be considered as management agent candidates, then out-of-band management must be used.

For this discussion we assume that the environment is purely Linux, and that out-of-band control will be used to manage the FAStT.

Is there an RDAC driver for Linux?

At the time of this writing, the Redundant Disk Array Controller (RDAC) is not available for Linux. This means that Linux hosts using storage on the FAStT have volume transfers within the FAStT managed by the FAStT itself. Multipathing is done using the appropriate driver for the Host Bus Adapter (HBA) and additional software called FAStT MSJ, that is used to manage the paths with the Host Bus Adapters. FAStT MSJ is described in Chapter 10., "FAStT MSJ (Management Suite for Java)" on page 211.

9.1.2 FAStT Storage Manager client: Getting started

The FAStT Storage Manager client is the graphical user interface (GUI) that controls one or more FAStT storage subsystems. Because this is a Java application, a Java runtime environment (provided on the FAStT Storage Manager CD) must be installed as well.

Linux Install of the FAStT Storage Manager client

Linux distributions of interest to us in this redbook come equipped with the **rpm** package manager. This tool maintains a dependency list such that if one package requires another, the user will be notified. When installing the Storage Manager client under Linux, you should install the Java runtime environment before the client software. In our implementation, the FAStT Storage Manager client resides on a Linux management station that can communicate with the FAStT controllers via the network. Such a management station may also be a host that is attached to the FAStT storage via Fibre Channel, but this is not required. To be clear, in the Linux environment, there is no Storage Manager agent available.

So, even if the Storage Manager client is installed on a host that is attached via Fibre Channel to the storage, the management still takes place out-of-band and requires network connectivity to the FAStT controllers. Both the client and the runtime environment are provided on the FAStT CD, but the latest versions (as well as the latest firmware for the FAStT controllers and expansion modules; please see Appendix 9.1.3, "Updating the FAStT firmware" on page 195 for detailed instructions on updating firmware) can be obtained from the following URL:

http://www.storage.ibm.com

From there, navigate to Disk Storage Systems,[3] then FAStT Storage Server, and then select your **FAStT model**; use the pull-down menu to select downloads, then select **FAStT Storage Manager Downloads** as appropriate for your version.

Read and accept the terms and conditions, and you now see a selection of both software and documentation. Please obtain and read the documentation entitled *"IBM TotalStorage FAStT Storage Manager Version X.Y Installation and Support Guide for Linux"* (where X.Y corresponds to your version number). Click your browser's "**back**" button and then select **LINUX operating system**; this action takes you to the download page for the code and the README file. Please look at the README file before installing, as it contains the latest information for the code release.

The package is rather large (over 30 MB) and requires another 70 MB to install, including the Java Runtime Environment. Make sure you allow enough room for unpacking the code, storing the rpm files, and the installed code. Also, please note that the Storage Manager client GUI requires the X window system be running on the management workstation.

You need to have root privilege to install the client. Site administrators may make their own decisions regarding access control to the Storage Manager client. The FAStT itself can be protected by password from changes to the storage configuration independently of access to the Storage Manager client.

First, uninstall any previous version of Storage Manager before installing the new version. To check for prior installations, query the packages that are installed and select only those containing the letters "**SM**" with the following command:[4]

```
rpm --query --all | grep SM
```

Use the output from this command to remove any prior version of the Storage Manager software by package name. For example:

```
rpm --uninstall SMruntime-XX.YY.ZZ-1
```

When you have obtained the latest version of Storage Manager, extract the contents of the zipped tar file:

```
tar zxvf *.tgz
```

The required `rpm` files for installation are extracted into `./Linux/SM`**`X`**`clientcode` where `X`=Storage Manager major version.

The Storage Manager components should be installed in the following order:

[3] Navigation details may change from time to time, these instructions were correct as of April 2003

[4] We have used the long option names for clarity. Experienced users will know of shorthand versions of the options. Please see the rpm documentation (**`man rpm`**) for more information

```
rpm --install --verify SMruntime-<version>.rpm
rpm --install --verify SMclient-LINUX-<version>.rpm
rpm --install --verify SMutil-<version>.rpm
```

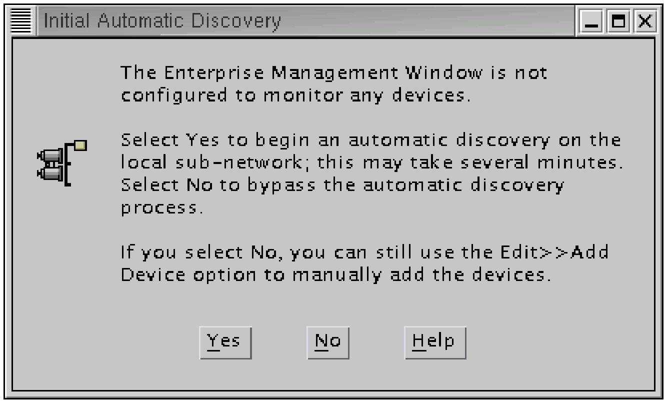
The Storage Manager software is installed in `/opt/IBM_FAStT`, with the client and command line interface, SMcli, installed in the `/opt/IBM_FAStT/client` directory.

Launching Storage Manager and connecting to the FAStT

To launch Storage Manager, open a terminal window and type **`SMclient`** [5] Alternatively, you can launch Storage Manager from the graphical interface menus:

Click the **Gnome** or **KDE** icon **-> Programs -> Utilities -> IBM FAStT Storage Manager**.

The screen shown in Figure 9-3 appears, asking if you wish to automatically discover the storage. This method will not find direct-attached storage controllers. Close this screen to see the Enterprise Management Window (Figure 9-4).



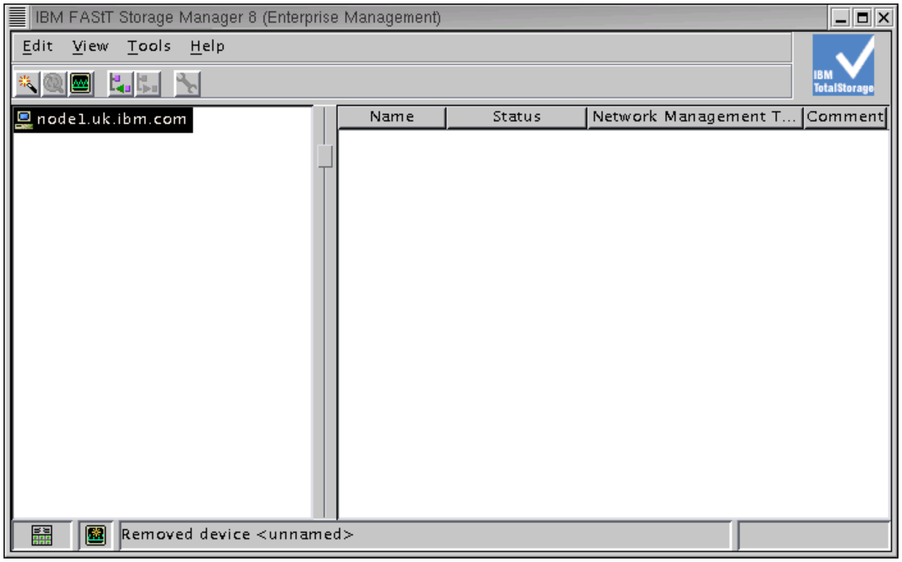
The Enterprise Management Window is not configured to monitor any devices.

Select Yes to begin an automatic discovery on the local sub-network; this may take several minutes. Select No to bypass the automatic discovery process.

If you select No, you can still use the Edit>>Add Device option to manually add the devices.

Figure 9-3 Automatic Discovery screen

[5] You may need to add the directory to your path. Otherwise, type the full path to SMclient

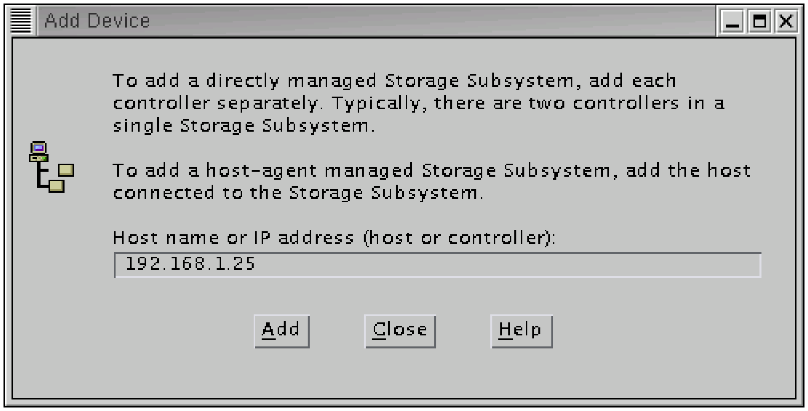


Removed device <unnamed>

Figure 9-4 Storage Manager Enterprise Management window

Select **Edit -> Add Device** from the menu. On the next screen, enter a Hostname or IP Address for each controller blade in the FAStT (Figure 9-5).

Enter a Host name or IP address you set on the FAStT and click **Add**. Repeat the process for the second controller blade.



To add a directly managed Storage Subsystem, add each controller separately. Typically, there are two controllers in a single Storage Subsystem.

To add a host-agent managed Storage Subsystem, add the host connected to the Storage Subsystem.

Host name or IP address (host or controller): 192.168.1.25

Figure 9-5 Add Device screen

Highlight **Storage Subsystem <unnamed>** and select **Tools -> Manage Device.**

If there was no configuration previously on the storage, you should see a Subsystem Management screen similar to Figure 9-6.

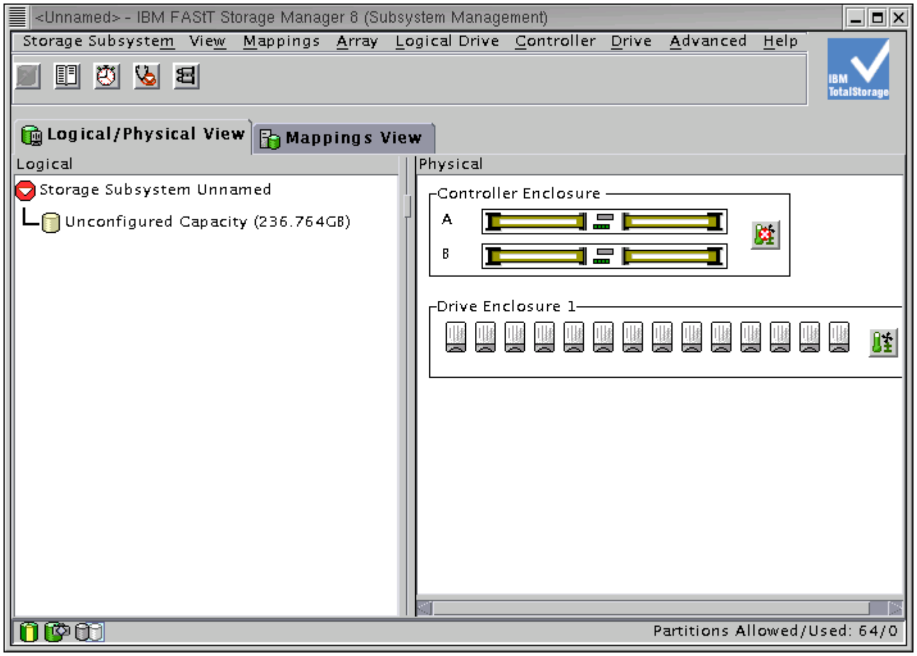


Figure 9-6 Subsystem Management screen

9.1.3 Updating the FAStT firmware

Obtain the latest firmware and NVSRAM for the FAStT controller as described in “Linux Install of the FAStT Storage Manager client” on page 191. Unzip the file to a temporary folder:

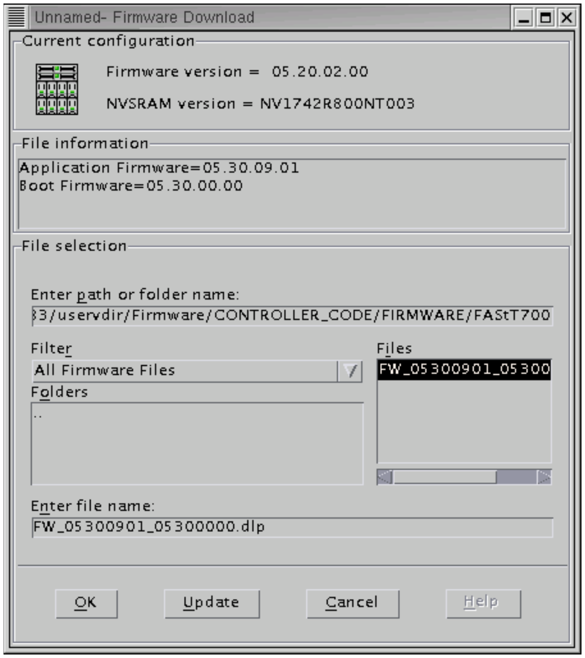
```
unzip *.zip
```

This extracts the firmware and NVSRAM files for all of the current models of FAStT controllers to the following folders:

```
./CONTROLLER CODE/FIRMWARE/<controller model>
./CONTROLLER CODE/NVSRAM/<controller model>
```

Note: You must first update the firmware and then the NVSRAM.

To update the controller firmware select **Storage Subsystem -> Download -> Firmware.** The screen shown in Figure 9-7 displays. Enter the location of the update file and click **OK**.


Unnamed- Firmware Download

Current configuration

Firmware version = 05.20.02.00

NVSRAM version = NV1742R800NT003

File information

Application Firmware=05.30.09.01
Boot Firmware=05.30.00.00

File selection

Enter path or folder name: 3/uservdir/Firmware/CONTROLLER_CODE/FIRMWARE/FAStT700

Filter: All Firmware Files

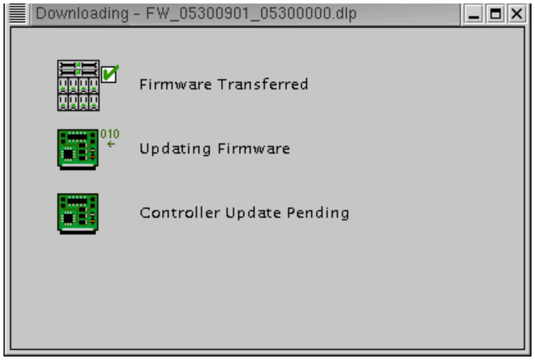
Files: FW_05300901_05300

Folders: ..

Enter file name: FW_05300901_05300000.dlp

Figure 9-7 Firmware Download Location screen

You now see a screen similar to Figure 9-8. This graphic persists until the firmware is transferred to the controllers and updated, which may take a few minutes.


Downloading - FW_05300901_05300000.dlp

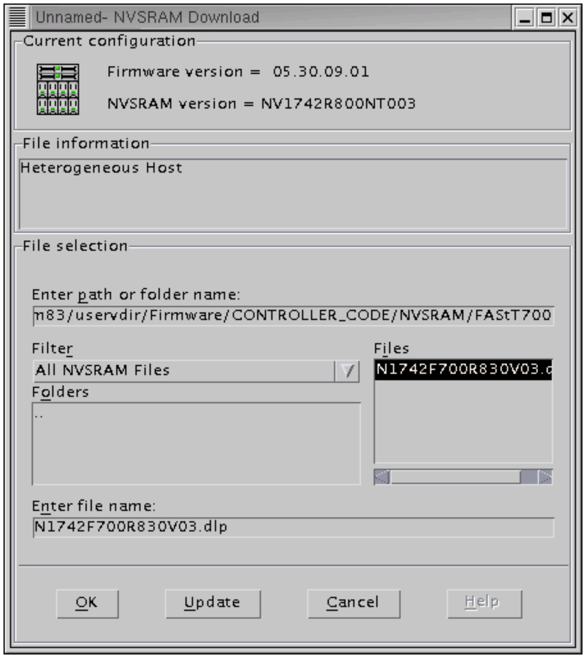
Firmware Transferred

Updating Firmware

Controller Update Pending

Figure 9-8 Firmware update screen

Next, update the NVSRAM. Select **Storage Subsystem -> Download -> NVSRAM**. Enter the location of the NVSRAM update file and click **Add** (Figure 9-9).



Unnamed- NVSRAM Download

Current configuration

Firmware version = 05.30.09.01

NVSRAM version = NV1742R800NT003

File information

Heterogeneous Host

File selection

Enter path or folder name: n83/uservdir/Firmware/CONTROLLER_CODE/NVSRAM/FAStT700

Filter: All NVSRAM Files

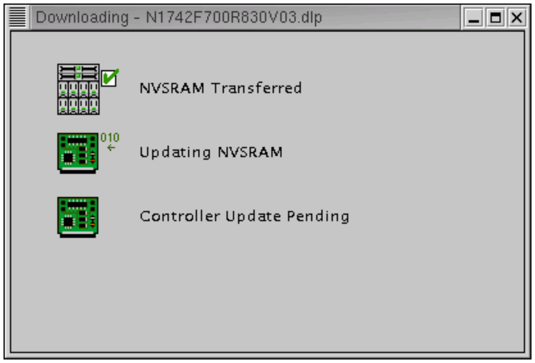
Files: N1742F700R830V03.c

Folders: ..

Enter file name: N1742F700R830V03.dlp

Figure 9-9 NVSRAM Download Location screen

A a screen similar to the one shown in Figure 9-10 displays.



Downloading - N1742F700R830V03.dlp

NVSRAM Transferred

Updating NVSRAM

Controller Update Pending

Figure 9-10 NVSRAM Update screen

Confirm both controllers are at the same level by selecting **View -> System Profile.**

Now that the controllers have been updated, we can use the Storage Manager to prepare the storage for attachment to the host systems.

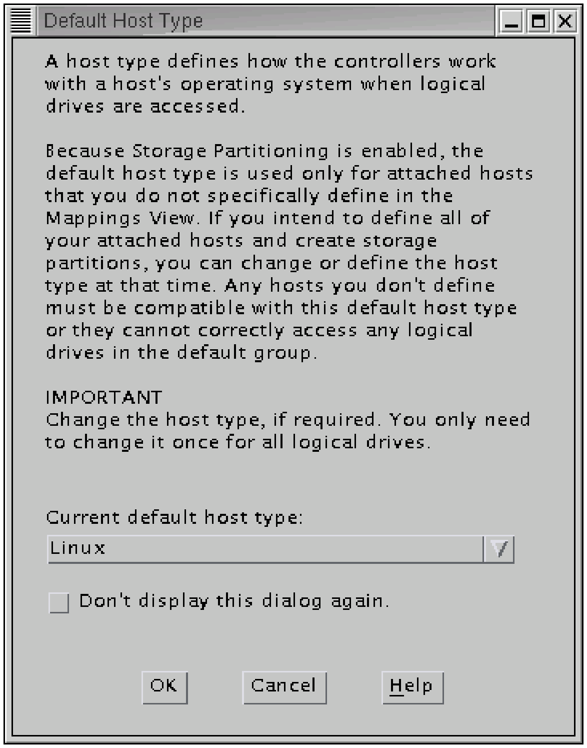
9.1.4 Setting up arrays, LUNs, and Storage Partitioning

This section takes you through the steps to set up storage arrays, LUNs, and Storage Partitioning. Refer to the redbook, *Fibre Array Storage Technology - A FAStT Introduction*, SG24-6246, for more on creating arrays and LUNs, selecting RAID levels, Storage Partitioning topics, for considerations on the Access LUN when defining the storage partitions (the sets of related logical drive mappings, host groups, and host ports).

The system is now ready to start creating arrays, LUNs, and Storage Partitioning. In this example we set up storage arrays for our High Availability cluster (see Chapter 6., “Red Hat Cluster Manager” on page 139). We also use small logical drive sizes to save on initialization time. Your requirements will likely be different in detail, but similar in many respects.

Highlight the **Unconfigured Capacity** and select **Logical Drive -> Create** from the menu.

Select **Linux** from the Default Host Type screen and click **OK** (Figure 9-11).



A host type defines how the controllers work with a host's operating system when logical drives are accessed.

Because Storage Partitioning is enabled, the default host type is used only for attached hosts that you do not specifically define in the Mappings View. If you intend to define all of your attached hosts and create storage partitions, you can change or define the host type at that time. Any hosts you don't define must be compatible with this default host type or they cannot correctly access any logical drives in the default group.

IMPORTANT
Change the host type, if required. You only need to change it once for all logical drives.

Current default host type: Linux

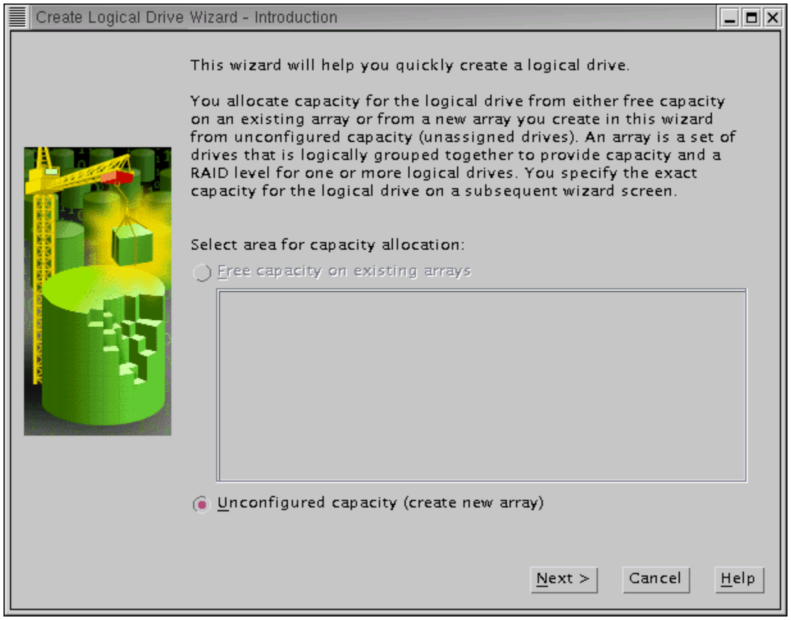
□ Don't display this dialog again.

Figure 9-11 Default Host Type screen

The Create Logical Drive wizard starts. The wizard takes you through the stages of creating arrays and logical drives. The first step is to select from the following options:

- Free capacity on existing arrays
- Unconfigured capacity (create new array)

As we have no existing arrays at the stage we selected **Unconfigured capacity,** and click **Next** (Figure 9-12).



This wizard will help you quickly create a logical drive.

You allocate capacity for the logical drive from either free capacity on an existing array or from a new array you create in this wizard from unconfigured capacity (unassigned drives). An array is a set of drives that is logically grouped together to provide capacity and a RAID level for one or more logical drives. You specify the exact capacity for the logical drive on a subsequent wizard screen.

Select area for capacity allocation:

○ Free capacity on existing arrays

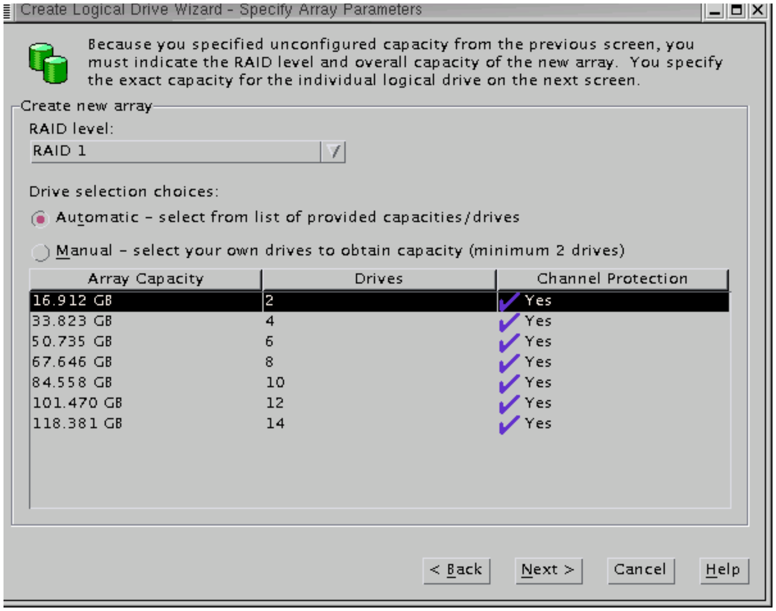
● Unconfigured capacity (create new array)

Figure 9-12 Logical Drive Wizard screen

Next, select a **RAID level**, and the number of drives to be included in the array. You also get the following choice:

- Automatic - Select from list of provided capacities/drives
- Manual - Select your own drives to obtain capacity

In most cases we recommend that you select **Automatic**. This is because the FAStT allocation heuristic does a reasonable job of distributing I/O traffic across available resources. However, manual selection may be preferable if there are but few physical disks to be allocated, or if other special circumstances warrant. With `Automatic` mode of this example,click **Next** (Figure 9-13).



Create Logical Drive Wizard - Specify Array Parameters

Because you specified unconfigured capacity from the previous screen, you must indicate the RAID level and overall capacity of the new array. You specify the exact capacity for the individual logical drive on the next screen.

Create new array

RAID level: RAID 1

Drive selection choices:

● Automatic - select from list of provided capacities/drives

○ Manual - select your own drives to obtain capacity (minimum 2 drives)

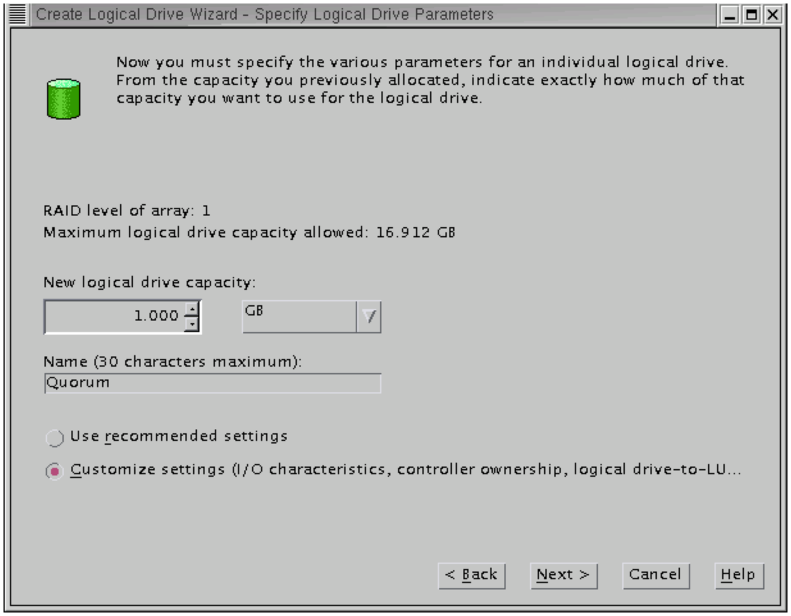
Array Capacity	Drives	Channel Protection
16.912 GB	2	✓ Yes
33.823 GB	4	✓ Yes
50.735 GB	6	✓ Yes
67.646 GB	8	✓ Yes
84.558 GB	10	✓ Yes
101.470 GB	12	✓ Yes
118.381 GB	14	✓ Yes

Figure 9-13 Create Array screen

Now, you must specify the logical drive parameters. Enter the logical drive capacity and a name. You also have the following options:

- Use recommended settings
- Customize settings (I/O characteristics, controller ownership, logical drive to LUN mappings)

Select **Customize** and click **Next** (Figure 9-14).


Create Logical Drive Wizard - Specify Logical Drive Parameters

Now you must specify the various parameters for an individual logical drive. From the capacity you previously allocated, indicate exactly how much of that capacity you want to use for the logical drive.

RAID level of array: 1
Maximum logical drive capacity allowed: 16.912 GB

New logical drive capacity:
1.000 GB

Name (30 characters maximum):
Quorum

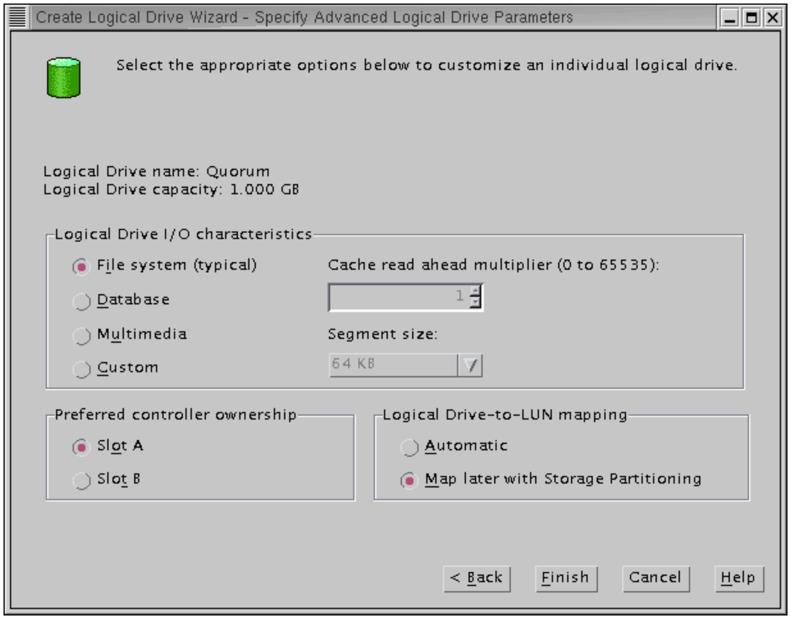
○ Use recommended settings
● Customize settings (I/O characteristics, controller ownership, logical drive-to-LU...

Figure 9-14 Logical Drive Parameters screen

You can then specify advanced logical drive parameters. You can set the drive I/O characteristics, preferred controller ownership, and drive-to-LUN mapping.

Make any changes to suit your environment. `Logical drive-to-LUN mapping` should be left at **Map later with Storage Partitioning**. Click **Finish** (Figure 9-15).

Note: Carefully consider any changes as they could dramatically affect the performance of your storage.


Create Logical Drive Wizard - Specify Advanced Logical Drive Parameters

Select the appropriate options below to customize an individual logical drive.

Logical Drive name: Quorum
Logical Drive capacity: 1.000 GB

Logical Drive I/O characteristics

● File system (typical)
○ Database
○ Multimedia
○ Custom

Cache read ahead multiplier (0 to 65535): 1

Segment size: 64 KB

Preferred controller ownership

● Slot A
○ Slot B

Logical Drive-to-LUN mapping

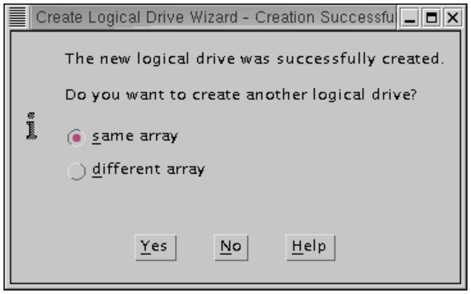
○ Automatic
● Map later with Storage Partitioning

Figure 9-15 Advanced Logical Drive Parameters screen

The next screen (Figure 9-16) prompts you: `Would you like to create another logical drive` and gives you the following options:

- Same array
- Different array

In our case, we selected **Different array** to create a RAID5 array and then selected **Same array** to create four logical drives within it (Figure 9-17).

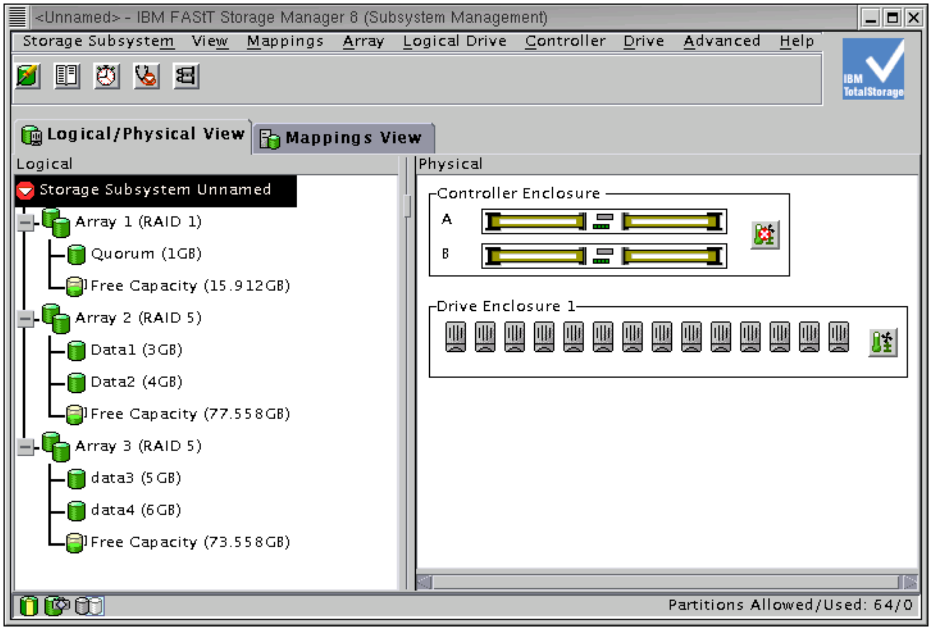

Create Logical Drive Wizard - Creation Successfu

The new logical drive was successfully created.

Do you want to create another logical drive?

● same array
○ different array

Figure 9-16 Create a New Logical Drive screen



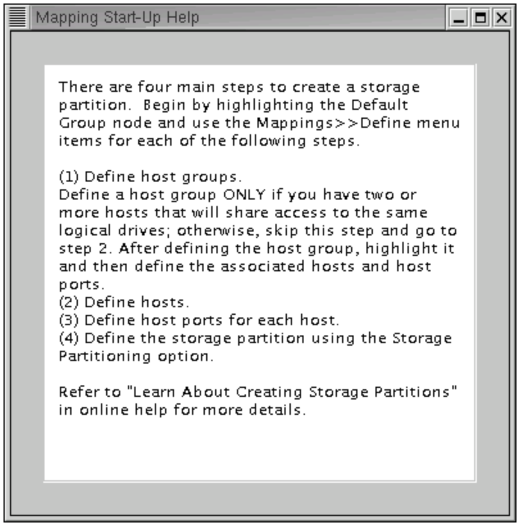
```
Storage Subsystem Unnamed
├ Array 1 (RAID 1)
│  ├ Quorum (1GB)
│  └ Free Capacity (15.912GB)
├ Array 2 (RAID 5)
│  ├ Data1 (3GB)
│  ├ Data2 (4GB)
│  └ Free Capacity (77.558GB)
└ Array 3 (RAID 5)
   ├ data3 (5GB)
   ├ data4 (6GB)
   └ Free Capacity (73.558GB)
```

Partitions Allowed/Used: 64/0

Figure 9-17 Arrays and logical drives

The systems involved in Storage Partitioning should be up and running in order to get the Host Port Identifiers. You should not use the `Default Group` for security reasons, but you should not delete this group.

Select the **Mappings View** tab in the Subsystem Management window. This displays the Mappings Startup Help message. Read the text and close the window (Figure 9-18).



Mapping Start-Up Help

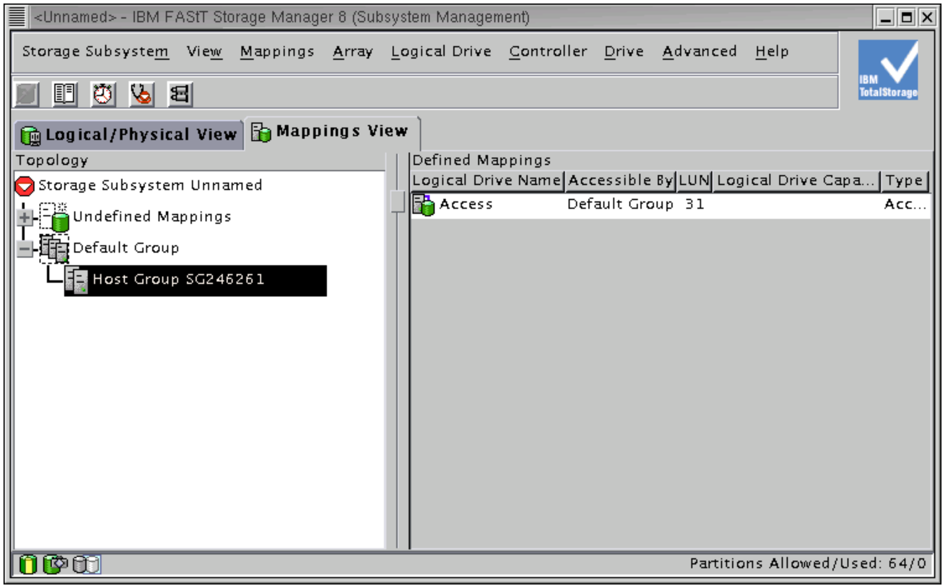
There are four main steps to create a storage partition. Begin by highlighting the Default Group node and use the Mappings>>Define menu items for each of the following steps.

(1) Define host groups.
Define a host group ONLY if you have two or more hosts that will share access to the same logical drives; otherwise, skip this step and go to step 2. After defining the host group, highlight it and then define the associated hosts and host ports.
(2) Define hosts.
(3) Define host ports for each host.
(4) Define the storage partition using the Storage Partitioning option.

Refer to "Learn About Creating Storage Partitions" in online help for more details.

Figure 9-18   Mappings Startup Help screen

Create a group for your Storage Partitioning. Select **Mappings -> Define -> Host Group**. Enter a name a click **Add** (Figure 9-19).



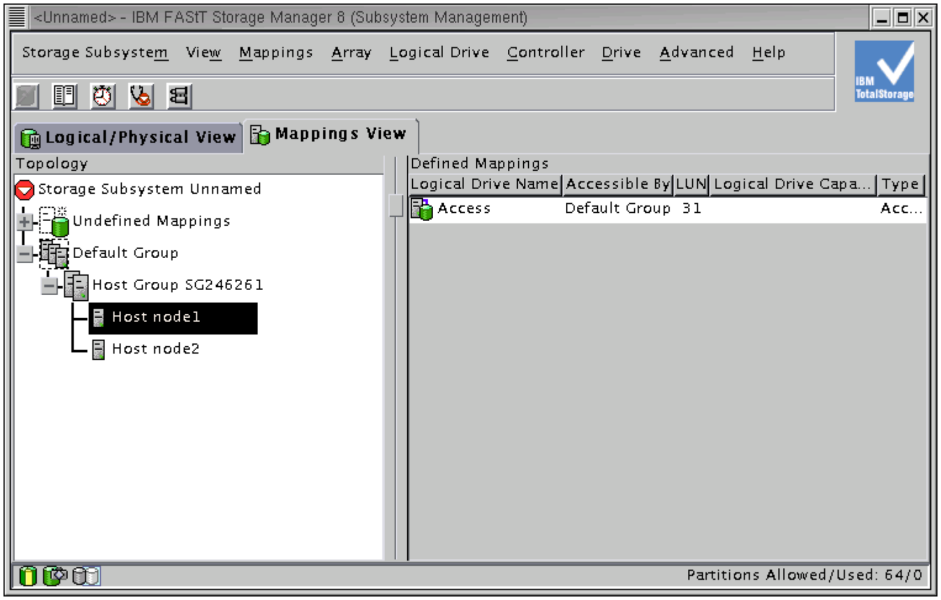
Defined Mappings

Logical Drive Name	Accessible By	LUN	Logical Drive Capa...	Type
Access	Default Group	31		Acc...

Partitions Allowed/Used: 64/0

Figure 9-19   Host Group added

Highlight the new Host Group and select **Mappings -> Define -> Host**. Enter a host name and click **Add**. Repeat this for each system, which is to be a part of this group. We are setting up a High Availability cluster, so we have two systems in the group (Figure 9-20).



Topology

- Storage Subsystem Unnamed
  - Undefined Mappings
  - Default Group
    - Host Group SG246261
      - Host node1
      - Host node2

Defined Mappings

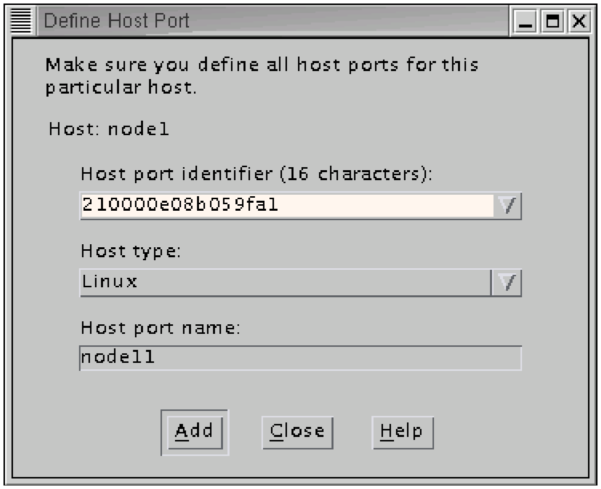
Logical Drive Name	Accessible By	LUN	Logical Drive Capa...	Type
Access	Default Group	31		Acc...

Partitions Allowed/Used: 64/0

Figure 9-20 Hosts added

Highlight the first **Host** in the new group. Select **Mappings -> Define -> Host Port**. Select the **Host Port Identifier** (collected earlier in the Fibre HBA BIOS setup) which matches the first card in this system. Select **Linux** as the `Host Type` (Figure 9-21).

> **Tip:** If the `Host Port Identifier` drop-down is empty or missing the correct entries, close Storage Manager, and restart the relevant system.



Define Host Port

Make sure you define all host ports for this particular host.

Host: node1

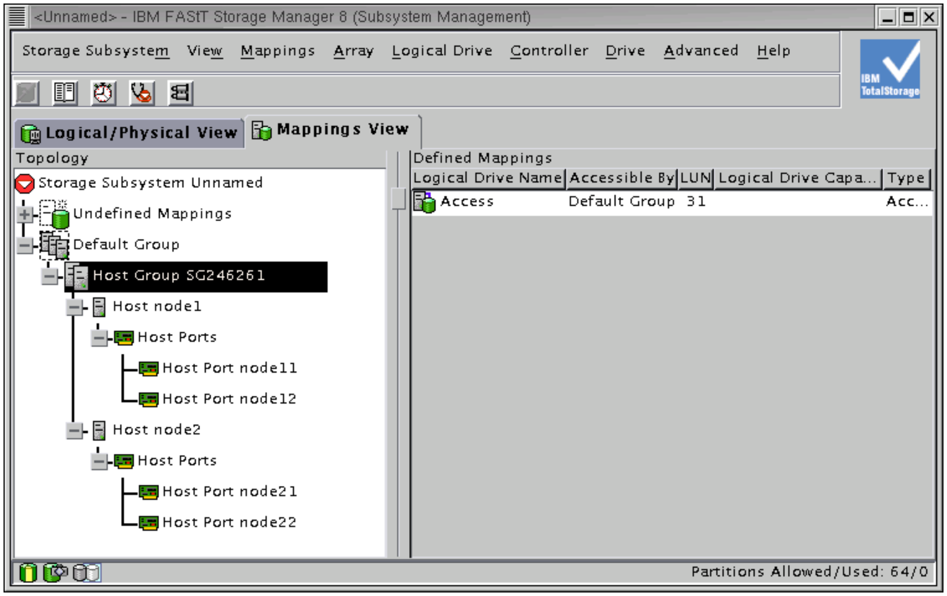
Host port identifier (16 characters):
210000e08b059fa1

Host type:
Linux

Host port name:
node11

Figure 9-21 Define Host Port screen

Repeat this process for each Fibre HBA in the system. Repeat the process again for each system in the group (Figure 9-22).



<Unnamed> - IBM FAStT Storage Manager 8 (Subsystem Management)

Topology

- Storage Subsystem Unnamed
  - Undefined Mappings
  - Default Group
    - Host Group SG246261
      - Host node1
        - Host Ports
          - Host Port node11
          - Host Port node12
      - Host node2
        - Host Ports
          - Host Port node21
          - Host Port node22

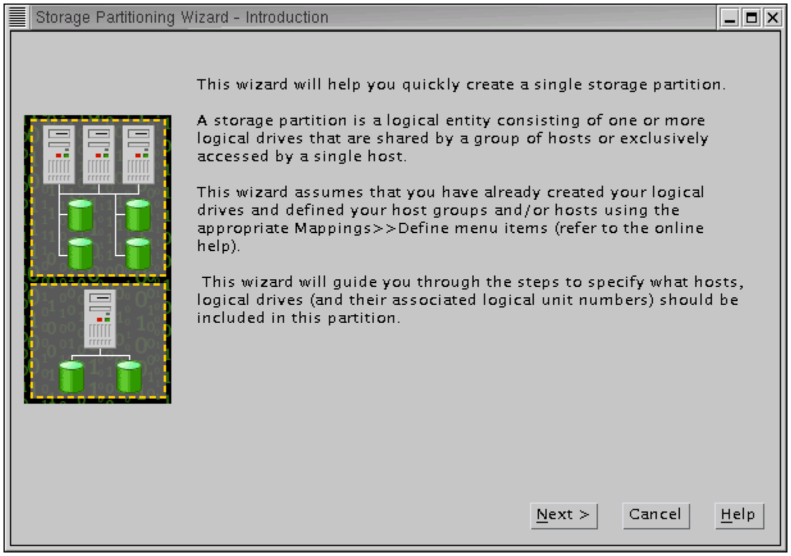
Defined Mappings

Logical Drive Name	Accessible By	LUN	Logical Drive Capa...	Type
Access	Default Group	31		Acc...

Partitions Allowed/Used: 64/0

Figure 9-22 Host Ports added

Highlight the new Host Group, select **Mappings -> Define -> Storage Partitioning**. The Storage Partitioning Wizard will appear. Read the text and click **Next** (Figure 9-23).



Storage Partitioning Wizard - Introduction

This wizard will help you quickly create a single storage partition.

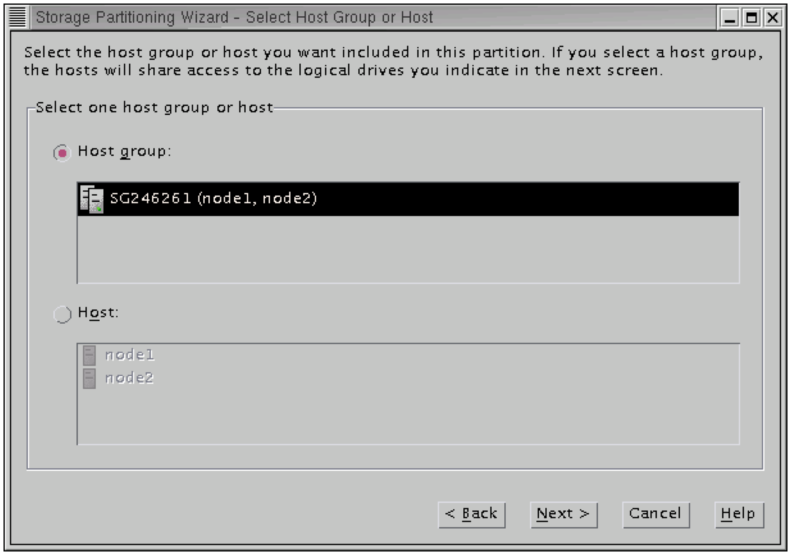
A storage partition is a logical entity consisting of one or more logical drives that are shared by a group of hosts or exclusively accessed by a single host.

This wizard assumes that you have already created your logical drives and defined your host groups and/or hosts using the appropriate Mappings>>Define menu items (refer to the online help).

This wizard will guide you through the steps to specify what hosts, logical drives (and their associated logical unit numbers) should be included in this partition.

Figure 9-23 Storage Partitioning Wizard screen

Select a **Host Group** or single **Host** for this partition. Click **Next** (Figure 9-24).



Storage Partitioning Wizard - Select Host Group or Host

Select the host group or host you want included in this partition. If you select a host group, the hosts will share access to the logical drives you indicate in the next screen.

Select one host group or host

● Host group:

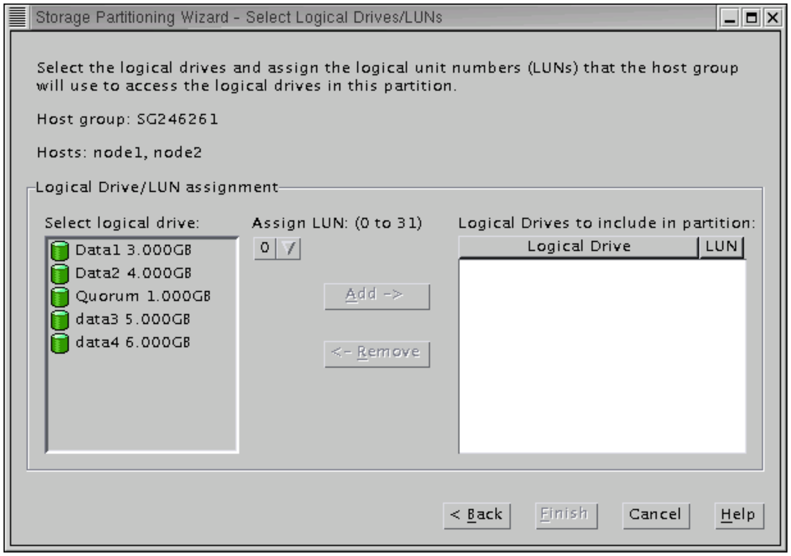
SG246261 (node1, node2)

○ Host:

node1
node2

Figure 9-24 Select Host Group or Host screen

Select the **logical drives** to include in the partition and assign LUN IDs. These IDs must start at zero and continue sequentially, you cannot miss a number out. If you later remove a LUN you must re-assign the LUN IDs to continue without skipping a number (Figure 9-25).



Storage Partitioning Wizard - Select Logical Drives/LUNs

Select the logical drives and assign the logical unit numbers (LUNs) that the host group will use to access the logical drives in this partition.

Host group: SG246261

Hosts: node1, node2

Logical Drive/LUN assignment

Select logical drive:

- Data1 3.000GB
- Data2 4.000GB
- Quorum 1.000GB
- data3 5.000GB
- data4 6.000GB

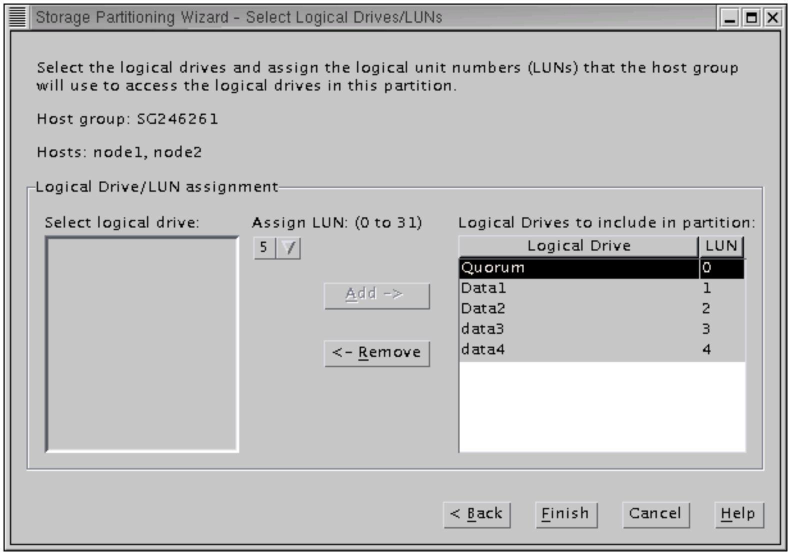
Assign LUN: (0 to 31) 0

Logical Drives to include in partition:

Logical Drive	LUN

Figure 9-25 Select Logical Drives/LUNs screen

Highlight the logical drives to be included and click **Add** (Figure 9-26). Once you have added all of the required logical drives, click **Finish**.


Storage Partitioning Wizard - Select Logical Drives/LUNs

Select the logical drives and assign the logical unit numbers (LUNs) that the host group will use to access the logical drives in this partition.

Host group: SG246261

Hosts: node1, node2

Logical Drive/LUN assignment

Select logical drive:

Assign LUN: (0 to 31) 5

Logical Drives to include in partition:

Logical Drive	LUN
Quorum	0
Data1	1
Data2	2
data3	3
data4	4

Figure 9-26 Logical drives added

The Access LUN (31) will probably be listed under `Defined Mapping` in your new Host Group. This LUN is used by some other operating systems. Delete this LUN as it may cause problems with multipathing under Linux. Highlight LUN31, press Delete, click **Yes** to confirm.

As final step you have to reboot your system. If you want an automatic mount of your volumes, you have to enter the appropriate parameters in /etc/fstab.

Important: Remember to make a copy of the FAStT profile.

For more about Storage Area Networks, please consult any of IBM's Redbooks on the topic.[6]

[6] Introduction to Storage Area Networks, SG24-5470-01 and Designing and Optimizing an IBM Storage Area Network, SG24-6419-00 are just two examples.

10

FAStT MSJ (Management Suite for Java)

This chapter contains a description of the FAStT Management Suite for Java Diagnostic and Configuration Utility. FAStT MSJ software is used to manage paths available from fibre-attached hosts to storage.

10.1 FAStT MSJ

If your installation uses the IBM FAStT Host Bus Adapter, then you will use the IBM FAStT HBA driver. If you have multiple paths to your storage, this driver can provide a failover capability, but the paths must be configured. This is the function of the IBM FAStT Management Suite Java (MSJ) Diagnostic and Configuration Utility. This tool allows you to view the current HBA configuration, attached nodes, LUNs, and paths. It will allow you to configure paths and perform some diagnostics.

Note: If you are not using IBM FAStT HBAs and the IBM FAStT driver, you can skip this part.

When using two adapters, it is necessary to hide one path to the LUN away from Linux. This is necessary because, like most OSs, Linux does not support multipathing by itself. At boot time, the adapters negotiate the preferred path that will be the only visible one. However, with the management tool, it is possible to make an adequate configuration instead of getting the results of FC protocol negotiations and driver design. Additionally, it is possible to choose the LUNs visible for the OS. Also included is static load balancing that distributes access to the individual LUNs via several paths.

FAStT MSJ has a GUI component. It also has an agent that runs on the host containing the HBAs. The agent, **`qlremote`**, is quite small but the GUI is rather large, so allow sufficient space (say 50 MB) for unpacking and installing the product.

Installing FAStT MSJ

Obtain the latest version of the IBM FAStT MSJ from the same Web location as described in FAStT Storage Manager client: Getting started.

The FAStT MSJ version is closely tied with the HBA version. Ensure that you have the correct version. Extract the gzipped tar file:

```
tar zxvf *.tgz
```

This will extract the required files to `./FAStT_MSJ/Linux`

As root, install FAStT_MSJ using the following command:

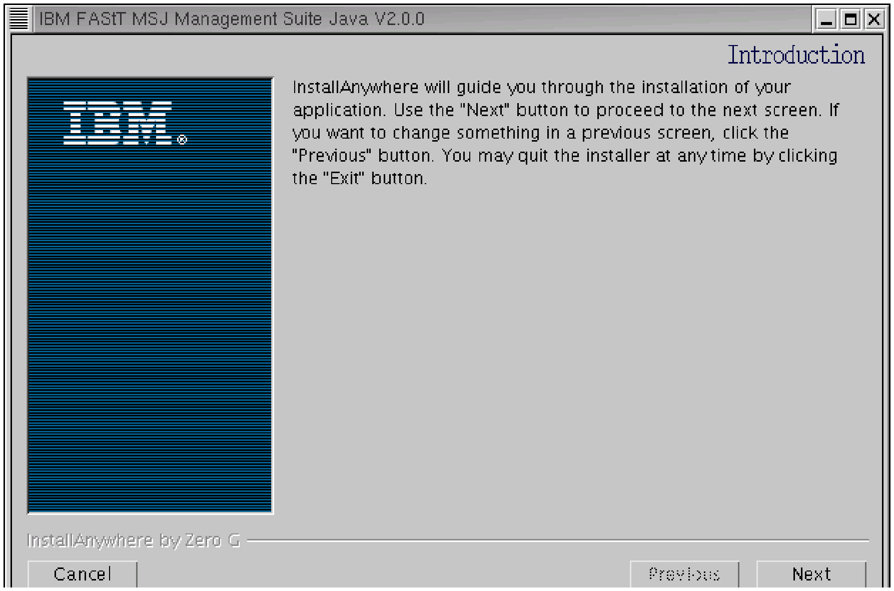
```
sh ./FAStTMSJ_install.bin
```

You should see a screen similar to Figure 10-1.

Figure 10-1 Installation Splash screen

You should then see the Introduction screen (Figure 10-2). Read the text, click **Next,** and continue through the License screen and the Readme screen.

.


IBM FAStT MSJ Management Suite Java V2.0.0
IBM

Introduction

InstallAnywhere will guide you through the installation of your application. Use the "Next" button to proceed to the next screen. If you want to change something in a previous screen, click the "Previous" button. You may quit the installer at any time by clicking the "Exit" button.

InstallAnywhere by Zero G
Cancel
Previous
Next

Figure 10-2 FAStT_MSJ Introduction screen

The next screen (Figure 10-3) allows you to choose which features of the product you wish to install. You have four choices:

- GUI and Linux Agent - This installs both on one system
- GUI - Console for remote administration
- Linux Agent - For systems to be managed remotely
- Custom - To customize the components to be installed

The default selection installs the GUI and Linux Agent. Click **Next**.

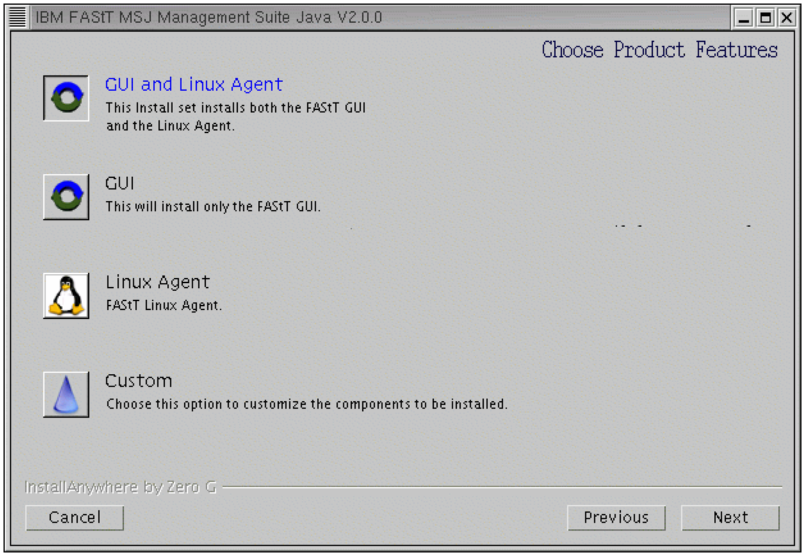

IBM FAStT MSJ Management Suite Java V2.0.0
Choose Product Features
GUI and Linux Agent
This Install set installs both the FAStT GUI and the Linux Agent.
GUI
This will install only the FAStT GUI.
Linux Agent
FAStT Linux Agent.
Custom
Choose this option to customize the components to be installed.
InstallAnywhere by Zero G
Cancel
Previous
Next

Figure 10-3 Product Features screen

Next, you are prompted for a location in which to install the software. The default folder is /opt/IBM_FAStT_MSJ. Click **Next** to accept the default, or choose a different location according to your preference. Software installation will take a few moments.

Note: During installation, a script is generated to uninstall the MSJ if desired.

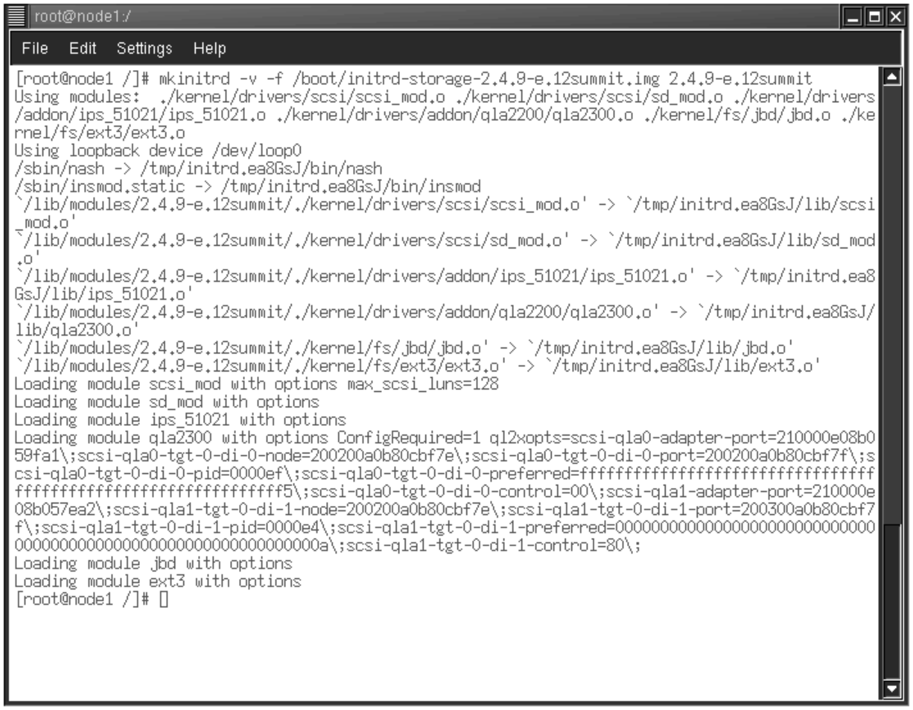
Once installation is complete, click **Done**.

10.1.1 Configuring the LUN attachment

This section introduces you to the basic steps for administering your Fibre Channel environment. Prior to describing how the settings are made, lets take a moment to understand how the fibre-attached devices (or any devices, in general) become effective. Device interfaces to the kernel are done using software components called device drivers. Linux has two different ways of integrating device drivers:

- Directly compiled to the kernel
- As loadable modules

In our work we used the second approach, and we used the FAStT drivers as loadable modules. If it is necessary to have those drivers available at boot time to launch the required scripts in time or to have access to the device during boot, they may be compiled into the kernel, or the boot loader (GRUB) can load an initial ramdisk, which contains all necessary drivers (see Figure 10-4).



```
[root@node1 /]# mkinitrd -v -f /boot/initrd-storage-2.4.9-e.12summit.img 2.4.9-e.12summit
Using modules:  ./kernel/drivers/scsi/scsi_mod.o ./kernel/drivers/scsi/sd_mod.o ./kernel/drivers
/addon/ips_51021/ips_51021.o ./kernel/drivers/addon/qla2200/qla2300.o ./kernel/fs/jbd/jbd.o ./ke
rnel/fs/ext3/ext3.o
Using loopback device /dev/loop0
/sbin/nash -> /tmp/initrd.ea8GsJ/bin/nash
/sbin/insmod.static -> /tmp/initrd.ea8GsJ/bin/insmod
`/lib/modules/2.4.9-e.12summit/./kernel/drivers/scsi/scsi_mod.o' -> `/tmp/initrd.ea8GsJ/lib/scsi
_mod.o'
`/lib/modules/2.4.9-e.12summit/./kernel/drivers/scsi/sd_mod.o' -> `/tmp/initrd.ea8GsJ/lib/sd_mod
.o'
`/lib/modules/2.4.9-e.12summit/./kernel/drivers/addon/ips_51021/ips_51021.o' -> `/tmp/initrd.ea8
GsJ/lib/ips_51021.o'
`/lib/modules/2.4.9-e.12summit/./kernel/drivers/addon/qla2200/qla2300.o' -> `/tmp/initrd.ea8GsJ/
lib/qla2300.o'
`/lib/modules/2.4.9-e.12summit/./kernel/fs/jbd/jbd.o' -> `/tmp/initrd.ea8GsJ/lib/jbd.o'
`/lib/modules/2.4.9-e.12summit/./kernel/fs/ext3/ext3.o' -> `/tmp/initrd.ea8GsJ/lib/ext3.o'
Loading module scsi_mod with options max_scsi_luns=128
Loading module sd_mod with options
Loading module ips_51021 with options
Loading module qla2300 with options ConfigRequired=1 ql2xopts=scsi-qla0-adapter-port=210000e08b0
59fa1\;scsi-qla0-tgt-0-di-0-node=200200a0b80cbf7e\;scsi-qla0-tgt-0-di-0-port=200200a0b80cbf7f\;s
csi-qla0-tgt-0-di-0-pid=0000ef\;scsi-qla0-tgt-0-di-0-preferred=fffffffffffffffffffffffffffffffff
ffffffffffffffffffffffffffffff5\;scsi-qla0-tgt-0-di-0-control=00\;scsi-qla1-adapter-port=210000e
08b057ea2\;scsi-qla1-tgt-0-di-1-node=200200a0b80cbf7e\;scsi-qla1-tgt-0-di-1-port=200300a0b80cbf7
f\;scsi-qla1-tgt-0-di-1-pid=0000e4\;scsi-qla1-tgt-0-di-1-preferred=00000000000000000000000000000
0000000000000000000000000000000000a\;scsi-qla1-tgt-0-di-1-control=80\;
Loading module jbd with options
Loading module ext3 with options
[root@node1 /]# []
```

Figure 10-4 Building a ramdisk image

You can see that **mkinitrd** scans `modules.conf` for the required modules, including all options for the particular modules. As you can see in this case, the drivers for the network, the ServeRAID, the Fibre Channel, and other modules are included. In addition, a long options list (generated by FAStT MSJ, we will see how to do this in 10.1.2, "Use FAStT MSJ to configure Fibre Channel paths" on page 216) is submitted, too. These options contain all the necessary path and LUN information to set up the FAStT multipath driver during boot. This leads to the following results:

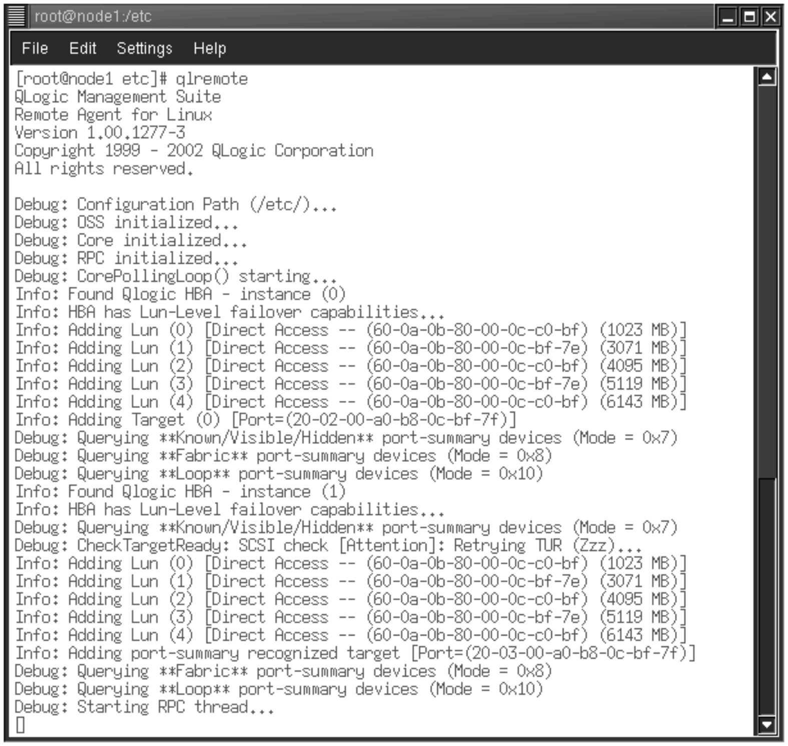
- Every time a change to your configuration occurs, you have to generate a new initial ramdisk and reboot (or unload and reload the Fibre driver).
- It is quite useful to keep an additional bootable ramdisk with no options for the FAStT HBA in reserve. After booting this ramdisk, the driver detects attached devices and sets preferred paths, but these are not overruled by the options. Then, with MSJ, you can generate a new, error free option string.

> **Note:** Any change in the configuration of your SAN causes you to reconfigure your paths and create a new initial ramdisk.

10.1.2 Use FAStT MSJ to configure Fibre Channel paths

In this section we set up the preferred and alternate paths for the LUNs.

The Linux agent is **`qlremote`**. This has to be running on the host before you open the GUI client software. Open a terminal window and run **`qlremote`** (Figure 10-5).



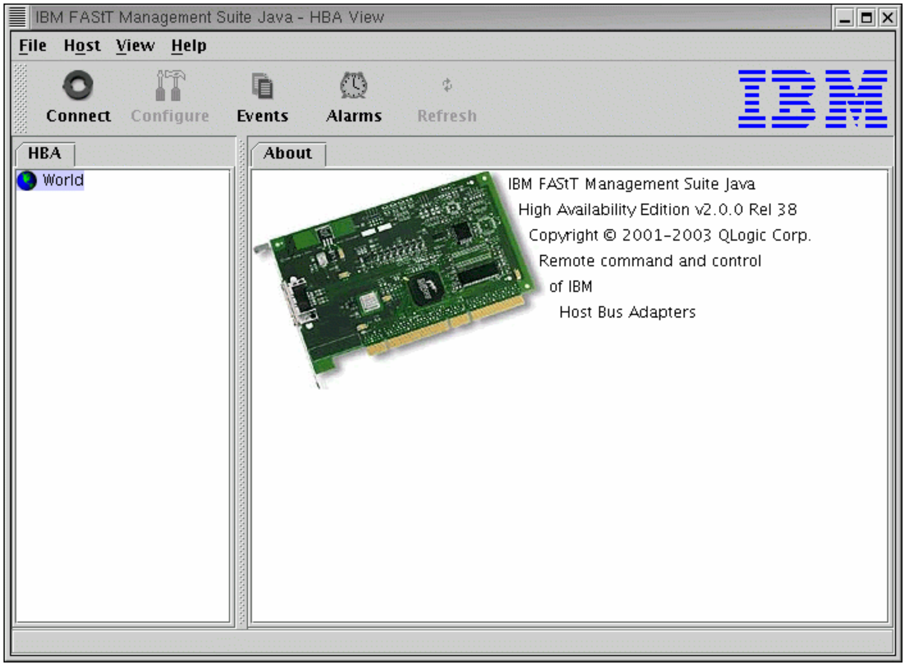
```
[root@node1 etc]# qlremote
QLogic Management Suite
Remote Agent for Linux
Version 1.00.1277-3
Copyright 1999 - 2002 QLogic Corporation
All rights reserved.

Debug: Configuration Path (/etc/)...
Debug: OSS initialized...
Debug: Core initialized...
Debug: RPC initialized...
Debug: CorePollingLoop() starting...
Info: Found Qlogic HBA - instance (0)
Info: HBA has Lun-Level failover capabilities...
Info: Adding Lun (0) [Direct Access -- (60-0a-0b-80-00-0c-c0-bf) (1023 MB)]
Info: Adding Lun (1) [Direct Access -- (60-0a-0b-80-00-0c-bf-7e) (3071 MB)]
Info: Adding Lun (2) [Direct Access -- (60-0a-0b-80-00-0c-c0-bf) (4095 MB)]
Info: Adding Lun (3) [Direct Access -- (60-0a-0b-80-00-0c-bf-7e) (5119 MB)]
Info: Adding Lun (4) [Direct Access -- (60-0a-0b-80-00-0c-c0-bf) (6143 MB)]
Info: Adding Target (0) [Port=(20-02-00-a0-b8-0c-bf-7f)]
Debug: Querying **Known/Visible/Hidden** port-summary devices (Mode = 0x7)
Debug: Querying **Fabric** port-summary devices (Mode = 0x8)
Debug: Querying **Loop** port-summary devices (Mode = 0x10)
Info: Found Qlogic HBA - instance (1)
Info: HBA has Lun-Level failover capabilities...
Debug: Querying **Known/Visible/Hidden** port-summary devices (Mode = 0x7)
Debug: CheckTargetReady: SCSI check [Attention]: Retrying TUR (Zzz)...
Info: Adding Lun (0) [Direct Access -- (60-0a-0b-80-00-0c-c0-bf) (1023 MB)]
Info: Adding Lun (1) [Direct Access -- (60-0a-0b-80-00-0c-bf-7e) (3071 MB)]
Info: Adding Lun (2) [Direct Access -- (60-0a-0b-80-00-0c-c0-bf) (4095 MB)]
Info: Adding Lun (3) [Direct Access -- (60-0a-0b-80-00-0c-bf-7e) (5119 MB)]
Info: Adding Lun (4) [Direct Access -- (60-0a-0b-80-00-0c-c0-bf) (6143 MB)]
Info: Adding port-summary recognized target [Port=(20-03-00-a0-b8-0c-bf-7f)]
Debug: Querying **Fabric** port-summary devices (Mode = 0x8)
Debug: Querying **Loop** port-summary devices (Mode = 0x10)
Debug: Starting RPC thread...
▯
```

Figure 10-5 Example of qlremote running

If the MSJ software is installed on the same host, leave this to run and open another terminal window. Otherwise, open a terminal window on the management workstation. In either case, type the following to run FAStT MSJ client:

```
/opt/IBM_FAStT_MSJ/FAStT
```

When the FAStT MSJ is launched a screen similar to Figure 10-6 should appear.

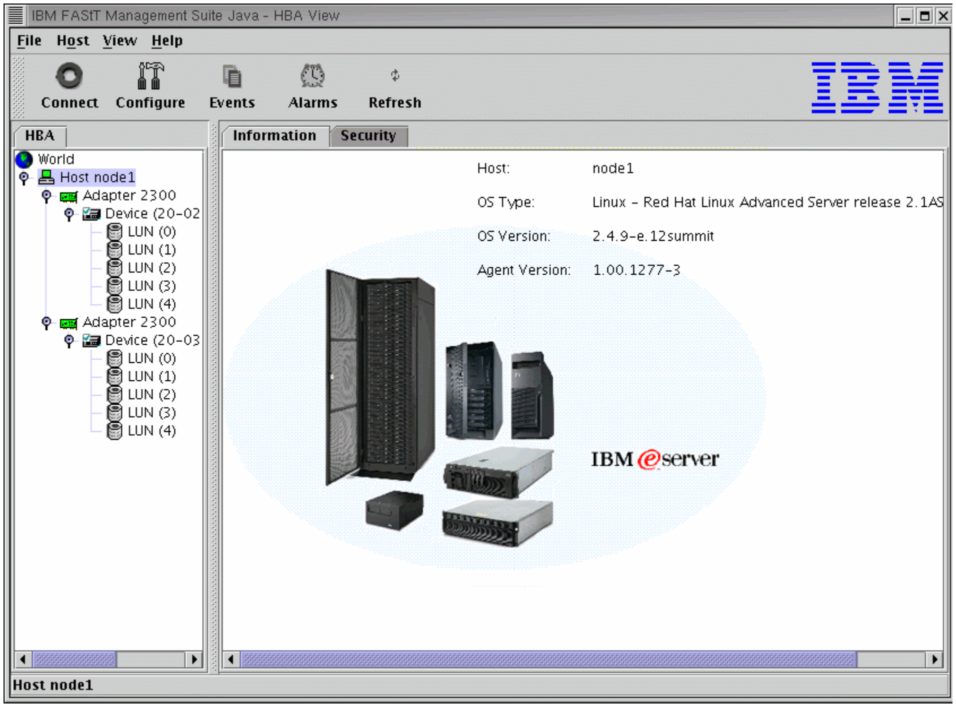


IBM FAStT Management Suite Java
High Availability Edition v2.0.0 Rel 38
Copyright © 2001-2003 QLogic Corp.
Remote command and control
of IBM
Host Bus Adapters

Figure 10-6 FAStT_MSJ screen

Click **Connect** to connect to the agent you wish to administer. Enter the IP address of the system where the **qlremote** agent is running and click **Connect**. You should see a screen similar to Figure 10-7 showing you storage and LUNs.

The management system's /etc/hosts file or the site DNS services should be up to date as this program will try to resolve a hostname to the IP address. You will get errors if this is not correct.



Host: node1

OS Type: Linux - Red Hat Linux Advanced Server release 2.1AS

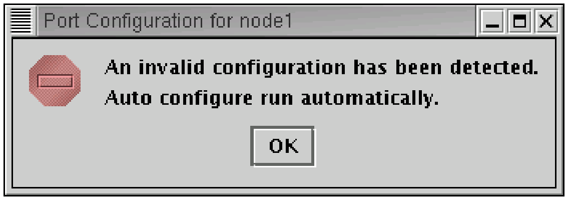
OS Version: 2.4.9-e.12summit

Agent Version: 1.00.1277-3

Figure 10-7 FAStT_MSJ connected

After connecting, the FAStT MSJ displays the adapters installed in the specific host. In addition, you can see the Fibre Channel nodes attached to the fibre HBAs.

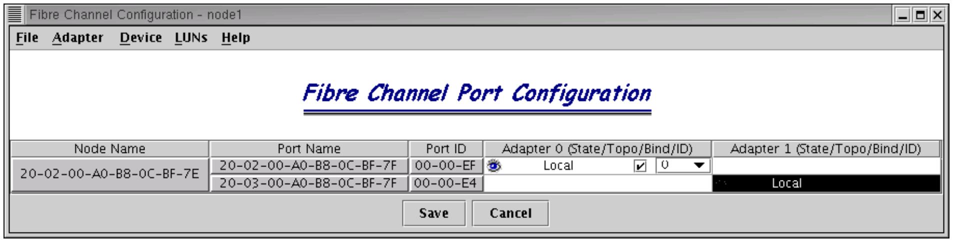
To configure your environment, highlight either the host machine or an adapter and click **Configure**. You will receive an error message (Figure 10-8). This is because there is no configuration setup yet. Click **OK** and continue.



An invalid configuration has been detected.
Auto configure run automatically.

Figure 10-8 Invalid Configuration screen

You should then see a screen similar to Figure 10-9. If your system is connected through an external switch, the labels will be slightly different (for example, displaying “Fabric” instead of “Local”).


Fibre Channel Configuration - node1
File Adapter Device LUNs Help
Save
Cancel

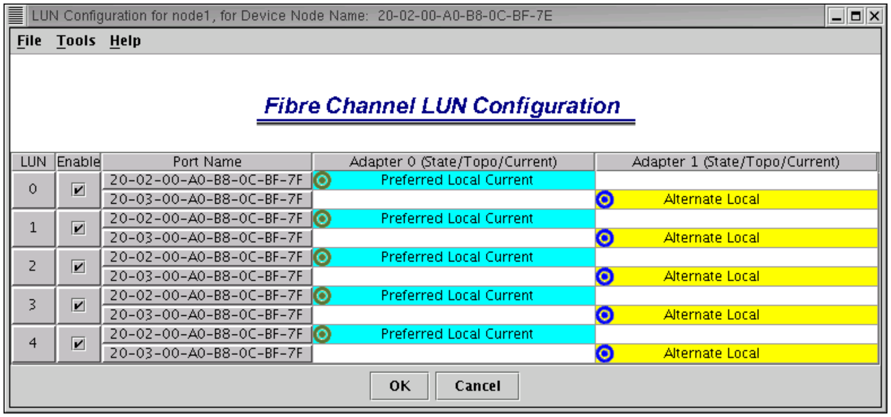
Fibre Channel Port Configuration

Node Name	Port Name	Port ID	Adapter 0 (State/Topo/Bind/ID)	Adapter 1 (State/Topo/Bind/ID)
20-02-00-A0-B8-0C-BF-7E	20-02-00-A0-B8-0C-BF-7F	00-00-EF	Local ☑ 0	
	20-03-00-A0-B8-0C-BF-7F	00-00-E4		Local

Figure 10-9 Port Configuration screen

For the easiest way to configure the ports, click **Device -> AutoConfigure**. You will receive a message asking: `Also configure LUNs` After clicking **Yes**, the preferred path is shown in white, while the hidden paths are shown in black. Hidden means that these paths are not visible to the operating system. Because the multipath driver is handling all I/O, they can be used as redundant paths without interfering with the operating system view.

Before balancing the LUNs, the configuration will look something like Figure 10-10. Select **Configure**, highlight the Node Name, select **Device -> Configure LUNs**. You will see the preferred paths are highlighted in blue, while the alternate paths are yellow. Also, you can see the preferred paths are marked with a green circle, because this matches the settings that the drivers did at boot time (and hence, the paths are "seen" by Linux).


LUN Configuration for node1, for Device Node Name: 20-02-00-A0-B8-0C-BF-7E
File Tools Help
OK
Cancel

Fibre Channel LUN Configuration

LUN	Enable	Port Name	Adapter 0 (State/Topo/Current)	Adapter 1 (State/Topo/Current)
0	☑	20-02-00-A0-B8-0C-BF-7F	● Preferred Local Current	
		20-03-00-A0-B8-0C-BF-7F		● Alternate Local
1	☑	20-02-00-A0-B8-0C-BF-7F	● Preferred Local Current	
		20-03-00-A0-B8-0C-BF-7F		● Alternate Local
2	☑	20-02-00-A0-B8-0C-BF-7F	● Preferred Local Current	
		20-03-00-A0-B8-0C-BF-7F		● Alternate Local
3	☑	20-02-00-A0-B8-0C-BF-7F	● Preferred Local Current	
		20-03-00-A0-B8-0C-BF-7F		● Alternate Local
4	☑	20-02-00-A0-B8-0C-BF-7F	● Preferred Local Current	
		20-03-00-A0-B8-0C-BF-7F		● Alternate Local

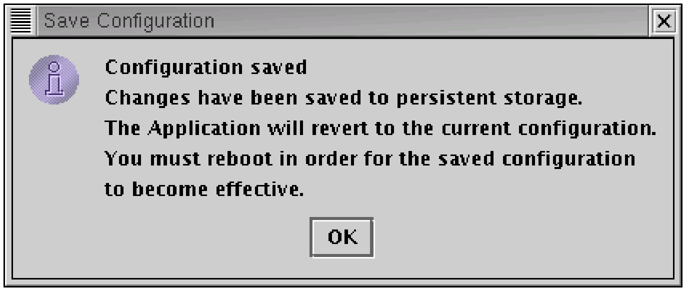
Figure 10-10 Paths before balancing LUNs

Load balancing by path allocation

Select **LUNs -> Load Balance -> All LUNs**. When prompted to accept the configuration and click **Yes**.

Click **Save**, enter the password, and click **OK.**

A screen will pop-up (Figure 10-11) informing you that the configuration has been saved and you must reboot for the changes to take effect. Of course, this reboot requirement refers to the system containing the HBAs.

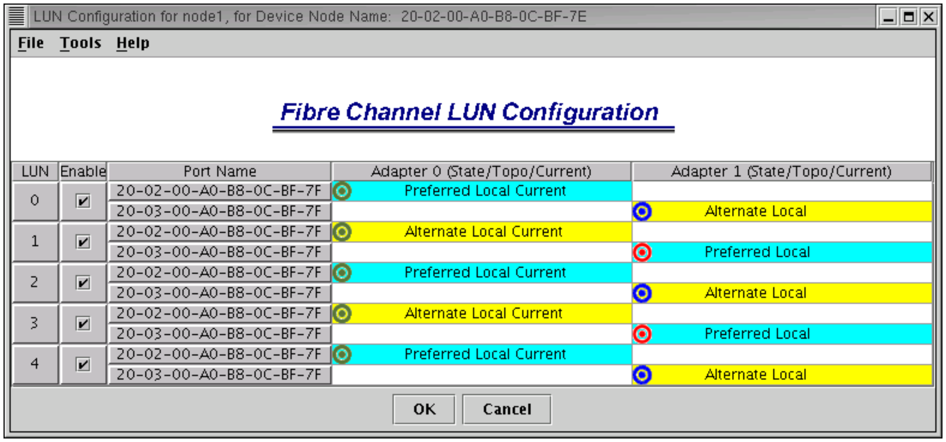


Save Configuration

Configuration saved
Changes have been saved to persistent storage.
The Application will revert to the current configuration.
You must reboot in order for the saved configuration
to become effective.

Figure 10-11   Configuration Saved screen

The current status of the paths will look something like Figure 10-12, the red circles indicate the new preferred path. This is marked red as you cannot change the paths during runtime.



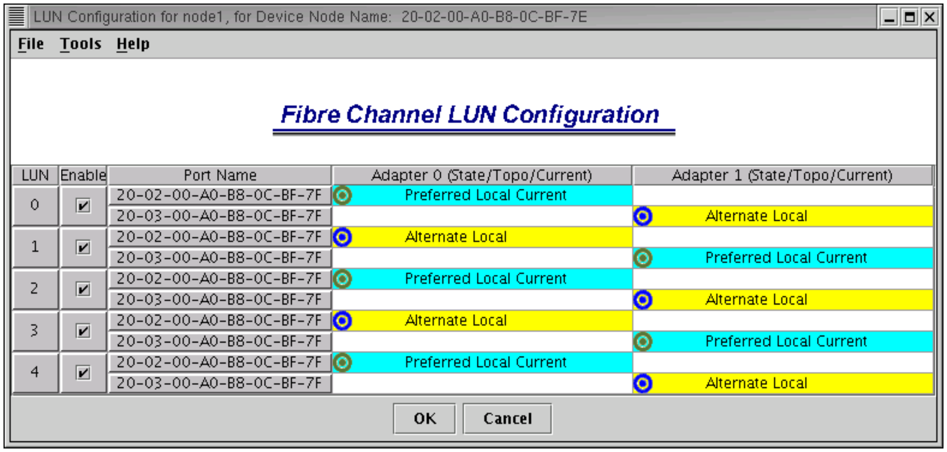
LUN Configuration for node1, for Device Node Name: 20-02-00-A0-B8-0C-BF-7E

Fibre Channel LUN Configuration

LUN	Enable	Port Name	Adapter 0 (State/Topo/Current)	Adapter 1 (State/Topo/Current)
0	☑	20-02-00-A0-B8-0C-BF-7F	◎ Preferred Local Current	
		20-03-00-A0-B8-0C-BF-7F		◎ Alternate Local
1	☑	20-02-00-A0-B8-0C-BF-7F	◎ Alternate Local Current	
		20-03-00-A0-B8-0C-BF-7F		◎ Preferred Local
2	☑	20-02-00-A0-B8-0C-BF-7F	◎ Preferred Local Current	
		20-03-00-A0-B8-0C-BF-7F		◎ Alternate Local
3	☑	20-02-00-A0-B8-0C-BF-7F	◎ Alternate Local Current	
		20-03-00-A0-B8-0C-BF-7F		◎ Preferred Local
4	☑	20-02-00-A0-B8-0C-BF-7F	◎ Preferred Local Current	
		20-03-00-A0-B8-0C-BF-7F		◎ Alternate Local

Figure 10-12   Balanced paths before rebooting

Once you have rebooted, you can launch the FAStT MSJ again and view how the paths have been configured. (Figure 10-13).


LUN Configuration for node1, for Device Node Name: 20-02-00-A0-B8-0C-BF-7E
File
Tools
Help
Fibre Channel LUN Configuration
OK
Cancel

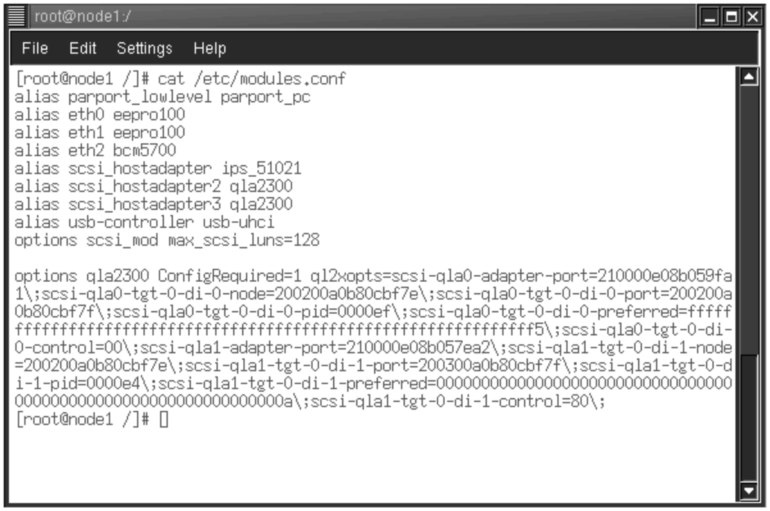
LUN	Enable	Port Name	Adapter 0 (State/Topo/Current)	Adapter 1 (State/Topo/Current)
0	☑	20-02-00-A0-B8-0C-BF-7F	◎ Preferred Local Current	
		20-03-00-A0-B8-0C-BF-7F		◎ Alternate Local
1	☑	20-02-00-A0-B8-0C-BF-7F	◎ Alternate Local	
		20-03-00-A0-B8-0C-BF-7F		◎ Preferred Local Current
2	☑	20-02-00-A0-B8-0C-BF-7F	◎ Preferred Local Current	
		20-03-00-A0-B8-0C-BF-7F		◎ Alternate Local
3	☑	20-02-00-A0-B8-0C-BF-7F	◎ Alternate Local	
		20-03-00-A0-B8-0C-BF-7F		◎ Preferred Local Current
4	☑	20-02-00-A0-B8-0C-BF-7F	◎ Preferred Local Current	
		20-03-00-A0-B8-0C-BF-7F		◎ Alternate Local

Figure 10-13   Balanced paths after rebooting

Exit from FAStT MSJ, click **File -> Exit**. Stop qlremote on the host system by pressing Ctrl-C in the correct terminal window or type the following:

```
killall -TERM qlremote
```

If you now view `/etc/modules.conf` you will notice that FAStT MSJ has added a large options string. This string is used by the Fibre HBA drivers (Figure 10-14).



```
[root@node1 /]# cat /etc/modules.conf
alias parport_lowlevel parport_pc
alias eth0 eepro100
alias eth1 eepro100
alias eth2 bcm5700
alias scsi_hostadapter ips_51021
alias scsi_hostadapter2 qla2300
alias scsi_hostadapter3 qla2300
alias usb-controller usb-uhci
options scsi_mod max_scsi_luns=128

options qla2300 ConfigRequired=1 ql2xopts=scsi-qla0-adapter-port=210000e08b059fa
1\;scsi-qla0-tgt-0-di-0-node=200200a0b80cbf7e\;scsi-qla0-tgt-0-di-0-port=200200a
0b80cbf7f\;scsi-qla0-tgt-0-di-0-pid=0000ef\;scsi-qla0-tgt-0-di-0-preferred=fffff
ffffffffffffffffffffffffffffffffffffffffffffffffffffffffff5\;scsi-qla0-tgt-0-di-
0-control=00\;scsi-qla1-adapter-port=210000e08b057ea2\;scsi-qla1-tgt-0-di-1-node
=200200a0b80cbf7e\;scsi-qla1-tgt-0-di-1-port=200300a0b80cbf7f\;scsi-qla1-tgt-0-d
i-1-pid=0000e4\;scsi-qla1-tgt-0-di-1-preferred=000000000000000000000000000000000
000000000000000000000000000000a\;scsi-qla1-tgt-0-di-1-control=80\;
[root@node1 /]# []
```

Figure 10-14   Example of modules.conf

As changes have been made to `modules.conf` we need to rebuild the ramdisk. First run **`depmod -a`** to update the `modules.dep file`.

Build the ramdisk as before:

```
mkinitrd -f /boot/<newname>.img <kernel-version>
```

The **-f** option allows you to build a ramdisk to a name that already exists, i.e overwrite it. In our case we used the following command with Red Hat Advanced Server (your naming convention may vary from this example):

```
mkinitrd -f /boot/initrd-storage-2.4.9-e.12summit.img 2.4.9-e.12summit
```

If you used a new name for the ramdisk, update /etc/grub and reboot the system.

10.1.3 Using FAStT as storage: A summary

As described before, you must set up your storage before building the initial ramdisk. Otherwise, the module will be loaded without a path configuration and no manual load balancing will be possible.

First you install the required management software. Install (at least) qlremote from the FAStT MSJ package on the host. After you have configured your storage, launch the qlremote agent and access it with the MSJ (either locally or from a management station). After you have set up your paths, MSJ instructs qlremote to write an option string to the modules.conf, which is used to configure the driver module during system boot.

Note: Run **qlremote** *only* during the usage of FAStT MSJ. Once you are done and the output string is written to the modules.conf, stop **qlremote** immediately!

You can verify your setup by unloading the module with **rmmod qla2x00** and reload with **modprobe qla2x00**. During this load the module will use the option string carrying the path definitions and activate them. Then start **qlremote** again and check the setting with FAStT MSJ.

SuSE notes

Before you build the new ramdisk you have to make the following change to the file /etc/sysconfig/kernel: add qla2200 or qla2300 to the string behind INITRD_MODULES, depending on your HBA designation:

```
INITRD_MODULES=”aic7xxx reiserfs qla2300”
```

The script /sbin/mk_initrd uses this string as input. By running **mk_initrd** a new version of the initial ramdisk is built in the /boot directory. Even though this is convenient you might build a new ramdisk instead of replacing the existing one. Have a look at the mk_initrd script, it offers you quite some useful (and sometimes colorful) options.

Red Hat notes

Before you restart you should add the kernel parameter max_scsi_luns=128 to the modules.conf file. This is necessary because the kernel does not probe the full range of LUNs by default. However, this is only critical if you are using a large amount of SAN volumes.

Note: Every change in the configuration of your SAN causes the need to reconfigure your paths and rebuild the initial ramdisk!

For more about Storage Area Networks, please consult any of IBM's several Redbooks on the topic.[1]

As final step, you have to reboot your system. If you want an automatic mount of your volumes, you have to enter the appropriate parameters in /etc/fstab.

Important: Keep in mind that the failure of a volume prior or during reboot may cause a change in the SCSI device numbering sequence.

[1] Introduction to Storage Area Networks, SG24-5470-01 and Designing and Optimizing an IBM Storage Area Network, SG24-6419-00 are just two examples.

11

Implementing VMware

VMware ESX Server is virtualization software that enables the deployment of multiple, secure, independent virtual machines on a single physical server. The purpose of this software is to provide an efficient and high performance platform for consolidation and accelerated deployment of services.

This chapter starts with a brief overview of VMware and the VMware ESX architecture, then focuses on the practical aspects of implementing VMware ESX 1.5.2 server on IBM eServer xSeries.

11.1 Introduction

VMware, Inc. was founded in 1998 to bring mainframe-class virtual machines to industry-standard computers. It offers the following products:

- VMware Workstation: This product allows you to run multiple operating systems and their applications simultaneously on a single Intel based PC.
- VMware GSX server: GSX runs in a server OS, like Linux or Windows Server Editions; it transforms the physical computer into a pool of virtual machines. Operating systems and applications are isolated in multiple virtual machines that reside on a single piece of hardware. System resources are allocated to any virtual machine based on need, delivering maximum capacity utilization and control over computing infrastructure. It can run up to 64 virtual machines (provided that the server hardware can sustain the load).
- VMware ESX Server: Is virtualization software that enables the deployment of multiple, secure, independent virtual machines on a single physical server, as illustrated in Figure 11-1. It runs directly on the hardware (in contrast to VMware Workstation and GSX Server products, which utilize host operating system to access hardware) to provide a secure, uniform platform for easy deployment, management, and remote control of other operating systems.

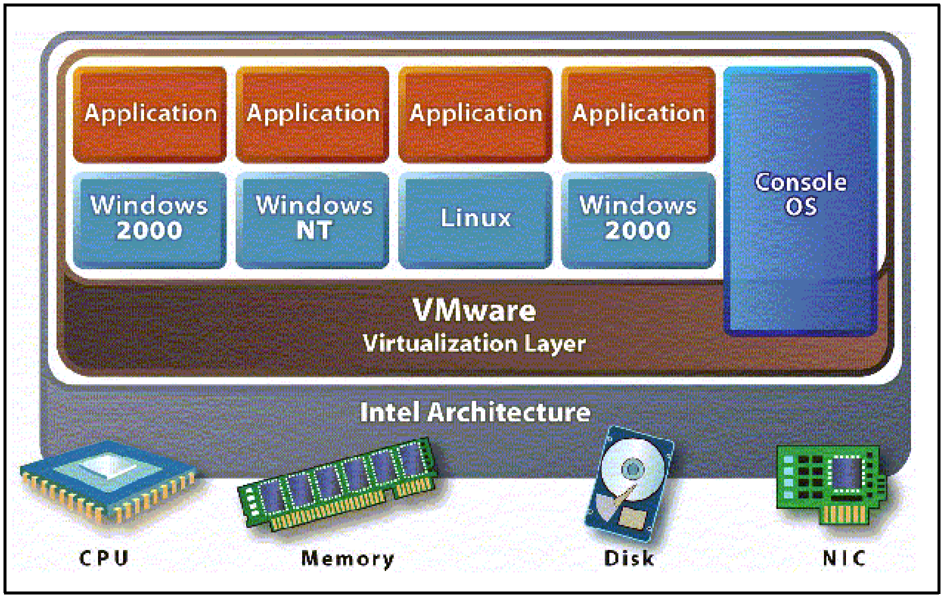


Figure 11-1 VMWare ESX server

The VMware ESX server can be used as platform for server consolidation. Many Intel based server are often used for services that do not fully utilize the capacity of the machine. Most systems run with an average system load of less than 25% of that capacity. Servers like the x440 with up to 16 processors offer enough power and scalability to move such services on one system.

Because VMware emulates the hardware, different operating systems, and the applications they run are presented with a consistent set of virtual hardware by each VM, irrespective of the physical hardware available in the system. VMs have no dependency on the physical hardware, and as such do not require device drivers specific to the physical hardware. This improves the mobility of a VM. An installation done in a virtual machine can be moved to another server.

11.1.1 VMware ESX architecture

The design of the VMware ESX Server core architecture implements the abstractions that allow hardware resources to be allocated to multiple workloads in fully isolated environments.

The key elements of the system design are:

- The VMware virtualization layer, which provides the idealized hardware environment and virtualization of underlying physical resources.
- The resource manager, which enables the partitioning and guaranteed delivery of CPU, memory, network bandwidth, and disk bandwidth to each virtual machine.
- The hardware interface components, including device drivers, which enable hardware-specific service delivery while hiding hardware differences from other parts of the system.

VMware ESX Server incorporates a resource manager and a Linux based service console that provides bootstrapping, management, and other services.

While the service console is basically used for administration purposes, the resource manager orchestrates access to physical resources and maps them to specific VMs. For instance, physical disks are virtualized. Virtual disks are created as regular files in the VMFS file system, and assigned to the VMs. These virtual disks are presented to the VMs as SCSI drives connected to a SCSI adapter.

This is the only disk storage controller used by the guest operating system, despite the wide variety of SCSI, RAID, and Fibre Channel adapters that might actually be used in the system.

11.1.2 IBM and VMware

Some milestones in excellent IBM and VMware cooperation that help customers to succeed with server consolidation projects include:

- December 2000: IBM becomes member of VMware Preferred Hardware Partner Program.
- December 2001: IBM certifies VMware ESX Server for IBM Server Proven Program.
- February 2002: IBM and VMware enter joint development agreement.
- August 2002: IBM announces xSeries + ESX Server bundles (with optional IBM support).

The IBM Server Proven Program helps to ensure customers that a whole proposed solution based on IBM technologies, including parts from other partners, will work together smoothly without any compatibility issues.

IBM offers outstanding service and support for VMware ESX products. Support packages are available through IBM Global Services that cover the entire solution including the hardware, VMware ESX Server, supported guest operating systems, and even the applications in most cases.

For more information about IBM and VMware relationship and VMware support on IBM eServer xSeries servers follow the listed URLs:

```
http://www.pc.ibm.com/ww/eserver/xseries/vmware.html
http://www.pc.ibm.com/ww/compat/nos/vmware.html
```

11.1.3 VMware system requirements

Table 11-1 on page 229 summarizes the VMware ESX 1.5.2 system requirements in regards to the server hardware and device support.

Consult `http://www.pc.ibm.com/us/compat/nos/matrix.shtml` for certification status.

Please visit `http://www.vmware.com` and `http://www.vmware.com/support` (restricted access) for supported guest operating systems.

Additionally, refer to the *VMware ESX Server User's Manual,* MIGR-44890 for details about the installation and setup procedures.

You can obtain an installation guide (document ID MIGR-44890) from: `http://www.ibm.com/support`

Table 11-1 Server requirements

Server hardware	Server device support
- **Processor**: – Intel Pentium II,III, 4 - **Minimum system RAM** – 512 MB - **Minimum disk space** – 2GB ++1GB per virtual machine - **Minimum networking** – 2 Network Interface Ports	- **Local area networking** – Broadcom 5700 Series Gigabit Ethernet – Intel e1000 Gigabit Ethernet – Common 10//100 Cards - **Direct attached storage** – IBM ServeRAID – HP SmartArray – Dell Perc – Common SCSI Adaptors - **Storage Area Networking** – Emulex 8000,9000 series – QLogic 2200,2300 series

11.2 Configuration

We used the following configuration for the purpose of this redbook:

Server:

IBM xSeries 330 with the following features:

- Two INTEL Xeon 1.1 GHz
- 2 GB of memory
- Two IBM FC2-133 Fibre Channel host adapter
- Two 18 GB hard disk drives

Storage:

- IBM SAN switch 2109-F16
- FAStT700
- One EXP700 with four HDDs

11.3 Implementing VMware ESX Server 1.5.2

This section describes the installation and configuration of the VMware ESX server.

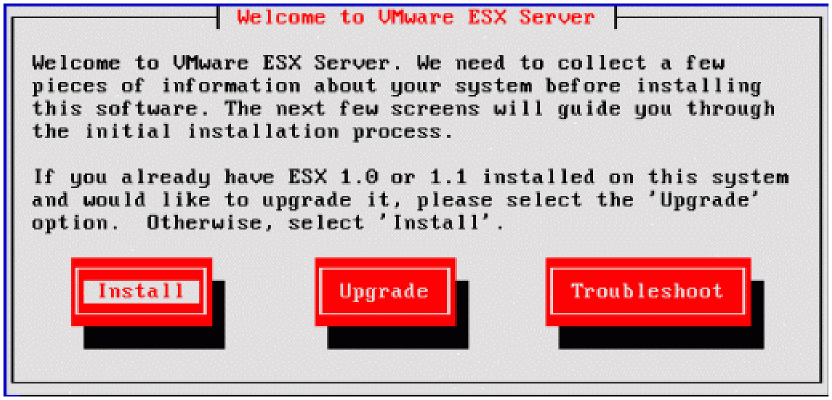
11.3.1 Installation steps

The installation of VMware is rather simple.

To begin the installation insert the VMware ESX CD and start up the server. The dialog shown in Figure 11-2 displays. Click **Install**.

> **Attention:** If you are planning to install with redundant paths you must disconnect all Fibre Channel HBAs from the fabric. Please be aware that loops are not supported by the QLogic device driver used with VMware ESX server. The VMware kernel supports up to 7 LUNs by default.
>
> However, you can set up your storage completely. Please refer to Chapter 9., “FAStT Storage Manager” on page 187 for further information.

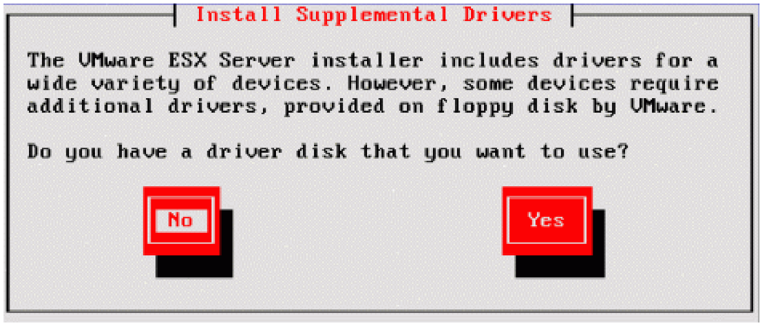
> **Important:** If you are installing on a x440 enter `esx apic` at the `boot:` prompt when it appears.



```
Welcome to VMware ESX Server

Welcome to VMware ESX Server. We need to collect a few
pieces of information about your system before installing
this software. The next few screens will guide you through
the initial installation process.

If you already have ESX 1.0 or 1.1 installed on this system
and would like to upgrade it, please select the 'Upgrade'
option.  Otherwise, select 'Install'.
```

Figure 11-2 Begin installation

The next dialog shown on Figure 11-3 asks for a driver disk. Actually, no additional driver is required: Answer **No**.

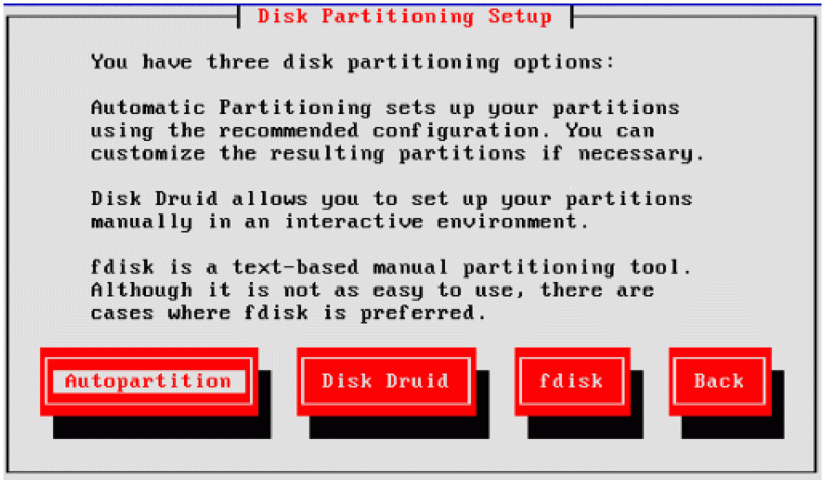


```
Install Supplemental Drivers

The VMware ESX Server installer includes drivers for a
wide variety of devices. However, some devices require
additional drivers, provided on floppy disk by VMware.

Do you have a driver disk that you want to use?
```

Figure 11-3 Driver disk

The system now starts probing the hardware, and installing appropriate modules for the detected hardware.

Upon completion of the hardware probing space, the Disk Partitioning Setup screen (shown in figure Figure 11-4) displays.



```
Disk Partitioning Setup

You have three disk partitioning options:

Automatic Partitioning sets up your partitions
using the recommended configuration. You can
customize the resulting partitions if necessary.

Disk Druid allows you to set up your partitions
manually in an interactive environment.

fdisk is a text-based manual partitioning tool.
Although it is not as easy to use, there are
cases where fdisk is preferred.
```

Figure 11-4 Partitioning of console OS

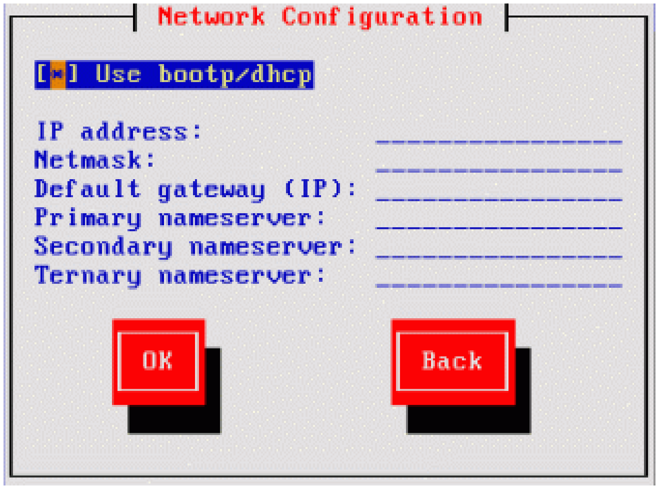
It is essential to keep in mind that the partitioning done at this point pertains to the console OS (storage partitioning for the virtual machines later will be done later). Even though **`Autopartitioning`** is convenient, we suggest that you partition the disk manually. Please keep the following recommendations in mind:

- The root partition should be at least 1.8 GB in size.
- For `/boot` about 25 MB are required.

- The /home directory is used for two purposes: storing the configurations of the virtual machines and storage for suspended virtual machines. This means that you require 10 MB + (size of VM memory) for each virtual machine (if your virtual machine has a memory size larger than 2 GB, the VMs must be attached to VMFS volumes; see Figure 11-27).
  - Apply the following rules for deciding the size of the swap partition:
    - 1-4 VMs -> 128 MB
    - 5-8 VMs -> 192 MB
    - 9-16 VMs -> 272 MB
    - 17-32 VMs -> 384 MB
    - > 32 VMs -> 512 MB

When the partitioning is done, the system prompts you as to where to locate the boot loader (LILO), usually in the Master Boot Record (MBR).

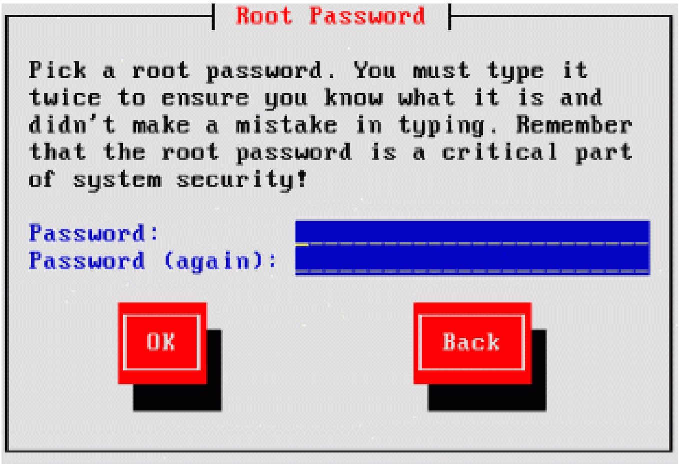
The installation proceeds with the setup of a network interface (Figure 11-5). This network is used by the console OS only. You need a static IP address for this interface because all configuration and setup communications are carried through this device.



```
Network Configuration

[*] Use bootp/dhcp

IP address:              ________________
Netmask:                 ________________
Default gateway (IP):    ________________
Primary nameserver:      ________________
Secondary nameserver:    ________________
Ternary nameserver:      ________________
```

Figure 11-5 Network setup of console OS

Next, you set up the time zone (screen not shown) and then specify a root password (Figure 11-6). This password is required for any fundamental modification to the VMware ESX server.



Root Password

Pick a root password. You must type it twice to ensure you know what it is and didn't make a mistake in typing. Remember that the root password is a critical part of system security!

Password:
Password (again):

Figure 11-6 Define root password

Click **OK**.

The the next dialog allows you to define additional users of the system ().

These additional users can be later defined as administrators for the virtual machines (creation, start, stop, and other maintenance tasks.). We strongly recommend *not* to use the root user ID for the setup and administration of virtual machines.

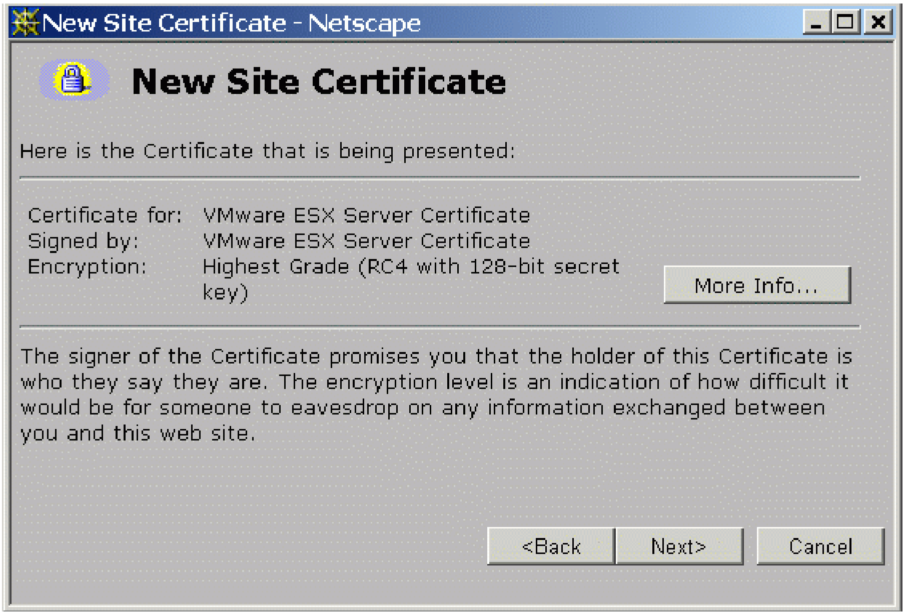
The installation continues until all the console OS and the VMware ESX server components have been copied to the disk. When completed, the server reboots and displays the LILO boot menu. If you do nothing the default setting (VMware kernel) is loaded.

11.3.2 Configuration of the ESX server

Once the VMware ESX server is loaded, you can access the system either locally via a console (press Alt+F2 to enter a console) or remotely. To do a remote access for the first time, start a Web browser and enter the following destination:

```
http://<VMware ESX IP address>
```

The VMware server responds to your request by transmitting a site certificate, as shown in Figure 11-7.



New Site Certificate

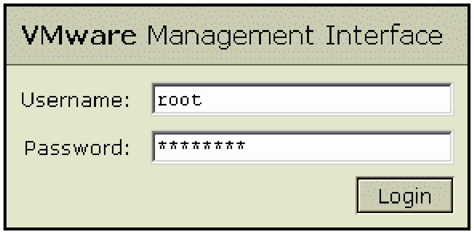
Here is the Certificate that is being presented:

Certificate for: VMware ESX Server Certificate
Signed by: VMware ESX Server Certificate
Encryption: Highest Grade (RC4 with 128-bit secret key)

The signer of the Certificate promises you that the holder of this Certificate is who they say they are. The encryption level is an indication of how difficult it would be for someone to eavesdrop on any information exchanged between you and this web site.

Figure 11-7 Security certificate of VMware ESX server

Accept the certificate, then enter the root user ID and password on the login screen shown in Figure 11-8.

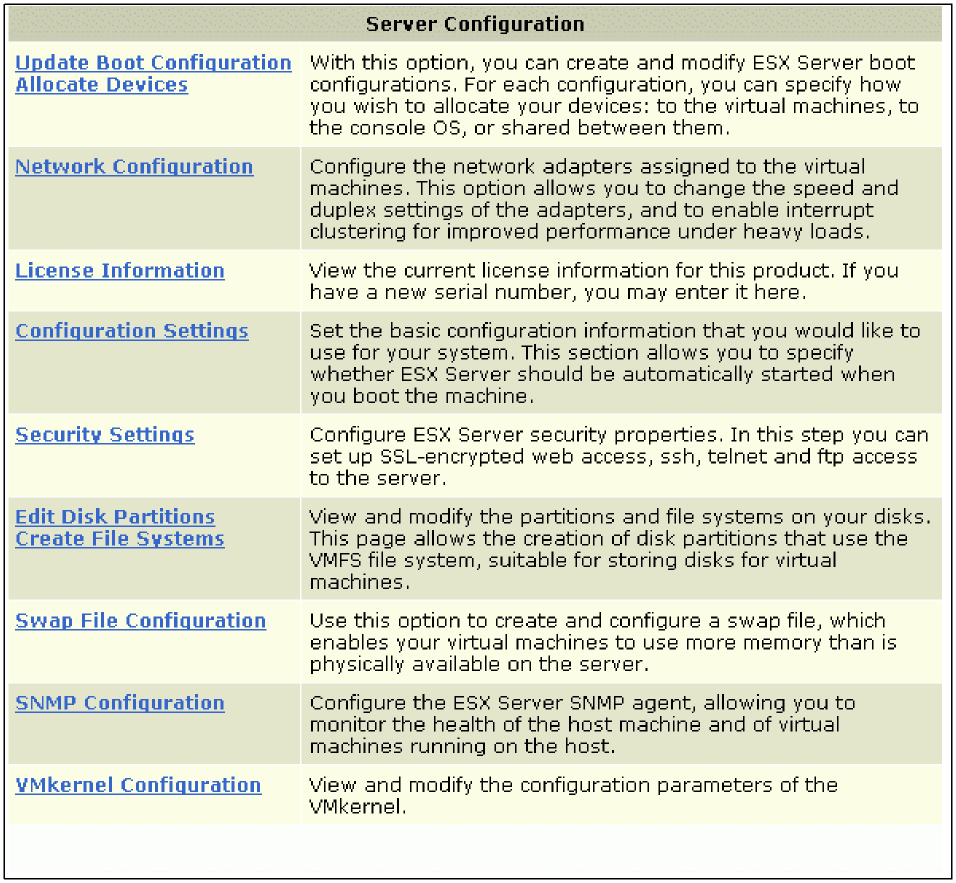

VMware Management Interface

Username: root

Password: *********

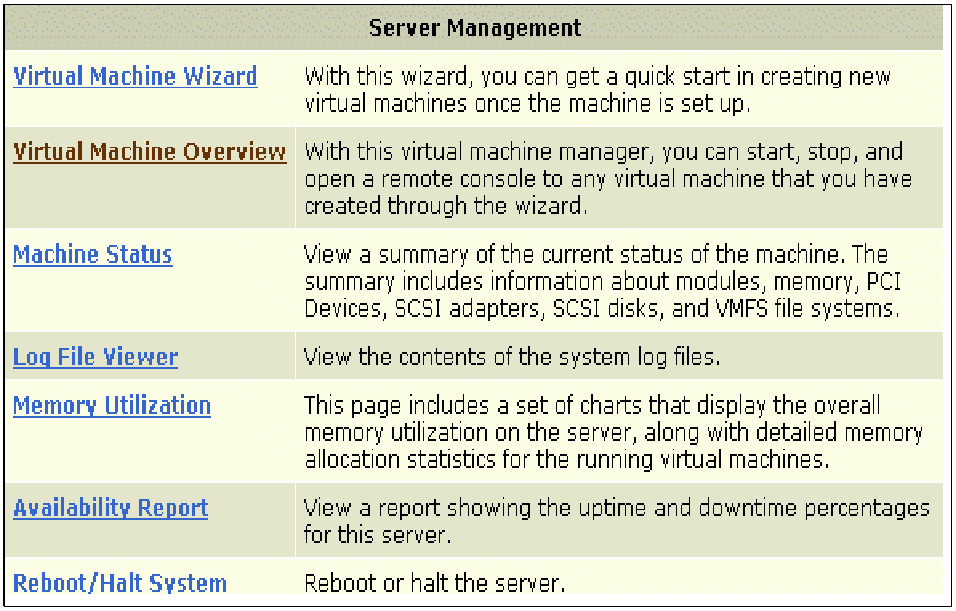
Figure 11-8 Login on VMware management console

If the login is successful, the next page that appears within your browser displays the ESX server configuration menu (Figure 11-10) and ESX server management menu (Figure 11-10).



Server Configuration	
Update Boot Configuration Allocate Devices	With this option, you can create and modify ESX Server boot configurations. For each configuration, you can specify how you wish to allocate your devices: to the virtual machines, to the console OS, or shared between them.
Network Configuration	Configure the network adapters assigned to the virtual machines. This option allows you to change the speed and duplex settings of the adapters, and to enable interrupt clustering for improved performance under heavy loads.
License Information	View the current license information for this product. If you have a new serial number, you may enter it here.
Configuration Settings	Set the basic configuration information that you would like to use for your system. This section allows you to specify whether ESX Server should be automatically started when you boot the machine.
Security Settings	Configure ESX Server security properties. In this step you can set up SSL-encrypted web access, ssh, telnet and ftp access to the server.
Edit Disk Partitions Create File Systems	View and modify the partitions and file systems on your disks. This page allows the creation of disk partitions that use the VMFS file system, suitable for storing disks for virtual machines.
Swap File Configuration	Use this option to create and configure a swap file, which enables your virtual machines to use more memory than is physically available on the server.
SNMP Configuration	Configure the ESX Server SNMP agent, allowing you to monitor the health of the host machine and of virtual machines running on the host.
VMkernel Configuration	View and modify the configuration parameters of the VMkernel.

Figure 11-9 VMware ESX server configuration



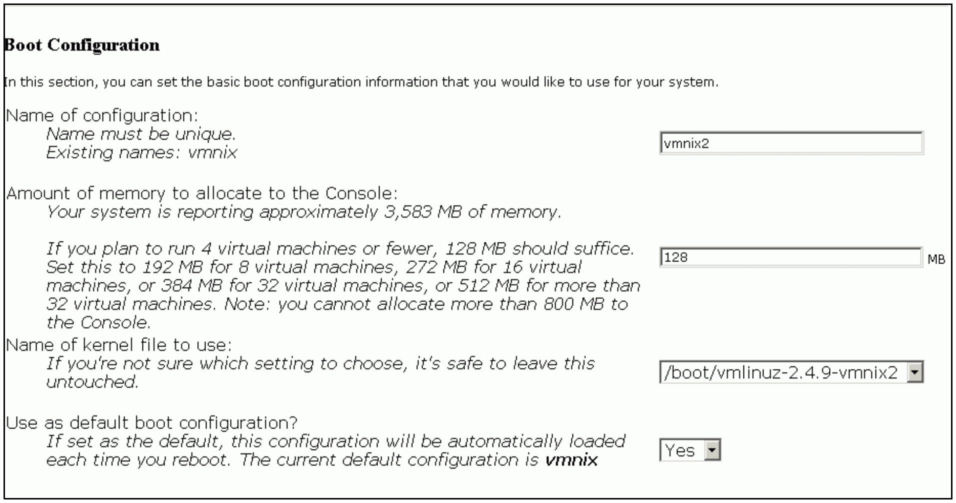
Server Management	
Virtual Machine Wizard	With this wizard, you can get a quick start in creating new virtual machines once the machine is set up.
Virtual Machine Overview	With this virtual machine manager, you can start, stop, and open a remote console to any virtual machine that you have created through the wizard.
Machine Status	View a summary of the current status of the machine. The summary includes information about modules, memory, PCI Devices, SCSI adapters, SCSI disks, and VMFS file systems.
Log File Viewer	View the contents of the system log files.
Memory Utilization	This page includes a set of charts that display the overall memory utilization on the server, along with detailed memory allocation statistics for the running virtual machines.
Availability Report	View a report showing the uptime and downtime percentages for this server.
Reboot/Halt System	Reboot or halt the server.

Figure 11-10 VMware ESX server management

Go to the server configuration menu first to finalize the server setup and be able to access the Fibre Channel storage.

We recommend that you start with license information (here you enter your activation key) and the security settings.

Next, select **Update Boot Configuration/Allocate Devices.** A new page in your browser displays the boot configuration (Figure 11-11) and the device allocation (Figure 11-12) dialogs.



Boot Configuration

In this section, you can set the basic boot configuration information that you would like to use for your system.

Name of configuration:
Name must be unique.
Existing names: vmnix
vmnix2

Amount of memory to allocate to the Console:
Your system is reporting approximately 3,583 MB of memory.

If you plan to run 4 virtual machines or fewer, 128 MB should suffice. Set this to 192 MB for 8 virtual machines, 272 MB for 16 virtual machines, or 384 MB for 32 virtual machines, or 512 MB for more than 32 virtual machines. Note: you cannot allocate more than 800 MB to the Console.
128 MB

Name of kernel file to use:
If you're not sure which setting to choose, it's safe to leave this untouched.
/boot/vmlinuz-2.4.9-vmnix2

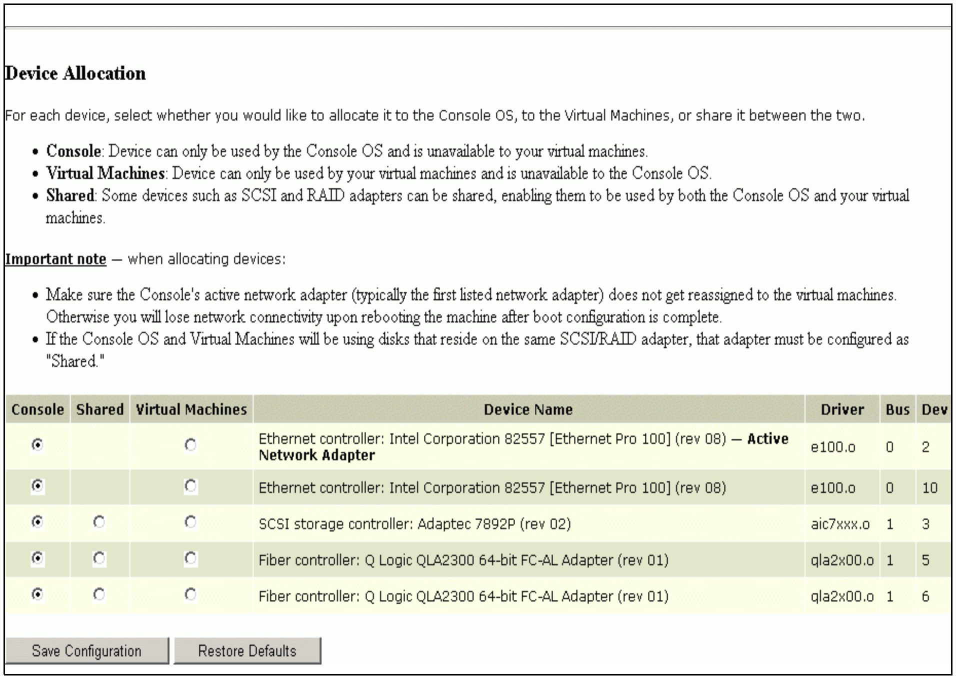
Use as default boot configuration?
If set as the default, this configuration will be automatically loaded each time you reboot. The current default configuration is ***vmnix***
Yes

Figure 11-11 Boot configuration

In the boot configuration you can specify the name to display on the LILO screen, the memory reserved for the console OS, and which VMware kernel file to load at startup. Additionally, you can mark this configuration as default. For the console OS memory, use the following rules:

- 1-4 VMs -> 128 MB
- 5-8 VMs -> 192 MB
- 9-16 VMs -> 272 MB
- 17-32 VMs -> 384 MB
- > 32 VMs -> 512 MB

The device allocation dialog in Figure 11-12 allows you to distribute your devices between console OS and virtual machines. Although SCSI devices only can be shared between VMs and console OS, we recommend that you keep them separate.



Device Allocation

For each device, select whether you would like to allocate it to the Console OS, to the Virtual Machines, or share it between the two.

- **Console**: Device can only be used by the Console OS and is unavailable to your virtual machines.
- **Virtual Machines**: Device can only be used by your virtual machines and is unavailable to the Console OS.
- **Shared**: Some devices such as SCSI and RAID adapters can be shared, enabling them to be used by both the Console OS and your virtual machines.

Important note — when allocating devices:

- Make sure the Console's active network adapter (typically the first listed network adapter) does not get reassigned to the virtual machines. Otherwise you will lose network connectivity upon rebooting the machine after boot configuration is complete.
- If the Console OS and Virtual Machines will be using disks that reside on the same SCSI/RAID adapter, that adapter must be configured as "Shared."

Console	Shared	Virtual Machines	Device Name	Driver	Bus	Dev
●		○	Ethernet controller: Intel Corporation 82557 [Ethernet Pro 100] (rev 08) — **Active Network Adapter**	e100.o	0	2
●		○	Ethernet controller: Intel Corporation 82557 [Ethernet Pro 100] (rev 08)	e100.o	0	10
●	○	○	SCSI storage controller: Adaptec 7892P (rev 02)	aic7xxx.o	1	3
●	○	○	Fiber controller: Q Logic QLA2300 64-bit FC-AL Adapter (rev 01)	qla2x00.o	1	5
●	○	○	Fiber controller: Q Logic QLA2300 64-bit FC-AL Adapter (rev 01)	qla2x00.o	1	6

Figure 11-12 Initial device allocation

If you are planning to install two Fibre Channel HBAs for redundancy, you have to take some details into account:

- By default the qla2x00 module is installed; this module does not support failover.
- To support failover, both HBAs must be identical.
- Only Qla2200 and Qla2300 adapters are supported for failover.
- Both cards must see exactly the same targets.
- There is no load balancing.
- Failover is provided for both ESS and FAStT.

To replace the driver modules, you have to either connect to the server via *ssh* or work directly from the console (assuming you did not change the security settings yet). Once you have access to the console OS shell, you need to edit or add (in case it does not exist yet) the file `vmware-devices.map.local`. Figure 11-13 shows the changes required for VMware to recognize the new settings. The file `vmware-devices.map` contains the hardware definition for the VMware kernel.

```
[root@helium root]# cd /etc/vmware
[root@helium vmware]# cat vmware-devices.map.local
device,0x1077,0x2200,vmhba,QLA2200,qla2200.o
device,0x1077,0x2300,vmhba,QLA2300,qla2300.o
device,0x1077,0x2312,vmhba,QLA2300,qla2300.o
[root@helium vmware]#
```

Figure 11-13 Extension required to use fail-over driver

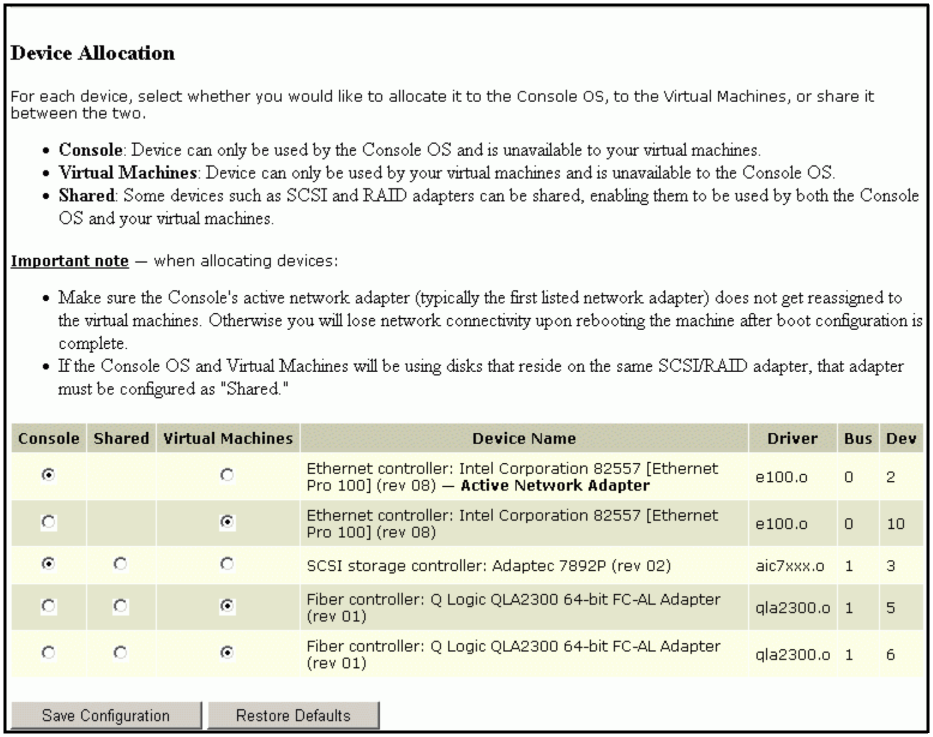
The file `vmware-devices.map.local` is used to select different device drivers. At boot time the file is evaluated. All entries override settings made in *`vmware-devices.map`*.

Once the changes are made, the device allocation console displays the new driver (Figure 11-15).

Now you can select the device with the correct device drivers for the virtual machines.

> **Tip:** To load the VMkernel driver manually, you have to enter at a shell `'/usr/sbin/vmkload_mod' /usr/lib/vmware/vmkmod/qla2xxx/qla2300.o vmhba`

Before you can reboot to commit the changes, you must set up your external storage. Refer to Chapter 9., “FAStT Storage Manager” on page 187 if you are using FAStT, or Chapter 7., “Configuring ESS for Linux” on page 155 if you are using ESS.



Device Allocation

For each device, select whether you would like to allocate it to the Console OS, to the Virtual Machines, or share it between the two.

- **Console**: Device can only be used by the Console OS and is unavailable to your virtual machines.
- **Virtual Machines**: Device can only be used by your virtual machines and is unavailable to the Console OS.
- **Shared**: Some devices such as SCSI and RAID adapters can be shared, enabling them to be used by both the Console OS and your virtual machines.

Important note — when allocating devices:

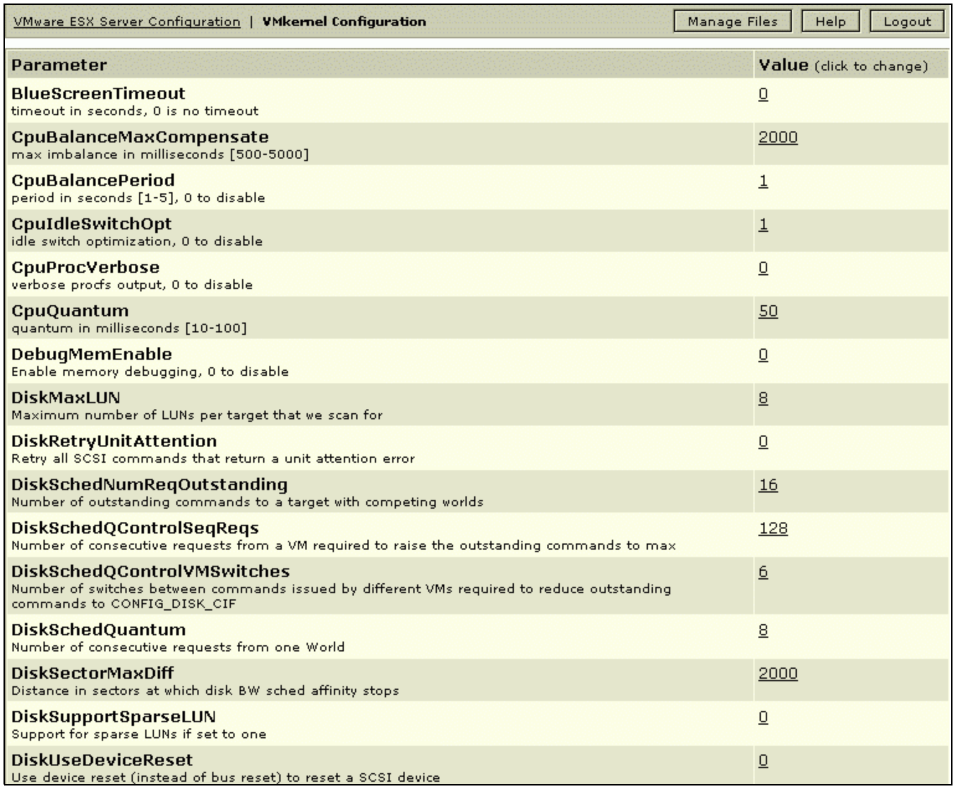
- Make sure the Console's active network adapter (typically the first listed network adapter) does not get reassigned to the virtual machines. Otherwise you will lose network connectivity upon rebooting the machine after boot configuration is complete.
- If the Console OS and Virtual Machines will be using disks that reside on the same SCSI/RAID adapter, that adapter must be configured as "Shared."

Console	Shared	Virtual Machines	Device Name	Driver	Bus	Dev
●		○	Ethernet controller: Intel Corporation 82557 [Ethernet Pro 100] (rev 08) — **Active Network Adapter**	e100.o	0	2
○		●	Ethernet controller: Intel Corporation 82557 [Ethernet Pro 100] (rev 08)	e100.o	0	10
●	○	○	SCSI storage controller: Adaptec 7892P (rev 02)	aic7xxx.o	1	3
○	○	●	Fiber controller: Q Logic QLA2300 64-bit FC-AL Adapter (rev 01)	qla2300.o	1	5
○	○	●	Fiber controller: Q Logic QLA2300 64-bit FC-AL Adapter (rev 01)	qla2300.o	1	6

Figure 11-14 Final assignment of devices

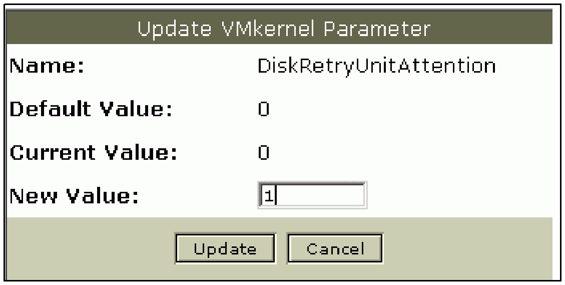
Now, attach the HBAs to your configured storage.

If you are using a FAStT you have to set one flag in the VMware kernel (Figure 11-15 on page 241, and Figure 11-16 on page 241). This is necessary to prevent VMware from misinterpreting status codes from FAStT as error codes (LUN exists, but is not accessible).



Parameter	Value (click to change)
BlueScreenTimeout timeout in seconds, 0 is no timeout	0
CpuBalanceMaxCompensate max imbalance in milliseconds [500-5000]	2000
CpuBalancePeriod period in seconds [1-5], 0 to disable	1
CpuIdleSwitchOpt idle switch optimization, 0 to disable	1
CpuProcVerbose verbose procfs output, 0 to disable	0
CpuQuantum quantum in milliseconds [10-100]	50
DebugMemEnable Enable memory debugging, 0 to disable	0
DiskMaxLUN Maximum number of LUNs per target that we scan for	8
DiskRetryUnitAttention Retry all SCSI commands that return a unit attention error	0
DiskSchedNumReqOutstanding Number of outstanding commands to a target with competing worlds	16
DiskSchedQControlSeqReqs Number of consecutive requests from a VM required to raise the outstanding commands to max	128
DiskSchedQControlVMSwitches Number of switches between commands issued by different VMs required to reduce outstanding commands to CONFIG_DISK_CIF	6
DiskSchedQuantum Number of consecutive requests from one World	8
DiskSectorMaxDiff Distance in sectors at which disk BW sched affinity stops	2000
DiskSupportSparseLUN Support for sparse LUNs if set to one	0
DiskUseDeviceReset Use device reset (instead of bus reset) to reset a SCSI device	0

Figure 11-15 Settings for FCal storage



Name:	DiskRetryUnitAttention
Default Value:	0
Current Value:	0
New Value:	1

Figure 11-16 Update VMkernel flag

Once you have saved the configuration, reboot the system to activate your settings. To reboot, you must select **Reboot/Halt System** from the Server Management dialog shown in Figure 11-10 on page 236.

11.3.3 Setup disk space for VMs

After reboot, login to the management console again to start setting up disk space for the virtual machines.

Important: This is a brief and summarized description. For details please refer to the *VMware ESX 1.52* server manual.

For our testing environment, we define three LUNs on the FAStT 700. Next, following instructions given in the previous section, we install and load an appropriate HBA driver and connect the server to the SAN. Figure 11-17 shows, for our installation, that the three LUNs are correctly detected as physical drives.

To partition these drives you have two options:

Create New Partition Partitions the disk automatically; additionally a partition for a VMware dump (crash dump space for VMware) is created (Figure 11-18 on page 243).

Expert Mode Fdisk Allows you to specify exactly, and based on your own input, how to partition the volumes (Figure 11-19 on page 244)

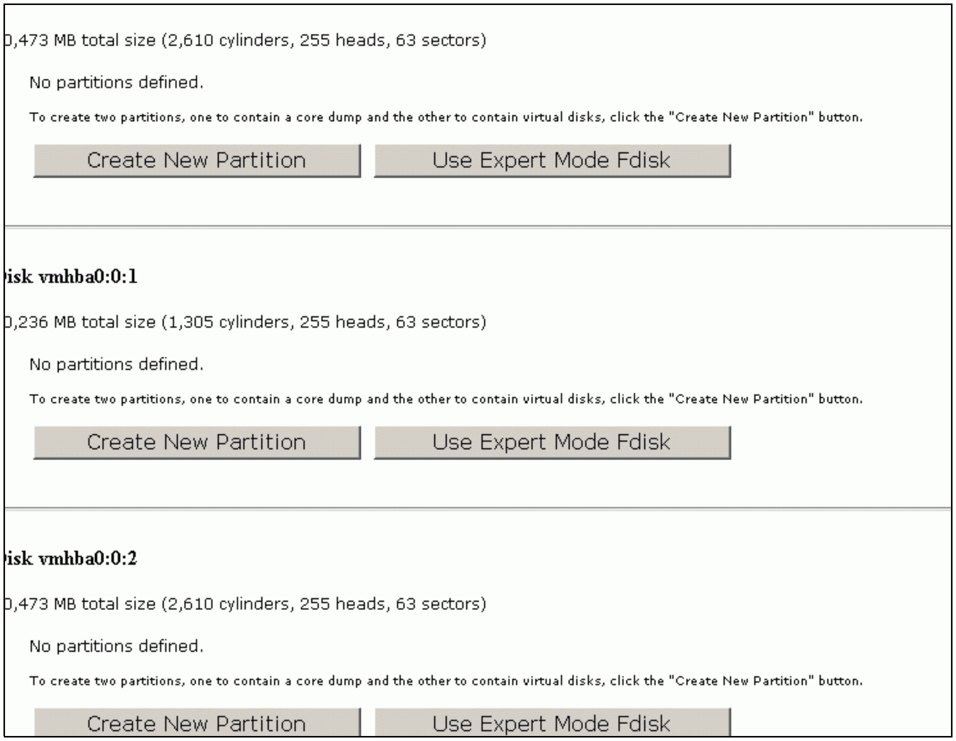
In addition to a label, you can specify the VMFS accessibility by three levels:

Private Only the specified virtual machines can access the partition.

Public The partition is available to multiple physical servers and their virtual machine, but only to one server at a single time.

Shared Shared access to the partition is possible; this setting is required for shared disk solutions like fail-over clustering between virtual machines.

The menu in Figure 11-9 on page 235 shows that you can create swap space for the virtual machines (Figure 11-20 on page 244). This feature allows you to assign more memory to your virtual machines than is physically available on the server. Please keep in mind that intensive swapping negatively impacts the system performance.



0,473 MB total size (2,610 cylinders, 255 heads, 63 sectors)

No partitions defined.

To create two partitions, one to contain a core dump and the other to contain virtual disks, click the "Create New Partition" button.

isk vmhba0:0:1

0,236 MB total size (1,305 cylinders, 255 heads, 63 sectors)

No partitions defined.

To create two partitions, one to contain a core dump and the other to contain virtual disks, click the "Create New Partition" button.

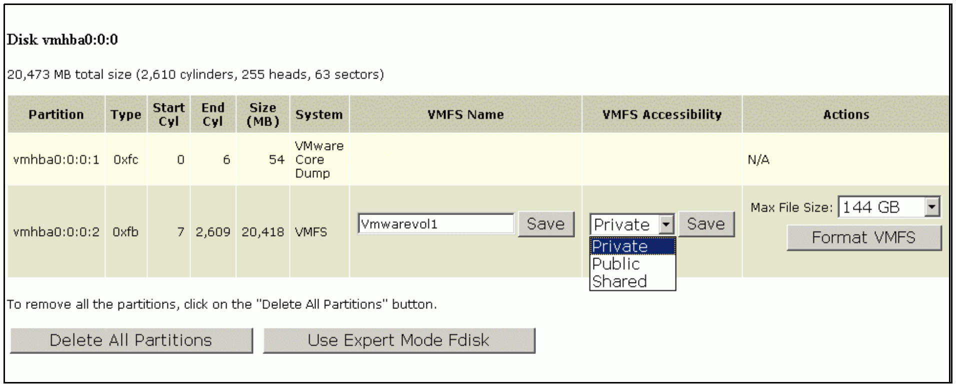
isk vmhba0:0:2

0,473 MB total size (2,610 cylinders, 255 heads, 63 sectors)

No partitions defined.

To create two partitions, one to contain a core dump and the other to contain virtual disks, click the "Create New Partition" button.

Figure 11-17 Volumes are visible



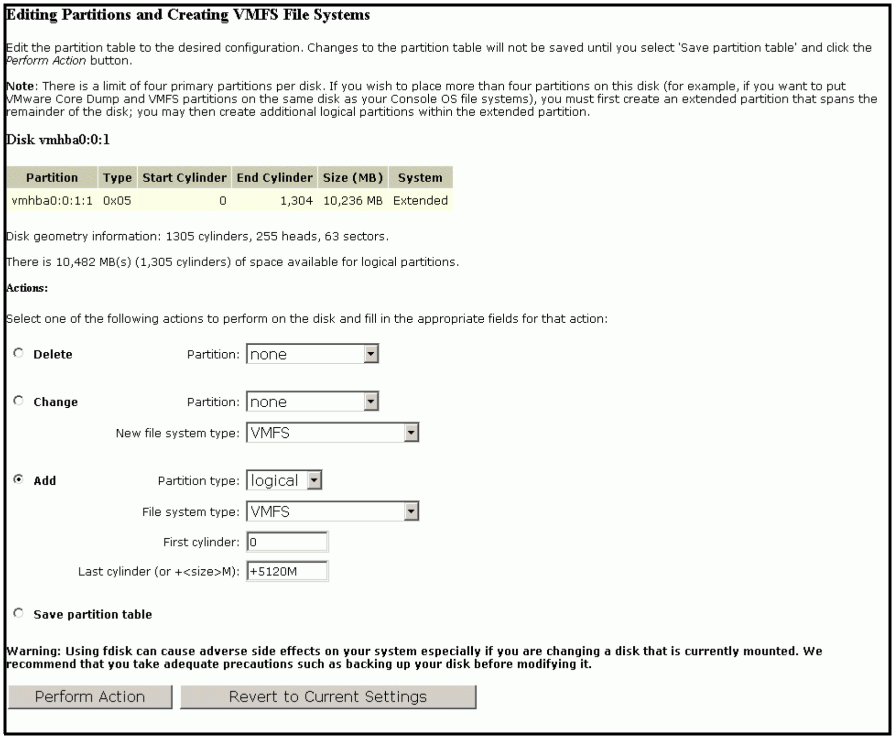
Disk vmhba0:0:0

20,473 MB total size (2,610 cylinders, 255 heads, 63 sectors)

Partition	Type	Start Cyl	End Cyl	Size (MB)	System	VMFS Name	VMFS Accessibility	Actions
vmhba0:0:0:1	0xfc	0	6	54	VMware Core Dump			N/A
vmhba0:0:0:2	0xfb	7	2,609	20,418	VMFS	Vmwarevol1 Save	Private Save Private Public Shared	Max File Size: 144 GB Format VMFS

To remove all the partitions, click on the "Delete All Partitions" button.

Figure 11-18 Create new partition



Editing Partitions and Creating VMFS File Systems

Edit the partition table to the desired configuration. Changes to the partition table will not be saved until you select 'Save partition table' and click the *Perform Action* button.

Note: There is a limit of four primary partitions per disk. If you wish to place more than four partitions on this disk (for example, if you want to put VMware Core Dump and VMFS partitions on the same disk as your Console OS file systems), you must first create an extended partition that spans the remainder of the disk; you may then create additional logical partitions within the extended partition.

Disk vmhba0:0:1

Partition	Type	Start Cylinder	End Cylinder	Size (MB)	System
vmhba0:0:1:1	0x05	0	1,304	10,236 MB	Extended

Disk geometry information: 1305 cylinders, 255 heads, 63 sectors.

There is 10,482 MB(s) (1,305 cylinders) of space available for logical partitions.

Actions:

Select one of the following actions to perform on the disk and fill in the appropriate fields for that action:

○ **Delete** Partition: none

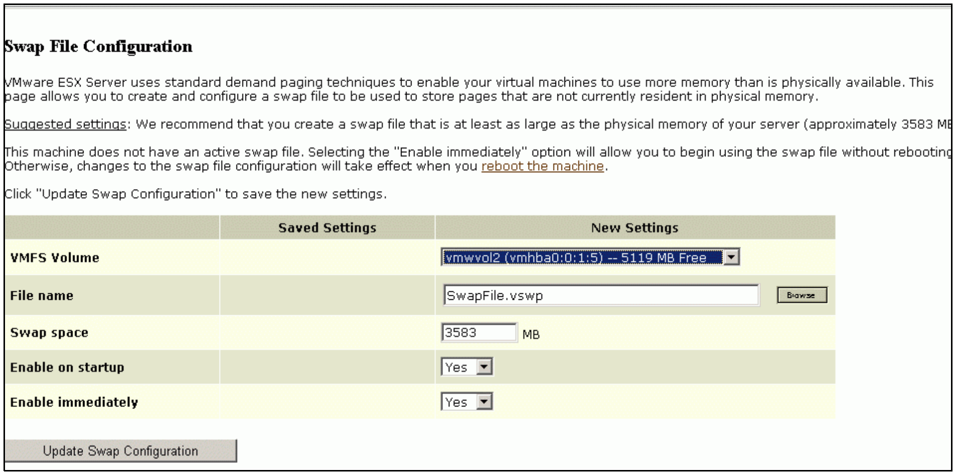
○ **Change** Partition: none
New file system type: VMFS

● **Add** Partition type: logical
File system type: VMFS
First cylinder: 0
Last cylinder (or +<size>M): +5120M

○ **Save partition table**

Warning: Using fdisk can cause adverse side effects on your system especially if you are changing a disk that is currently mounted. We recommend that you take adequate precautions such as backing up your disk before modifying it.

Figure 11-19 Expert mode



Swap File Configuration

VMware ESX Server uses standard demand paging techniques to enable your virtual machines to use more memory than is physically available. This page allows you to create and configure a swap file to be used to store pages that are not currently resident in physical memory.

Suggested settings: We recommend that you create a swap file that is at least as large as the physical memory of your server (approximately 3583 M

This machine does not have an active swap file. Selecting the "Enable immediately" option will allow you to begin using the swap file without rebootin
Otherwise, changes to the swap file configuration will take effect when you reboot the machine.

Click "Update Swap Configuration" to save the new settings.

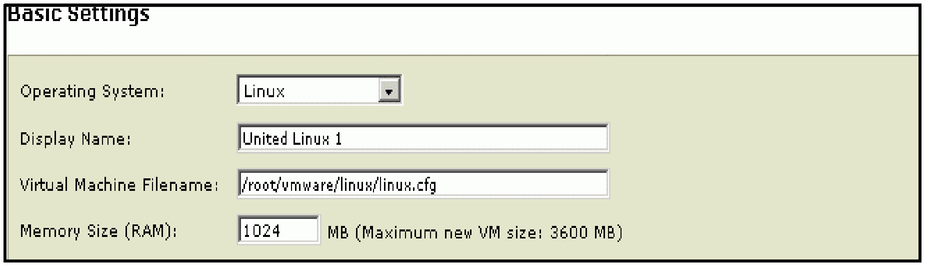
	Saved Settings	New Settings
VMFS Volume		vmwvol2 (vmhba0:0:1:5) -- 5119 MB Free
File name		SwapFile.vswp Browse
Swap space		3583 MB
Enable on startup		Yes
Enable immediately		Yes

Figure 11-20 Swap for VMware (more logical than physical memory)

11.3.4 Define a VM using the wizard

To create a virtual machine, click **Virtual Machine Wizard** in Figure 11-10 on page 236.The page that displays in your browser let you specify the major characteristics of a VM. The page contains all the elements shown in Figure 11-21 on page 245 through Figure 11-26 on page 247.

- In the Basic Settings ()section, you define the guest OS to install, select a **display name** for this VM, specify the path for the configuration file, and the memory size available to the VM.



Basic Settings

Operating System: Linux

Display Name: United Linux 1

Virtual Machine Filename: /root/vmware/linux/linux.cfg

Memory Size (RAM): 1024 MB (Maximum new VM size: 3600 MB)

Figure 11-21 Define VM basic settings

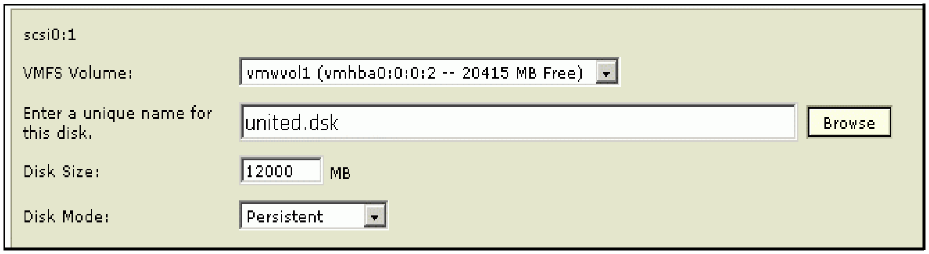
- In the SCSI Disk section, select one of the partitions you had set up before. There are four disk modes available:

Persistent All write accesses by the guest are written permanently to the disk immediately.

Nonpersistent All changes are discarded when a virtual machine is shut down and powered off.

Undoable All changes are written to a log file and can be discarded at anytime.

Append All changes are stored in a log file; the changes can be discarded or made permanent.



scsi0:1

VMFS Volume: vmwvol1 (vmhba0:0:0:2 -- 20415 MB Free)

Enter a unique name for this disk. united.dsk

Disk Size: 12000 MB

Disk Mode: Persistent

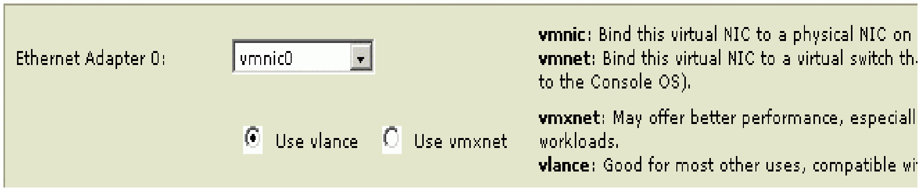
Figure 11-22 Define VM SCSI disks

- In the Networking section you can have a choice between two different types of NIC:

  vmnic Binds the virtual NIC to a physical NIC; each physical NIC has its own VMnic device number

  vmnet Binds the virtual NIC to a NIC of the virtual network provided by the VMware

  Additionally, you can select the type of the virtual NIC (vmlance or vmxnet).



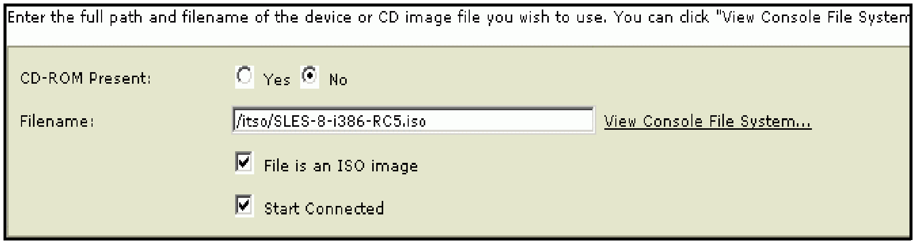
Ethernet Adapter 0: vmnic0

● Use vlance ○ Use vmxnet

vmnic: Bind this virtual NIC to a physical NIC on
vmnet: Bind this virtual NIC to a virtual switch th
to the Console OS).
vmxnet: May offer better performance, especiall
workloads.
vlance: Good for most other uses, compatible wi

Figure 11-23 Define VM networking

- The CD-ROM section allows you to manage the access of a virtual machine to the CD drive. Alternatively, you can specify an ISO file instead of a physical CD-ROM.



Enter the full path and filename of the device or CD image file you wish to use. You can click "View Console File System

CD-ROM Present: ○ Yes ● No

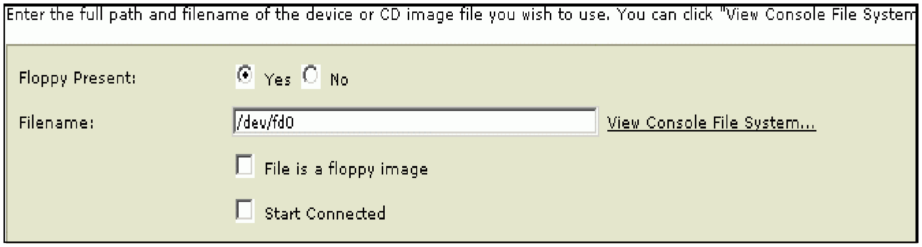
Filename: /itso/SLES-8-i386-RC5.iso View Console File System...

☑ File is an ISO image

☑ Start Connected

Figure 11-24 Define VM CD-ROM

- The floppy drive section is similar to the CD-ROM section.



Enter the full path and filename of the device or CD image file you wish to use. You can click "View Console File System

Floppy Present: ● Yes ○ No

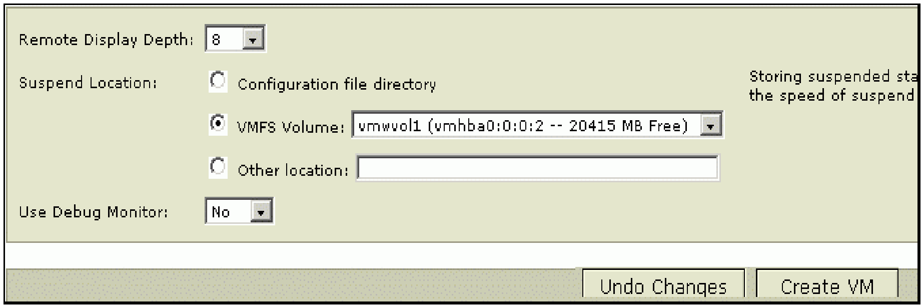
Filename: /dev/fd0 View Console File System...

□ File is a floppy image

□ Start Connected

Figure 11-25 Define VM floppy drive

- In the Misc section, you can specify the suspend location (necessary, if your virtual machines uses more than 2 GB of memory).



Remote Display Depth: 8

Suspend Location: ○ Configuration file directory

● VMFS Volume: vmwvol1 (vmhba0:0:0:2 -- 20415 MB Free)

○ Other location:

Use Debug Monitor: No

Storing suspended sta
the speed of suspend

Figure 11-26 Define VM miscellaneous options and create VM

When you are satisfied with the settings, click **Create VM**.

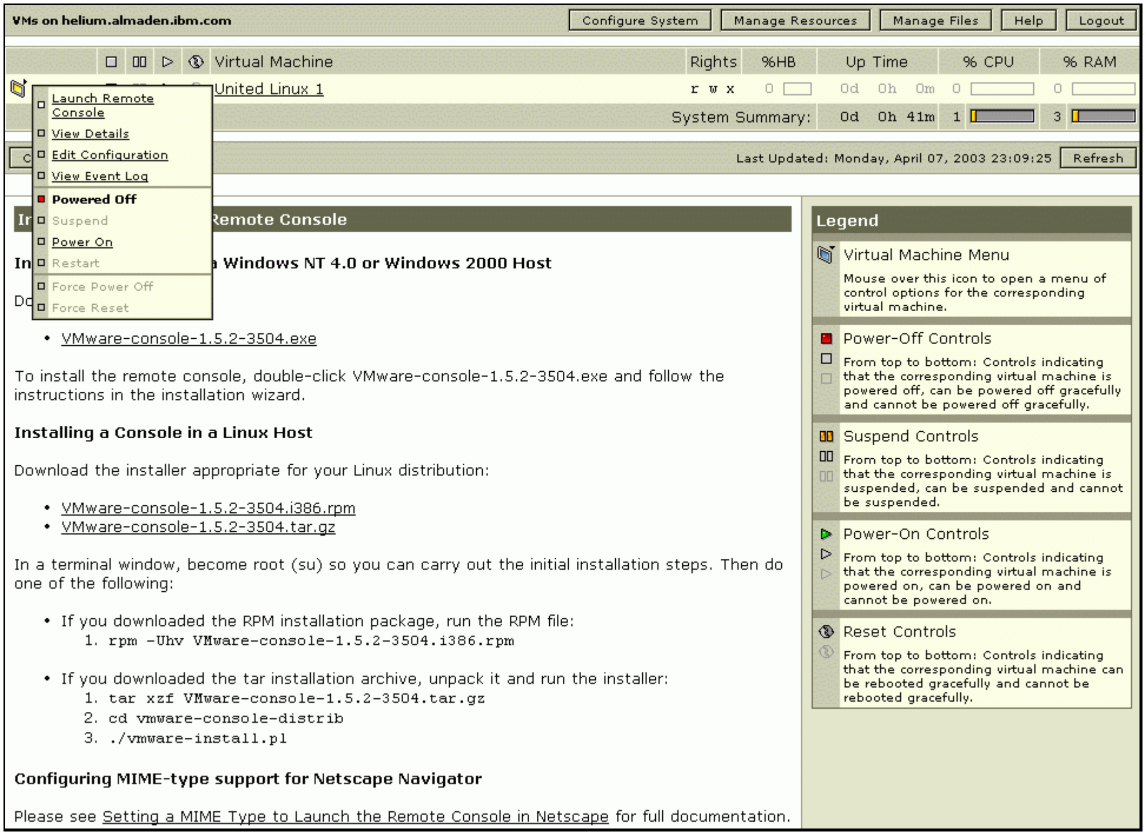
The page in your browser now displays a summary screen for the VM just created. Selecting **Return to Overview** on this page, takes you to the dialog shown in Figure 11-27 on page 248.

11.3.5 Install VMware remote console

Before you launch and monitor your virtual machine you have to install the VMware remote console (see middle left in Figure 11-27 on page 248) on your management workstation.

You can choose between a Windows or a Linux console. Please refer to the dialog for installation details.

If you install the remote console on Windows, you can launch it by right-clicking on the folder symbol located in the upper left corner of the screen (see Figure 11-27 on page 248) and then select **Launch Remote Console** from the pull-down menu.



Virtual Machine	Rights	%HB	Up Time	% CPU	% RAM
United Linux 1	r w x	0	0d 0h 0m	0	0
System Summary:			0d 0h 41m	1	3

Ir… Remote Console

In… a Windows NT 4.0 or Windows 2000 Host

Do…

- VMware-console-1.5.2-3504.exe

To install the remote console, double-click VMware-console-1.5.2-3504.exe and follow the instructions in the installation wizard.

Installing a Console in a Linux Host

Download the installer appropriate for your Linux distribution:

- VMware-console-1.5.2-3504.i386.rpm
- VMware-console-1.5.2-3504.tar.gz

In a terminal window, become root (su) so you can carry out the initial installation steps. Then do one of the following:

- If you downloaded the RPM installation package, run the RPM file:

```
1. rpm -Uhv VMware-console-1.5.2-3504.i386.rpm
```

- If you downloaded the tar installation archive, unpack it and run the installer:

```
1. tar xzf VMware-console-1.5.2-3504.tar.gz
2. cd vmware-console-distrib
3. ./vmware-install.pl
```

Configuring MIME-type support for Netscape Navigator

Please see Setting a MIME Type to Launch the Remote Console in Netscape for full documentation.

Legend

Virtual Machine Menu
Mouse over this icon to open a menu of control options for the corresponding virtual machine.

Power-Off Controls
From top to bottom: Controls indicating that the corresponding virtual machine is powered off, can be powered off gracefully and cannot be powered off gracefully.

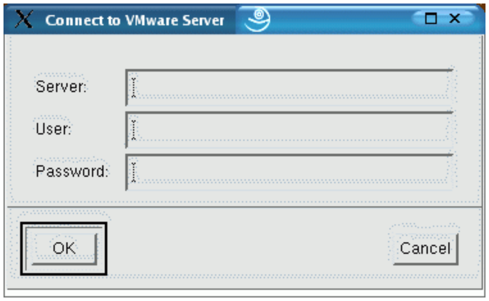
Suspend Controls
From top to bottom: Controls indicating that the corresponding virtual machine is suspended, can be suspended and cannot be suspended.

Power-On Controls
From top to bottom: Controls indicating that the corresponding virtual machine is powered on, can be powered on and cannot be powered on.

Reset Controls
From top to bottom: Controls indicating that the corresponding virtual machine can be rebooted gracefully and cannot be rebooted gracefully.

Figure 11-27 Installation of remote console and launch dialog

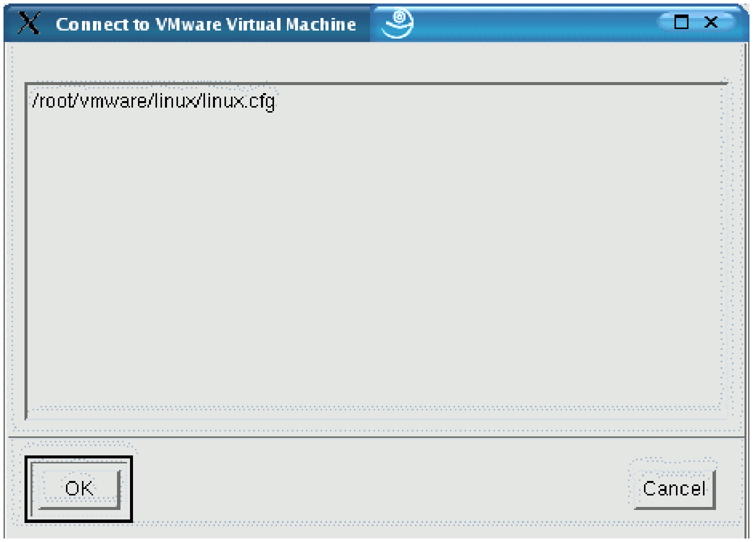
On Linux workstations you have entered the command **vmware-console** The remote console is launched and a dialog asks you for IP address and login information (user and password) to access the ESX server (Figure 11-28).



Server:
User:
Password:

Figure 11-28 Linux Remote Console: connecting to VMware server

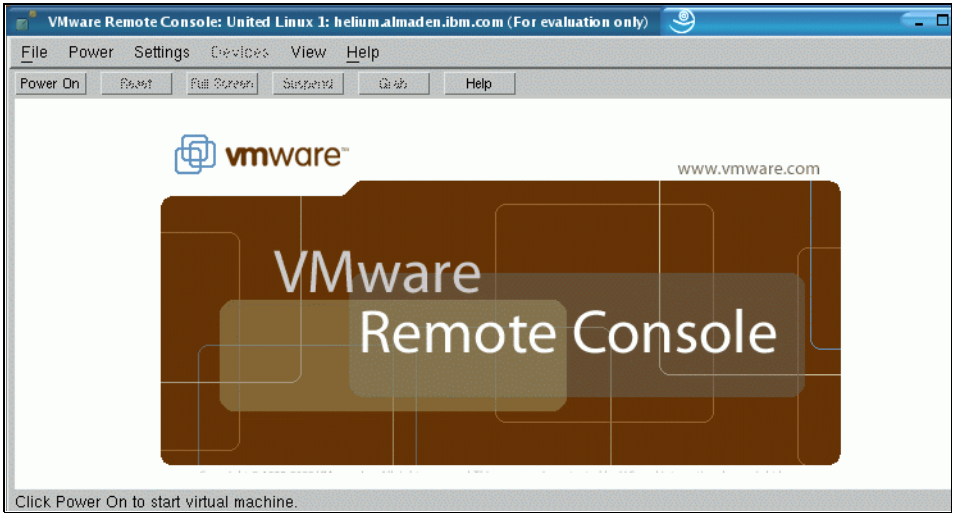
Clicking **OK** connects to the specified server. The next window shows you the list of available VM configurations (Figure 11-29). There is only one defined in our illustration.



/root/vmware/linux/linux.cfg

Figure 11-29 Linux Remote Console: select configuration

After selecting a configuration the console displays as shown in Figure 11-30. To start the virtual machine, click **Power On**.

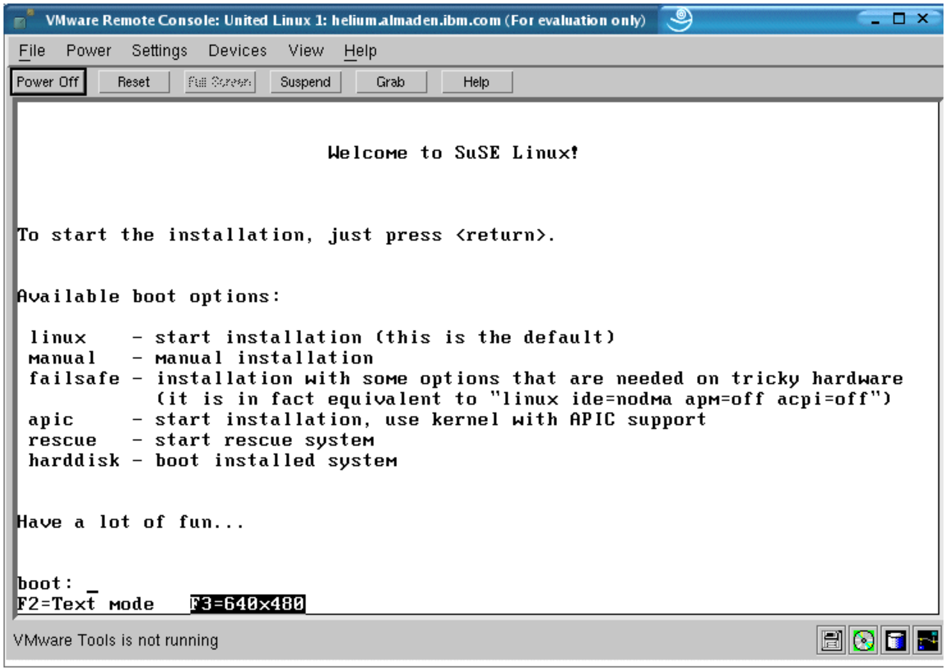


Click Power On to start virtual machine.

Figure 11-30 Linux Remote Console

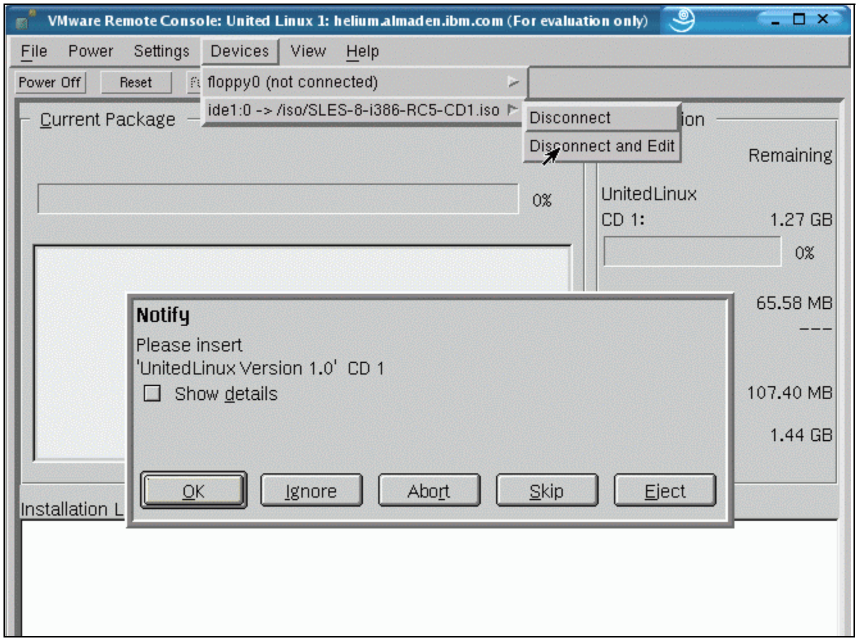
The virtual machine boots like a regular system. The pictures in Figure 11-31, Figure 11-32 on page 251, and Figure 11-33 on page 251) illustrate what appears within the remote console window when you boot a VM defined as a Linux system that is about to be installed.

In this scenario, we had previously copied the ISO-images of the Linux installation CDs to our ESX server and connected the (virtual) CD drive of our Linux VM to an ISO-image (see Figure 11-24 on page 246). This means that whenever the guest OS demands a different CD (Figure 11-31) you have to disconnect the existing ISO file (Figure 11-32 on page 251), define a new one, and connect it to the virtual machine (Figure 11-33 on page 251).


VMware Remote Console: United Linux 1: helium.almaden.ibm.com (For evaluation only)
File
Power
Settings
Devices
View
Help
Power Off
Reset
Full Screen
Suspend
Grab
Help
VMware Tools is not running

```
                        Welcome to SuSE Linux!


To start the installation, just press <return>.


Available boot options:

 linux    - start installation (this is the default)
 manual   - manual installation
 failsafe - installation with some options that are needed on tricky hardware
            (it is in fact equivalent to "linux ide=nodma apm=off acpi=off")
 apic     - start installation, use kernel with APIC support
 rescue   - start rescue system
 harddisk - boot installed system


Have a lot of fun...


boot: _
F2=Text mode    F3=640x480
```

Figure 11-31   Linux Remote Console: Start guest OS installation


VMware Remote Console: United Linux 1: helium.almaden.ibm.com (For evaluation only)
File Power Settings Devices View Help
Power Off Reset
floppy0 (not connected)
ide1:0 -> /iso/SLES-8-i386-RC5-CD1.iso
Disconnect
Disconnect and Edit
Current Package
0%
ion
Remaining
UnitedLinux
CD 1: 1.27 GB
0%
65.58 MB

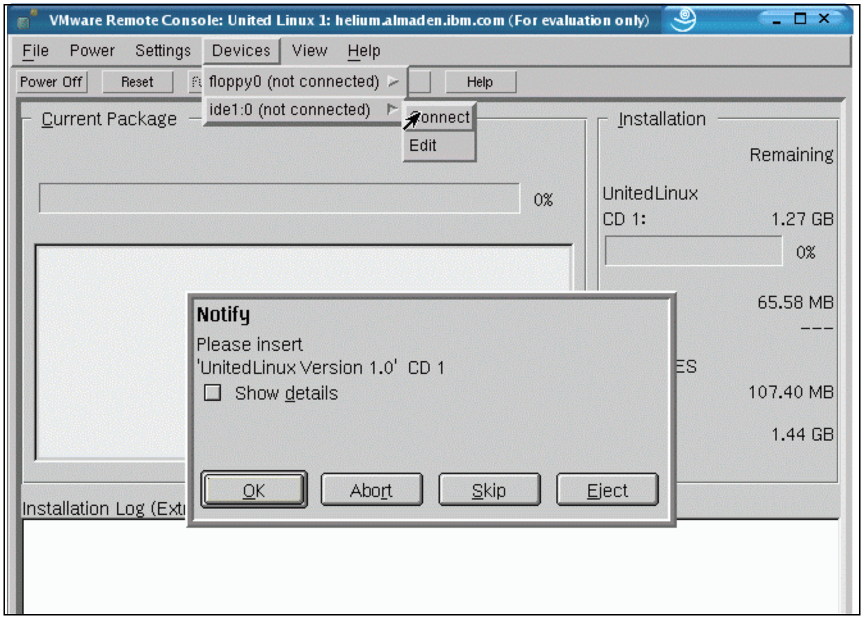
107.40 MB
1.44 GB
Installation L
OK Ignore Abort Skip Eject

Notify

Please insert
'UnitedLinux Version 1.0' CD 1

□ Show details

Figure 11-32 Connect to different ISO-file part 1



Notify

Please insert
'UnitedLinux Version 1.0' CD 1

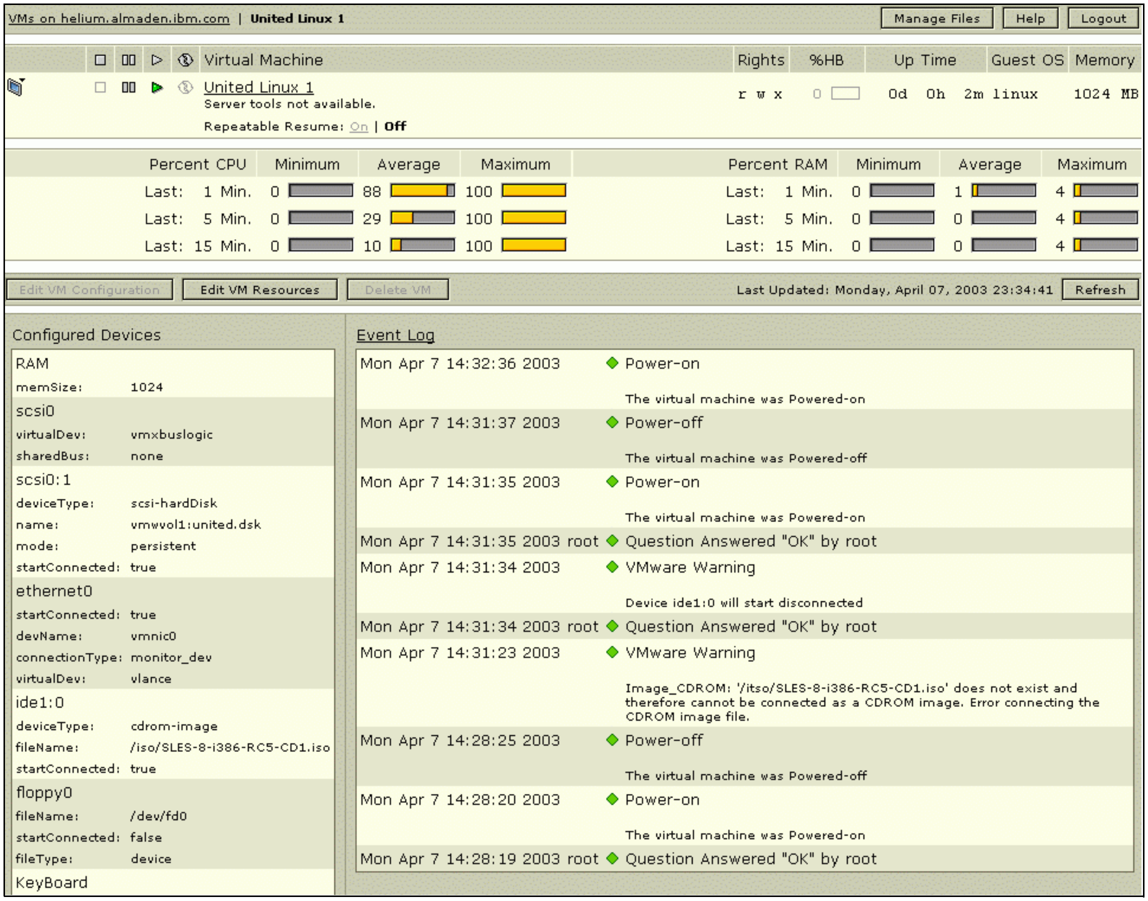
□ Show details

Figure 11-33 Connect to different ISO-file part 2

11.3.6 Monitoring a VM

Once you set up virtual machines and start using them (i.e. in a production environment) VMware allows you to monitor the virtual machines. Figure 11-34 below shows the VM Overview. For tuning purposes you can modify some parameters of the virtual machine by clicking on **Edit VM Resources**.

Please refer to the manual for details on optimizing your resources.



Virtual Machine	Rights	%HB	Up Time	Guest OS	Memory
United Linux 1 Server tools not available. Repeatable Resume: On | Off	r w x	0	0d 0h 2m	linux	1024 MB

Percent CPU	Minimum	Average	Maximum	Percent RAM	Minimum	Average	Maximum
Last: 1 Min.	0	88	100	Last: 1 Min.	0	1	4
Last: 5 Min.	0	29	100	Last: 5 Min.	0	0	4
Last: 15 Min.	0	10	100	Last: 15 Min.	0	0	4

Last Updated: Monday, April 07, 2003 23:34:41

Configured Devices

RAM	
memSize:	1024
scsi0	
virtualDev:	vmxbuslogic
sharedBus:	none
scsi0:1	
deviceType:	scsi-hardDisk
name:	vmwvol1:united.dsk
mode:	persistent
startConnected:	true
ethernet0	
startConnected:	true
devName:	vmnic0
connectionType:	monitor_dev
virtualDev:	vlance
ide1:0	
deviceType:	cdrom-image
fileName:	/iso/SLES-8-i386-RC5-CD1.iso
startConnected:	true
floppy0	
fileName:	/dev/fd0
startConnected:	false
fileType:	device
KeyBoard	

Event Log

```
Mon Apr 7 14:32:36 2003       Power-on
                              The virtual machine was Powered-on
Mon Apr 7 14:31:37 2003       Power-off
                              The virtual machine was Powered-off
Mon Apr 7 14:31:35 2003       Power-on
                              The virtual machine was Powered-on
Mon Apr 7 14:31:35 2003 root  Question Answered "OK" by root
Mon Apr 7 14:31:34 2003       VMware Warning
                              Device ide1:0 will start disconnected
Mon Apr 7 14:31:34 2003 root  Question Answered "OK" by root
Mon Apr 7 14:31:23 2003       VMware Warning
                              Image_CDROM: '/itso/SLES-8-i386-RC5-CD1.iso' does not exist and
                              therefore cannot be connected as a CDROM image. Error connecting the
                              CDROM image file.
Mon Apr 7 14:28:25 2003       Power-off
                              The virtual machine was Powered-off
Mon Apr 7 14:28:20 2003       Power-on
                              The virtual machine was Powered-on
Mon Apr 7 14:28:19 2003 root  Question Answered "OK" by root
```

Figure 11-34 Active VM

A

Storage from the OS view

In this appendix we discuss the concept of storage from the perspective of the operating system, and how this storage is presented to an application. Our intent is to introduce to the person familiar with Windows some of the differences evident in Linux. We will also briefly survey the file systems available with Linux. We will show how to use a feature of Linux called the /proc file system to view attached storage, and briefly mention the file system table /etc/fstab.

Disk storage in the Windows environment

As is often the case, one cannot grasp the state of the art (or at least its idiosyncrasies) without examining the roots of the technology. So, it is with disk storage in Windows. We will not return to the absolute beginning of computer technology. Instead, we will begin slightly later to look at drive letters. The Windows discussion is oversimplified by design, our apologies to readers who know the development history and technical architecture better than we describe. Our emphasis, though, is on Linux, so we discuss Windows only in contrast.

Disk letters

In DOS,[1] a disk is assigned a letter. Traditionally, “A” is the boot floppy, “B” is the second floppy. It is a bit of a stretch to recall that this once was the extent of external storage available. To add non-removable disk storage, a hardware method was needed. The BIOS provides hard disk access via a hardware interrupt (INT13). Through the BIOS, a disk looks like one big file. DOS breaks up this storage into files and directories by means of a File Allocation Table (FAT). The first unused letter (historically speaking) was “C” thus the notion of the “C” drive being the first hard disk on an Intel-based system survives to this day. A subtle but important implication is that the BIOS has a single interrupt to service this storage type, so only one hard drive can be presented to the operating system at any given time. This is of no consequence to DOS, as it cannot multitask.

But the notion of an “attached volume” persisted even as complexity evolved. With the advent of multiple hard drives, each “volume” was independent of the others, there was no hierarchy that spanned multiple volumes. The operating system image had to be resident in memory to permit the mounted volume to be changed. Again, not an issue with DOS. The FAT was sufficient to locate a file on the attached volume, and the operating system had no intrinsic recourse to the disk once it was loaded. This scheme was adequate until the operating system evolved to use virtual memory, when it had occasion to use disk storage for its own purposes.

Until Windows 3.11, the underlying disk accesses were all handled through the BIOS by DOS. Windows NT employed a server-oriented design derived from OS/2 that was not constrained in this manner. Even so, Microsoft’s desire for commonality between desktop and server environments ensured the disk naming convention (if not the access method) would be carried into Windows NT and beyond.

[1] Using letters for disk attachment is not original to DOS, but that is not relevant to this introduction.

Disk partitions

Fundamental constraints inherited from earlier designs limited the addressable size of a mounted volume. Disk capacities outstripped physical addressability at a precipitous rate. It became necessary to subdivide individual hard drives into smaller, completely addressable pieces. The concept of disk partitioning is common to both Windows and Linux, and we will revisit it in subsequent sections. Suffice it to say that Windows partitioning permits up to four physical divisions of a disk. The last division may be subdivided, but these subdivisions are not visible at boot. Disks are not presented to Windows based on their adapter type or access protocol as is done in Linux.

Disk storage in the Linux environment

Although Linux is (rather emphatically) not UNIX, it is UNIX-like and shares the UNIX disk access paradigm. Linux, in its simplest invocation, needs only to be given its bootstrap and a “root device.” The operating system expects to find everything it needs to run on this root device, which can be a floppy disk, a ramdisk, a CD-ROM, hard drive, or network. In addition to the initial load of the kernel, are two necessary elements: a device driver that is particular to the media to be viewed as a disk, and a nexus in the file system by which the media can be accessed. It is in some sense a paradox that the operating system should need an entry in the file system to access the device on which it has a file system. This is because the access point of the device driver looks like a file. In fact, a design element of UNIX-like systems is that everything looks like a file such as disks, pipes, directories, and printers, which mainly look the same.

Characteristically, the disk devices are specified in the file system as “special” file entries in the /dev directory. With each of these “special” files is associated a set of indices (major and minor numbers, see Appendix A-1, “/proc/partitions” on page 259, or the output of the command **`ls -l /dev`**) that correlate to the physical attachment (for example, the controller, target, and logical unit) and the driver semantics associated with the desired access (for example, character or block mode access). The first device to be accessed is declared by the hardware and is termed “xxy” where xx indicates the device adapter type, and y enumerates the first device on that adapter. A SCSI disk, for example, might be seen as sda, while the second IDE hard disk might be seen as hdb. Once Linux is booted, we may expect to see entries such as /dev/sdb0 for the first partition of the second SCSI disk, or /dev/hda2 for the third (0,1,2...) partition on the first IDE hard disk, and so forth.

Disk partitions

Linux does not use the BIOS, but other systems that do might have to share a disk with Linux. Thus, Linux is aware of data structures and BIOS limitations that might influence such co-residency. For a purely Linux disk, these issues do not affect partition choices except for the adherence to four physical partitions. Since the topic of sharing a Linux disk with other operating systems (especially the boot disk) is well documented[2] elsewhere, we will not pursue it here. When unconstrained by co-residency requirements, Linux disks can be partitioned according to your taste, but there are some things to remember.

First, Linux disks, like Windows disks, can have only four physical partitions. Additional logical partitions can be created in one of these physical partitions, known as an extended partition, similar to the method used for Windows. Second, Linux disks are assigned names in a manner that can cause great difficulty if changes are made later. Lower SCSI ID numbers are assigned lower-order letters. If you remove one drive from the chain, the names of the higher ID number drives will change. A related note that we point out in several places in this book is to be careful to avoid gaps in the LUN numbering, as Linux will “stop counting” while probing when it reaches the first unavailable LUN.

Partitioning disks is done to achieve a physical separation of data according to type or function. A common tool to do this is **`fdisk`** but there are others such as **`cfdisk`** or even third-party tools like Partition Magic. No matter what the method, the goal is the same - to create a structure that can accommodate a file hierarchy. A hierarchical arrangement of directories (or folders) containing files can be as simple as a single tree that spans an entire disk. This is often done with trial installations but is not recommended for production use[3]. There are some useful sizing guidelines that we will discuss.

File hierarchy

Work by the Filesystem Hierarchy Standard Group has been published[4] to assist in interoperability of UNIX-like systems (including Linux). This group recommends segregating data into major subdivisions that derive from the shareability and variability of data.

For example, data that is shared between multiple systems should be segregated from data that is private to a single machine. Similarly, data that is highly variable should be segregated as much as possible from data that is static. Partitions

[2] See The Linux Documentation Project, Large Disk HOWTO and the Partition HOWTO at http://www.tldp.org

[3] To be clear, we are discussing the system disk. Additional disks can be subdivided as required to accommodate the applications.

[4] http://www.pathname.com/fhs/2.2/

enforce the implementation of this hierarchy by imposing physical constraints on the segments (files cannot span partitions) and allowing control over the access protocols to the segments (partitions are a convenient unit of access control as well as share policies, just as within Windows).

For a Linux example, /usr could be sharable and static, while /boot would be static, but unshareable. Similarly, /var/mail may be sharable and variable, while /var/lock would be variable but unshareable.

Regardless of the division policies, all branches of a hierarchical structure must eventually share a common point of origin, called the "root" (represented by a forward slash, "/"). The root partition contains enough information to boot, repair, and recover a system. Because system installation and recovery may require booting from smaller media, the root partition may be physically quite small. Essentially, the root partition contains directory entries, some of which may (or should) be separate physical partitions. The hierarchy is then assembled using the concept of "mounting" to be discussed in Appendix , "The file system table /etc/fstab" on page 262.

You can explore the rationale in more detail depending on your interest[5] but we suggest segregating partitions and size ranges for the following mount points:

- / - Small to medium (tens to 100s of MB)
- /usr - Large (several GB)
- /opt - Medium to large (100s of MB up to several GB)
- /var - Medium (100s of MB)
- /tmp - Medium (100s of MB) or larger, depending on application requirements
- /home - Very small to large, depending on user private space requirements

A separate partition not shown here (because it contains no file hierarchy) is declared for system swap, usually 1 or 2 times the size of system memory. Also, a small partition may be mounted on /boot for ease of location on systems with BIOS constraints causing addressing limitations.

There are some interesting constituents of the file hierarchy. We have briefly discussed the /dev directory,[6] where device special nodes are conventionally aggregated. In "Identifying your SCSI storage" on page 259 we will introduce the /proc pseudo-hierarchy. These are not disk partitions; they merely exploit the "everything-looks-like-a-file" paradigm discussed above. But every partition where you want to write a file hierarchy needs to be allocated a branch of the tree that begins with "root" (/). Their names can be arbitrary (although conventions exist), and they are managed via a table of entries in the plain-text file /etc/fstab. Before discussing the content of that file in "The file system table /etc/fstab" on

[5] *ibid.*

[6] Linux Allocated Devices, http://www.kernel.org/pub/linux/docs/device-list/devices.txt

page 262, let’s talk about possible file systems that may be selected for disk storage with Linux.

File systems

We have discussed a hierarchy of files, which is sometimes referred to as a file system. There is a more fundamental structure that underlies this hierarchy that is properly known as a file system, and it is important to choose one that meets your needs. Just as Windows has the FAT and NTFS file systems, Linux also has more than one type. In fact, the selection is quite rich. In addition to being able to read file systems from a number of other operating systems including Windows, UNIX and OS/2, Linux can reside in and make read/write use of several varieties including ext2, ext3, reiserfs, xfs, JFS, and others. What are these and why are there so many choices? The Linux System Administrator’s Handbook introduces file systems this way:

> “A file system is the methods and data structures that an operating system uses to keep track of files on a disk or partition; that is, the way the files are organized on the disk. The word is also used to refer to a partition or disk that is used to store the files or the type of the file system. Thus, one might say “I have two file systems" meaning one has two partitions on which one stores files, or that one is using the ``extended file system'', meaning the type of the file system.
>
> The difference between a disk or partition and the file system it contains is important. A few programs (including, reasonably enough, programs that create file systems) operate directly on the raw sectors of a disk or partition; if there is an existing file system there it will be destroyed or seriously corrupted. Most programs operate on a file system, and therefore will not work on a partition that does not contain one (or that contains one of the wrong type).
>
> Before a partition or disk can be used as a file system, it needs to be initialized, and the bookkeeping data structures need to be written to the disk. This process is called making a file system.”[7]

SuSE Linux uses the rreiserFS as its default file system, other distributions may default to ext2 or ext3. The reasons may be historical, but that is seldom the best criterion for selection. The ext2 file system has a 10-year production history, so its stability is attractive. It suffers from fragmentation when called upon to store a large number of small (a few kilobyte) files. The ext3 file system is ext2 plus journalling, however, the journalling implementation is dependent on kernel journalling, which can be considered fragile, yet ext3 can also journal both data and metadata, which is a feature not available with other journalling file systems. ReiserFS is a journalling file system that excels in both space and time efficiency

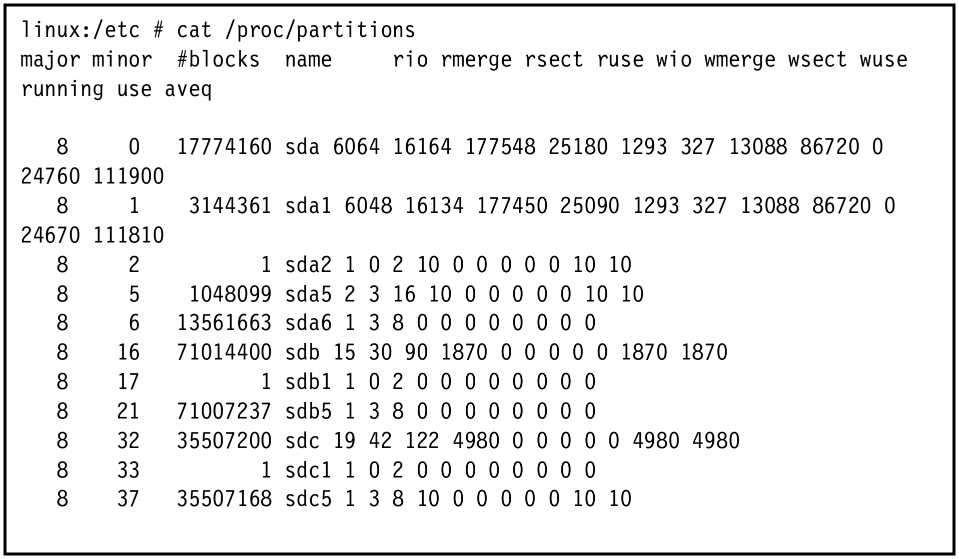
[7] http://tldp.org/LDP/sag/ - chapter 6.8

with variably-sized files because of its internal structure. IBM's JFS is very efficient in throughput for server-like environments with large files. Considerations imposed by higher-level disk aggregation programs and volume managers such as LVM, md, and Veritas VxVM are beyond the scope of this discussion, although their use may affect the choice of an underlying file system. Rather than suggest a default "one-size-fits-all" choice, we encourage you to research file system types to determine the best one for your application. Dan Robbins, CEO of gentoo, Inc. has posted a thirteen-part tutorial on file systems on the IBM developerWorks™ site[8]. Look for more useful information at IBM's developerWorks tutorial site[9].

Identifying your SCSI storage

The Linux /proc-file system offers you a large variety of possibilities to checkout your connections. In this appendix we want to introduce to them shortly.

To check out the SCSI HBAs operating in your system, view the file /proc/partitions. This output contains information about which SCSI storage device are detected, and to which device it is associated. Figure A-1 shows you the look-a-like in our example.



```
linux:/etc # cat /proc/partitions
major minor  #blocks  name     rio rmerge rsect ruse wio wmerge wsect wuse
running use aveq

   8     0   17774160 sda 6064 16164 177548 25180 1293 327 13088 86720 0
24760 111900
   8     1    3144361 sda1 6048 16134 177450 25090 1293 327 13088 86720 0
24670 111810
   8     2          1 sda2 1 0 2 10 0 0 0 0 0 10 10
   8     5    1048099 sda5 2 3 16 10 0 0 0 0 0 10 10
   8     6   13561663 sda6 1 3 8 0 0 0 0 0 0 0 0
   8    16   71014400 sdb 15 30 90 1870 0 0 0 0 0 1870 1870
   8    17          1 sdb1 1 0 2 0 0 0 0 0 0 0 0
   8    21   71007237 sdb5 1 3 8 0 0 0 0 0 0 0 0
   8    32   35507200 sdc 19 42 122 4980 0 0 0 0 0 4980 4980
   8    33          1 sdc1 1 0 2 0 0 0 0 0 0 0 0
   8    37   35507168 sdc5 1 3 8 10 0 0 0 0 0 10 10
```

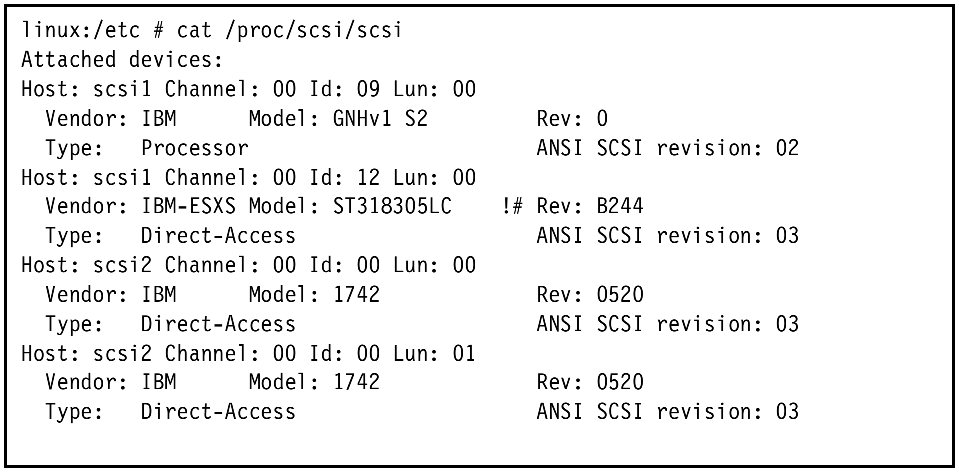
Figure A-1 /proc/partitions

[8] Start here: http://www-106.ibm.com/developerworks/linux/library/l-fs.html

[9] http://www-106.ibm.com/developerworks/views/linux/tutorials.jsp

The system has three disks attached (sda, sdb, and sdc). Each disk is partitioned.

The folder `/proc/scsi` contains informations about the SCSI adapter as well as about the attached components. For each SCSI HBA device driver is a folder in `/proc/scsi`. Within this folder you will find a file for each HBA using this driver. By viewing these files you can learn which adapter corresponds with which SCSI host ID. The file `scsi` informs you about the physical devices attached to your SCSI buses including ID and LUN. Figure 11-35 on page 260 shows an example: on SCSI HBA 2 is a device attached. The device is of the type 1742, which is the model number of the FAStT700.



```
linux:/etc # cat /proc/scsi/scsi
Attached devices:
Host: scsi1 Channel: 00 Id: 09 Lun: 00
  Vendor: IBM      Model: GNHv1 S2         Rev: 0
  Type:   Processor                        ANSI SCSI revision: 02
Host: scsi1 Channel: 00 Id: 12 Lun: 00
  Vendor: IBM-ESXS Model: ST318305LC    !# Rev: B244
  Type:   Direct-Access                    ANSI SCSI revision: 03
Host: scsi2 Channel: 00 Id: 00 Lun: 00
  Vendor: IBM      Model: 1742             Rev: 0520
  Type:   Direct-Access                    ANSI SCSI revision: 03
Host: scsi2 Channel: 00 Id: 00 Lun: 01
  Vendor: IBM      Model: 1742             Rev: 0520
  Type:   Direct-Access                    ANSI SCSI revision: 03
```

Figure 11-35 Content of /proc/scsi/scsi

This device provides two LUNs: 0 and 1. Actually, if the storage is invisible to your system this is the perfect spot to look for. Linux does not probe for LUNs beyond gaps. For example, during Storage Partitioning on a FAStT you can assign individual LUNs to the logical drives associated to a Host Group. However, if there is either a gap in between (e.g. 0,1,3,4) or the assignment incorrect (2,3,4,5) Linux does not detect all LUNs.

At the same time both files can help you finding errors in your SAN setup (for example, wrong port mapping causing seeing devices double).

For more in depth investigation you can collect informations about a generic devices by using the **`sg_scan`** command.

Figure 11-36 shows the sample output of the command.

```
linux:/etc # sg_scan
/dev/sg0: scsi1 channel=0 id=9 lun=0  type=3
/dev/sg1: scsi1 channel=0 id=12 lun=0  type=0
/dev/sg2: scsi2 channel=0 id=0 lun=0  type=0
/dev/sg3: scsi2 channel=0 id=0 lun=1  type=0
```

Figure 11-36 sg_scan

It delivers the generic devices attached including their ID and LUN. For further details, launch the command **sg_inq** with the corresponding device delivered by **sg_scan**.

```
linux:/etc # sg_inq /dev/sg2
standard INQUIRY:
  PQual=0, Device type=0, RMB=0, ANSI version=3, [full version=0x03]
  AERC=0, TrmTsk=0, NormACA=1, HiSUP=1, Resp data format=2, SCCS=0
  BQue=0, EncServ=1, MultiP=0, MChngr=0, ACKREQQ=0, Addr16=0
  RelAdr=0, WBus16=1, Sync=1, Linked=0, TranDis=0, CmdQue=1
    length=36 (0x24)
 Vendor identification: IBM
 Product identification: 1742
 Product revision level: 0520
 Product serial number: 1T23563008
```

Figure 11-37 sg_inq

Figure 11-37 shows the output of such an inquiry.

Using your storage on Linux

Once you have installed the OS and driver, the next step is to set up multipathing if it is available. Having a driver we were able to verify that the system “sees” the storage in a way that is planned or at least expected. For the OS, it does not matter how the storage is organized (single disks, logical drives in an array, etc.). All devices are addressed as a physical disk. This means that in order to mount and use the drive, appropriate partitioning is required. This is usually done with the command **fdisk /dev/sd?**

where ? stands for the device as stated in /proc/partitions (e.q. fdisk /dev/sdc). Once the drive is partitioned the new partition table has to be loaded. This is usually done by the boot of the system. However, during the implementation phase it is sometimes feasible to unload the device driver with the command **rmmod <modulename>**. The command **lsmod** lists you all loaded modules. Now reload the module with **modprobe <modulename>** again. Well, sure enough you might simply reboot the system.

The final step is writing a file system to the partition (if it is not used as raw device). To make a ext2 file system, for example, you simply have to enter the command **mkfs -t ext2 /dev/sd?*** where ? is the device number and * the partition number. In order to mount it, you need a mount point which you can create by the command **mkdir /path/my-mount-point**. After you have mounted the drive by **mount -t fs /dev/sd?* /path/my-mount-point** your storage is now accessible for the system. To make the mount persistent, make an entry in /etc/fstab.

To promote a new LUN to the /proc file system from the command line, use the following command:

```
echo “scsi add-single-device $host $channel $id $lun” > /proc/scsi/scsi
```

This is the proper way if you cannot afford to either unload the driver module or reboot the system.

The file system table /etc/fstab

Mounting a file system is nothing more than attaching a branch to the file hierarchy, where the branch is contained on a partition. As is common in Linux, many things can be “mounted” that are not truly partitions (for example, the /proc pseudo-file system). Our interest here is partitions, and the file that controls the attachment is /etc/fstab. The structure of this file is described in the online documentation (**man 5 fstab**) and we can synopsize it here with an example.

# filesystem	mount-point	fs-type	options	dump	fsck-order
/dev/sda0	/	reiserfs	rw	1	1
/devsda1	swap	swap	pri=1	0	0
proc	/proc	proc	defaults	0	0

Table A-1 Example /etc/fstab entries

The entry under the file system heading specifies the device (or pseudo device, as in the example of proc) to be attached (“mounted”). The mount point specifies the location in the tree to which the branch will be attached. The fs-type specifies the file system type that Linux should expect. Options vary according to file system type, we will not enumerate the options here. The final two fields specify the backup-priority and the order in which file systems are checked at boot. These fields are gradually losing currency as file systems and archiving strategies evolve. Reading the online documentation is your best source of further information on these fields, as well as the options field.

Glossary

A

Access logical drive. The Access Logical Drive is a special drive that uses none of the physical disk drives and should be assigned to the last (highest) available LUN number. Typically, the LUN number will be LUN 31. The Access Logical Drive allows the Storage Manager Agent to communicate to the Fibre Channel RAID controllers through the fibre connection for storage management services. These services include monitoring, configuring, and maintaining the RAID storage device.

allocated storage. On the ESS, this is the space that you have allocated to volumes, but not yet assigned.

array.The group of physical disk drive modules (DDMs) that are divided into logical drives and allocated to hosts.

assigned storage. On the ESS, this is the space that you have allocated to volumes, and assigned to a port.

asynchronous operation. A type of operation in which the remote copy XRC function copies updates to the secondary volume of an XRC pair at some time after the primary volume is updated. Contrast with synchronous operation.

availability. The degree to which a system or resource is capable of performing its normal function.

B

bay. Physical space on an ESS rack. A bay contains SCSI, ESCON, or Fibre Channel / FICON interface cards.

backup. The process of creating a copy of data to ensure against accidental loss.

C

cache. A random access electronic storage in selected storage controls used to retain frequently used data for faster access by the host.

CCW. Channel command word.

chunksize. The number of data blocks, assigned by a system administrator, written to the primary RAIDset or stripeset member before the remaining data blocks are written to the next RAIDset or stripeset member.

channel. (1) A path along which signals can be sent; for example, data channel and output channel. (2) A functional unit, controlled by the processor, that handles the transfer of data between processor storage and local peripheral equipment.

channel connection address (CCA). The input/output (I/O) address that uniquely identifies an I/O device to the channel during an I/O operation.

channel interface. The circuitry in a storage control that attaches storage paths to a host channel.

channel path. The ESA/390 term for the interconnection between a channel and its associated controllers.

channel subsystem. The ESA/390 term for the part of host computer that manages I/O communication between the program and any attached controllers.

CKD. Count key data (CKD) is the disk architecture used by zSeries (and S/390) servers. Because data records can be variable length, they all have a count field that indicates the record size. The key field is used to enable a hardware search on a key, however, this is not generally used for most data anymore. ECKD is a more recent version of CKD that uses an enhanced S/390 channel command set. The commands used by CKD are called Channel Command Words (CCWs); these are equivalent to the SCSI commands

CKD Server. This term is used to describe a zSeries 900 server (or a S/390 server) that is ESCON or FICON connected to the storage enclosure. In these environments the data in the storage enclosure for these servers is organized in CKD format. These servers run the z/OS, OS/390, MVS, z/VM, VM, VSE and TPF family of operating systems

cluster. See storage cluster

cluster processor complex (CPC). The unit within a cluster that provides the management function for the storage server. It consists of cluster processors, cluster memory, and related logic.

concurrent copy. A copy services function that produces a backup copy and allows concurrent access to data during the copy.

concurrent maintenance. The ability to service a unit while it is operational.

consistent copy. A copy of data entity (for example a logical volume) that contains the contents of the entire data entity from a single instant in time.

control unit address (CUA). The high order bits of the storage control address, used to identify the storage control to the host system.

Note: The control unit address bits are set to zeros for ESCON attachments.

CUA. Control unit address.

D

DA. The SSA loops of the ESS are physically and logically connected to the Device Adapters (also see device adapter)

DASD. Acronym for Direct Access Storage Device. This term is common in the z/OS environment to designate a disk or z/OS volume

DASD subsystem. A DASD storage control and its attached direct access storage devices.

data availability. The degree to which data is available when needed. For better data availability when you attach multiple hosts that share the same data storage, configure the data paths so that data transfer rates are balanced among the hosts.

data sharing. The ability of homogenous or divergent host systems to concurrently utilize information that they store on one or more storage devices. The storage facility allows configured storage to be accessible to any attached host systems, or to all. To use this capability, you need to design the host program to support data that it is sharing.

DDM. See disk drive module.

dedicated storage. Storage within a storage facility that is configured such that a single host system has exclusive access to the storage.

device. The zSeries 900 and S/390 term for a disk drive.

device adapter. A physical sub unit of a storage controller that provides the ability to attach to one or more interfaces used to communicate with the associated storage devices.

Device Support Facilities program (ICKDSF). A program used to initialize DASD at installation and perform media maintenance in zSeries 900 and S/390 environments.

DFDSS. Data Facility Data Set Services (see DFSMSdss)

DFSMSdss. A functional component of DFSMS/MVS used to copy, dump, move, and restore data sets and volumes.

disaster recovery. Recovery after a disaster, such as a fire, that destroys or otherwise disables a system. Disaster recovery techniques typically involve restoring data to a second (recovery) system, then using the recovery system in place of the destroyed or disabled application system. See also recovery, backup, and recovery system.

disk drive module (DDM). The primary nonvolatile storage medium that you use for any host data that is stored within a subsystem. Number and type of DDMs within a storage facility may vary. DDMs are the whole replaceable units (FRUs) that hold the HDDs

disk-group. In the ESS, a group of 7 or 8 DDMs that are not yet formatted as ranks.

disk 8-pack. In the ESS, a group of 7 or 8 DDMs that are not yet formatted as ranks.

drawer. A unit that contains multiple DDMs, and provides power, cooling, and related interconnection logic to make the DDMs accessible to attached host systems.

dump. A capture of valuable storage information at the time of an error.

disk. Disks (or maybe logical disks) are the logical representations of a SCSI or FCP disk as seen from the server system. In reality, a disk may span multiple physical disks, and the size of the disk is set when the disk is defined to the storage enclosure.

Disk drive. See disk drive module (DDM).

E

ESCON. Enterprise Systems Connection Architecture. An zSeries 900 and S/390 computer peripheral interface. The I/O interface utilizes S/390 logical protocols over a serial interface that configures attached units to a communication fabric.

ESCON Channel. The ESCON channel is a hardware feature on the zSeries and S/390 servers that controls data flow over the ESCON Link. An ESCON channel is usually installed on an ESCON channel card which may contain up to four ESCON channels

ESCON Host Adapter. The host adapter (HA) is the physical component of the storage server used to attach one or more host I/O interfaces. The ESS can be configured with ESCON host adapters (HA). The ESCON host adapter is connected to the ESCON channel and accepts the host CCWs (channel command words) that are sent by the host system.The ESCON HA in the ESS has two ports for ESCON channel connection.

ESCON Port. The ESCON port is the physical interface into the ESCON channel. An ESCON port has an ESCON connector interface. You have an ESCON port wherever you plug in an ESCON Link

ESCON Link. An ESCON link is the fiber connection between the zSeries 900 (or S/390) server and the storage enclosure. An ESCON link can also exist between a zSeries 900 (or S/390) processor and an ESCON Director (fibre switch), and between an ESCON Director and the storage enclosure (or other ESCON capable devices)

extended remote copy (XRC). A hardware- and software-based remote copy service option that provides an asynchronous volume copy across storage subsystems for disaster recovery, device migration, and workload migration.

F

Fabric Fibre Channel employs a fabric to connect devices. A fabric can be as simple as a single cable connecting two devices. The term is most often used to describe a more complex network utilizing hubs, switches and gateways.

Fabric-Loop Ports, FL_ports. These ports are just like the F_ports, except that they connect to an FC-AL topology. FL_ports can only attach to NL_ports. The ESS Fibre Channel adapters do not support the FL_port functionality, which is found only in fabrics or hubs.

Fabric Ports, F_ports. These ports are found in Fibre Channel Switched Fabrics. They are not the source or destination of IUs, but instead function only as a "middleman" to relay the IUs from the sender to the receiver. F_ports can only attach to N_ports. The ESS Fibre Channel adapters do not support the F_port functionality, which is found only in fabrics.

FC See Fibre Channel

FC-AL. Fibre Channel Arbitrated Loop. Description of Fibre Channel connection topology when SAN fabric consists of hubs. This implementation of the Fibre Channel standard uses a ring topology for the communication fabric.

FC Adapter. A FC Adapter is a card installed in a host system. It connects to the Fibre Channel (fibre) through a connector. The FC adapter allows data to be transferred over fibre links at very high speeds (currently 100 MB/s) and over greater distances than SCSI. According to its characteristics and its configuration, they allow the server to participate in different connectivity topologies

FC Host Adapter. The ESS has a FC Host Adapter (HA). The FC Host Adapter is connected to the Fibre Channel and accepts the FC commands that are sent by the host system

FCP. Fibre Channel Protocol. When we are mapping SCSI to the Fibre Channel transport (FC-4 Upper Layer) then we say FCP

FCS. See Fibre Channel standard

FC-SW. Fibre Channel Switched Fabric. Description of Fibre Channel connection topology when SAN fabric consists of switches.

Fibre Channel. Some people refer to Fibre Channel as the Fibre version of SCSI. Fibre Channel is capable of carrying IPI traffic, IP traffic, FICON traffic, FCP (SCSI) traffic, and possibly traffic using other protocols, all at the same level in the standard FC transport. Two types of cable can be used: copper and fiber. Copper for shorter distances and fiber for the longer distances

Fibre Channel Host. A server that uses Fibre Channel I/O adapter cards to connect to the storage enclosure

Fibre Channel standard. An ANSI standard for a computer peripheral interface. The I/O interface defines a protocol for communication over a serial interface that configures attached units to a communication fabric. The protocol has two layers. The IP layer defines basic interconnection protocols. The upper layer supports one or more logical protocols (for example FCP for SCSI command protocols, ESCON for ESA/390 command protocols).

FICON. An IO interface based on the Fibre Channel architecture. In this new interface, the ESCON protocols have been mapped to the FC-4 layer, i.e. the Upper Level Protocol layer, of the Fibre Channel Architecture. It is used in the S/390 and z/series environments.

FICON Channel. A channel that has a FICON channel-to-controller I/O interface that uses optical cables as a transmission medium.

Fixed Block Architecture (FBA). FC and SCSI disks use a fixed block architecture, that is, the disk is arranged in fixed size blocks or sectors. With an FB architecture the location of any block can be calculated to retrieve that block. The concept of tracks and cylinders also exists, because on a physical disk we have multiple blocks per track, and a cylinder is the group of tracks that exists under the disk heads at one point-in-time without doing a seek

FlashCopy. A point-in-time copy services function that can quickly copy data from a source location to a target location.

F_Node Fabric Node - a fabric attached node.

F_Port Fabric Port - a port used to attach a NodePort (N_Port) to a switch fabric.

G

GB. See gigabyte.

gigabyte. 1 073 741 824 bytes.

group. A group consist of eight DDMs.

H

hard disk drive (HDD). A nonvolatile storage medium within a storage server used to keep the data records. In the ESS these are 3.5” disks, coated with thin layers of special substances, where the information is magnetically recorded and read. HDDs are packed in replaceable units called DDMs.

HDD. See hard disk drive.

heterogeneous. Term used to describe an environment consisting of multiple types of hosts concurrently (i.e. Windows NT, Linux and IBM AIX)

homogenous. Term used to describe an environment consisting entirely of one type of host (i.e. Windows 2000 only)

host. The server where the application programs run.

host adapter (HA). A physical sub unit of a storage controller that provides the ability to attach to one or more host I/O interfaces.

I

ICKDSF. See Device Support Facilities program.

Intel Servers. When we say Intel servers (or Intel based servers) we are referring to all the different server makes that run on Intel processors. This includes servers running Windows NT and Windows 2000, as well as Novell Netware and Linux. This term applies to the IBM Netfinity servers, the most recent xSeries family of the IBM @serverbrand, and the various non-IBM server makes available on the market that run on Intel processors

Internet Protocol (IP). A protocol used to route data from its source to its destination in an Internet environment.

I/O device. An addressable input/output unit, such as a direct access storage device, magnetic tape device, or printer.

I/O interface. An interface that you define in order to allow a host to perform read and write operations with its associated peripheral devices.

ITSO. International Technical Support Organization

J

JBOD. Just a Bunch Of Disks. A disk group configured without the disk redundancy of the RAID-5 arrangement. When configured as JBOD, each disk in the disk group is a rank in itself.

JCL. See job control language.

Job control language (JCL). A problem-oriented language used to identify the job or describe its requirements to an operating system.

K

KB. See kilot .

ioeer.Oner tho(s)-6.9 andemee(s)17.5(.)]TJ /F1 1 Tf 9.908

logical volume. The storage medium associated with a logical disk. A logical volume typically resides on one or more DDMs. For CKD the logical volume size is defined by the device emulation mode (3390 or 3380 track format). For open systems hosts, the size is 0.5 GB to the maximum capacity of a rank

LUN. See logical unit number

M

MB. See megabyte.

megabyte (MB). 1 048 576 bytes.

mirrorset. A RAID storageset of two or more physical disks that maintains a complete and independent copy of the entire virtual disk's data. Also referred to as RAID-1

N

native Linux. An environment where Linux is installed and operating on the native hardware platform. This is the most common of Linux implementations however, the distinction needs to be drawn as the zSeries servers allow for Linux to be run as a guest under the z/VM (or VM) native operating system

Node-Loop Ports, NL_ports.These ports are just like the N_ports, except that they connect to a Fibre Channel Arbitrated Loop (FC-AL) topology. NL_ports can only attach to other NL_ports or to FL_ports. The ESS Fibre Channel adapters support the NL_port functionality when connected directly to a loop.

Node Ports, N_ports. These ports are found in Fibre Channel Nodes, which are defined to be the source or destination of Information Units (IUs). I/O devices and host systems interconnected in point-to-point or switched topologies use N_ports for their connections. N_ports can only attach to other N_ports or to F_ports. The ESS Fibre Channel adapters support the N_port functionality when connected directly to a host or to a fabric.

non-disruptive. The attribute of an action or activity that does not result in the loss of any existing capability or resource, from the customer's perspective.

nonvolatile storage (NVS). Random access electronic storage with a backup battery power source, used to retain data during a power failure. Nonvolatile storage, accessible from all cached IBM storage clusters, stores data during DASD fast write, dual copy, and remote copy operations.

NVS. See Nonvolatile storage.

O

open systems. (or open servers) refers to the Windows NT, WIndows 2000, Linux, and various flavors of UNIX operating environments. Both for IBM and non IBM servers

operating system. Software that controls the execution of programs. An operating system may provide services such as resource allocation, scheduling, input/output control, and data management.

P

peer-to-peer remote copy (PPRC). A hardware based remote copy option that provides a synchronous volume copy across storage subsystems for disaster recovery, device migration, and workload migration.

port. (1) An access point for data entry or exit. (2) A receptacle on a device to which a cable for another device is attached.

PPRC. See peer-to-peer remote copy.

primary device. One device of a dual copy or remote copy volume pair. All channel commands to the copy logical volume are directed to the primary device. The data on the primary device is duplicated on the secondary device. See also secondary device.

PTF. Program temporary fix. A fix to a bug in a program or routine.

R

rack. A unit that houses the components of a storage subsystem, such as controllers, disk drives, and power.

RAIDset. A storageset that stripes data and parity across three or more members in a disk array. Also referred to as RAID-5

rank. A disk group upon which a RAID-5 array is configured. For JBOD, each DDM becomes a rank.

random access. A mode of accessing data on a medium in a manner that requires the storage device to access nonconsecutive storage locations on the medium

RDAC. Redundant Disk Array Controller. Controller failover facility provided for some operating systems with the FAStT product line

read hit. When data requested by the read operation is in the cache.

read miss. When data requested by the read operation is not in the cache.

recovery. The process of rebuilding data after it has been damaged or destroyed. In the case of remote copy, this involves applying data from secondary volume copies.

recovery system. A system that is used in place of a primary application system that is no longer available for use. Data from the application system must be available for use on the recovery system. This is usually accomplished through backup and recovery techniques, or through various DASD copying techniques, such as remote copy.

ReiserFS. Reiser FS is a journaling filesystem which means logs all changes to a disk and can play them back after a system failure. Hence long filesystem checks like with **ext2** are obsolete**.**

remote copy. A storage-based disaster recovery and workload migration function that can copy data in real time to a remote location. Two options of remote copy are available. See peer-to-peer remote copy and extended remote copy.

restore. Synonym for recover.

re-synchronization. A track image copy from the primary volume to the secondary volume of only the tracks which have changed since the volume was last in duplex mode.

S

SCSI. SCSI (Small Computer Systems Interface) is the protocol that the SCSI adapter cards use. Although SCSI protocols can be used on Fibre Channel (then called FCP) most people mean the parallel interface when they say SCSI. This is the ANSI standard for a logical interface to computer peripherals and for a computer peripheral interface. The interface utilizes a SCSI logical protocol over an I/O interface that configures attached targets and initiators in a multi-drop topology.

SCSI adapter. A SCSI adapter is an I/O card installed in a host system. It connects to the SCSI bus through a SCSI connector. There are different versions of SCSI, some of which can be supported by the same adapter. The protocols that are used on the SCSI adapter (the command set) can be either SCSI-2 or SCSI-3

SCSI ID. An unique identifier (ID) assigned to a SCSI device, that is used in protocols on the SCSI interface to identify or select the device. The number of data bits on the SCSI bus determines the number of available SCSI IDs. A wide interface has 16 bits, with 16 possible IDs. A SCSI device is either an initiator or a target

SCSI host. An open systems server that uses SCSI adapters to connect to the storage enclosure

SCSI host adapter. The host adapter (HA) is the physical component of the storage server used to attach one or more host I/O interfaces.The ESS can be configured with SCSI host adapters (HA). The SCSI host adapter is connected to the SCSI bus and accepts the SCSI commands that are sent by the host system.The SCSI HA has two ports for SCSI bus connection

SCSI port. A SCSI Port is the physical interface into which you connect a SCSI cable. The physical interface varies, depending on what level of SCSI is supported

Seascape architecture. A storage system architecture developed by IBM for open system servers and S/390 host systems. It provides modular storage solutions that integrate software, storage management, and technology for disk, tape, and optical storage.

secondary device. One of the devices in a dual copy or remote copy logical volume pair that contains a duplicate of the data on the primary device.

server. A type of host that provides certain services to other hosts that are referred to as clients.

spare. A disk drive that is used to receive data from a device that has experienced a failure that requires disruptive service. A spare can be pre-designated to allow automatic dynamic sparing. Any data on a disk drive that you use as a spare is destroyed by the dynamic sparing copy process.

Spareset. A collection of disk drives made ready by the controller to replace failed members of a storageset.

SSA. Serial Storage Architecture. An IBM standard for a computer peripheral interface. The interface uses a SCSI logical protocol over a serial interface that configures attached targets and initiators in a ring topology.

SSR. System Support Representative. The IBM person who does the hardware installation and maintenance.

storage cluster. A partition of a storage server that is capable of performing all functions of a storage server. When a cluster fails in a multiple-cluster storage server, any remaining clusters in the configuration can take over the processes of the cluster that fails (fail-over)

storage device. A physical unit which provides a mechanism to store data on a given medium such that it can be subsequently retrieved.

storage server. A unit that manages attached storage devices and provides access to that storage and storage-related functions for one or more attached host systems

Striped mirrorset. A storageset that stripes data across an array of two or more mirrorsets. Also referred to as RAID-0+1

Stripeset. A storageset that stripes data across an array of two or more disk drives. Also referred to as RAID-0

striping. A technique that distributes data in bit, byte, multi-byte, record, or block increments across multiple disk drives.

synchronization. An initial volume copy. This is a track image copy of each primary track on the volume to the secondary volume.

synchronous operation. A type of operation in which the remote copy PPRC function copies updates to the secondary volume of a PPRC pair at the same time that the primary volume is updated. Contrast with asynchronous operation.

T

TCP/IP. Transmission Control Protocol/Internet Protocol.

timeout. The time in seconds that the storage control remains in a “long busy” condition before physical sessions are ended.

U

Ultra-SCSI. An enhanced small computer system interface.

V

volume. An ESA/390 term for the information recorded on a single unit of recording medium. Indirectly, it can refer to the unit of recording medium itself. On a non-removable medium storage device, the terms may also refer, indirectly, to the storage device that you associate with the volume. When you store multiple volumes on a single storage medium transparently to the program, you may refer to the volumes as logical volumes.

VTOC. Volume table of contents. In a DASD, the place where the space and allocation information of the volume is maintained.

W

World Wide Name (WWN). unique number assigned to Fibre Channel devices (including hosts and I/O adapter ports) - analogous to a MAC address on a network card.

write hit. A write operation where the data requested is in the cache.

write miss. A write operation where the data requested is not in the cache.

X

XRC. Extended remote copy.

Y

YaST. Acronym for Yet another Setup Tool. YaST provides integrated user and group administration and simplifies system administration.

Z

z/Architecture. The IBM 64-bit real architecture implemented in the new IBM eserver zSeries 900 enterprise e-business servers.

z/OS. The IBM operating systems for the z/Architecture family of processors.

zSeries 900. Refers to the zSeries 900 family of servers from the IBM eserver brand. Also referred to as zSeries. These servers are the successors to the IBM 9672 G5 and G6 family of processors.

Related publications

The publications listed in this section are considered particularly suitable for a more detailed discussion of the topics covered in this redbook.

IBM Redbooks

For information on ordering these publications, see "How to get IBM Redbooks" on page 278. Note that some of the documents referenced here may be available in softcopy only:

- *IBM TotalStorage Enterprise Storage Server Model 800*, SG24-6424
- *IBM TotalStorage Enterprise Storage Server: Implementing the ESS in Your Environment*, SG24-5420
- *IBM TotalStorage Enterprise Storage Server Web Interface User's Guide*, SC26-7448
- *Fibre Array Storage Technology - A FAStT Introduction*, SG24-6246
- *IBM TotalStorage FAStT700 and Copy Services*, SG24-6808
- *Linux for S/390*, SG24-4987
- *Linux for IBM zSeries and S/390: Distributions*, SG24-6264
- *Linux on IBM zSeries and S/390: Large Scale Linux Deployment*, SG24-6824
- *Tuning IBM eServer xSeries Servers for Performance*, SG24-5287

- *IBM Storage Area Network, SG24-6419-00* , SG24-5420
- *The Cutting Edge: The IBM eServer BladeCenter*, REDP-3581
- *Getting Started with zSeries Fibre Channel Protocol*, REDP-0205

Other publications

These publications are also relevant as further information sources:

- *IBM TotalStorage Enterprise Storage Server Web Interface User's Guide*, SC26-7448
- *IBM TotalStorage Enterprise Storeage Server: Subsystem Device Driver User's Guide,* SC26--7478

Online resources

These Web sites and URLs are also relevant as further information sources:

- IBM Linux Web site

  http://www.ibm.com/linux/
- The Linux Technology Center

  http://www.ibm.com/linux/ltc
- The IBM TotalStorage Web site

  http://www.storage.ibm.com/
- The IBM TotalStorage SAN fabric Web site

  http://www.storage.ibm.com/ibmsan/products/sanfabric.html
- The IBM eServer Web site

  http://www.ibm.com/eserver

How to get IBM Redbooks

You can search for, view, or download 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

Index

Implementing Linux with IBM Disk Storage

(0.5” spine)
0.475”<->0.875”
250 <-> 459 pages

Implementing Linux with IBM Disk Storage

Your guide to Linux implementations on IBM eServers

Use Linux efficiently with FAStT and ESS

Explore options relating to Linux in SANs

This IBM Redbook explains considerations, requirements, pitfalls, and possibilities when implementing Linux with IBM disk storage products. This redbook presents the reader with a practical overview of the tasks involved in installing Linux on a variety of IBM @server platforms, and it also introduces the people who are already familiar with Linux to the powerful IBM TotalStorage servers: ESS and FAStT. The book also provides the steps required to prepare the storage system for Linux host attachment.

Among other things, we review the recently added support for FCP attachment when running SuSE Linux (SLES 8) in an LPAR on a zSeries machine.

On pSeries we cover the implementation of SLES 8, either natively or in an LPAR on a p690, and explain how to configure the Emulex Host Bus Adapter, for attachment to FAStT and ESS.

For Intel based machines, we look at the implementation of Red Hat Enterprise AS 2.1 and SLES 8 on xSeries servers and BladeCenter; in addition, we explain how to set up the Red Hat High Availability cluster using shared external storage.

The book also contains a chapter on the VMware ESX Server, focusing on the practical aspects of its implementation and configuration with external storage.

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

SG24-6261-01 ISBN 0738453471