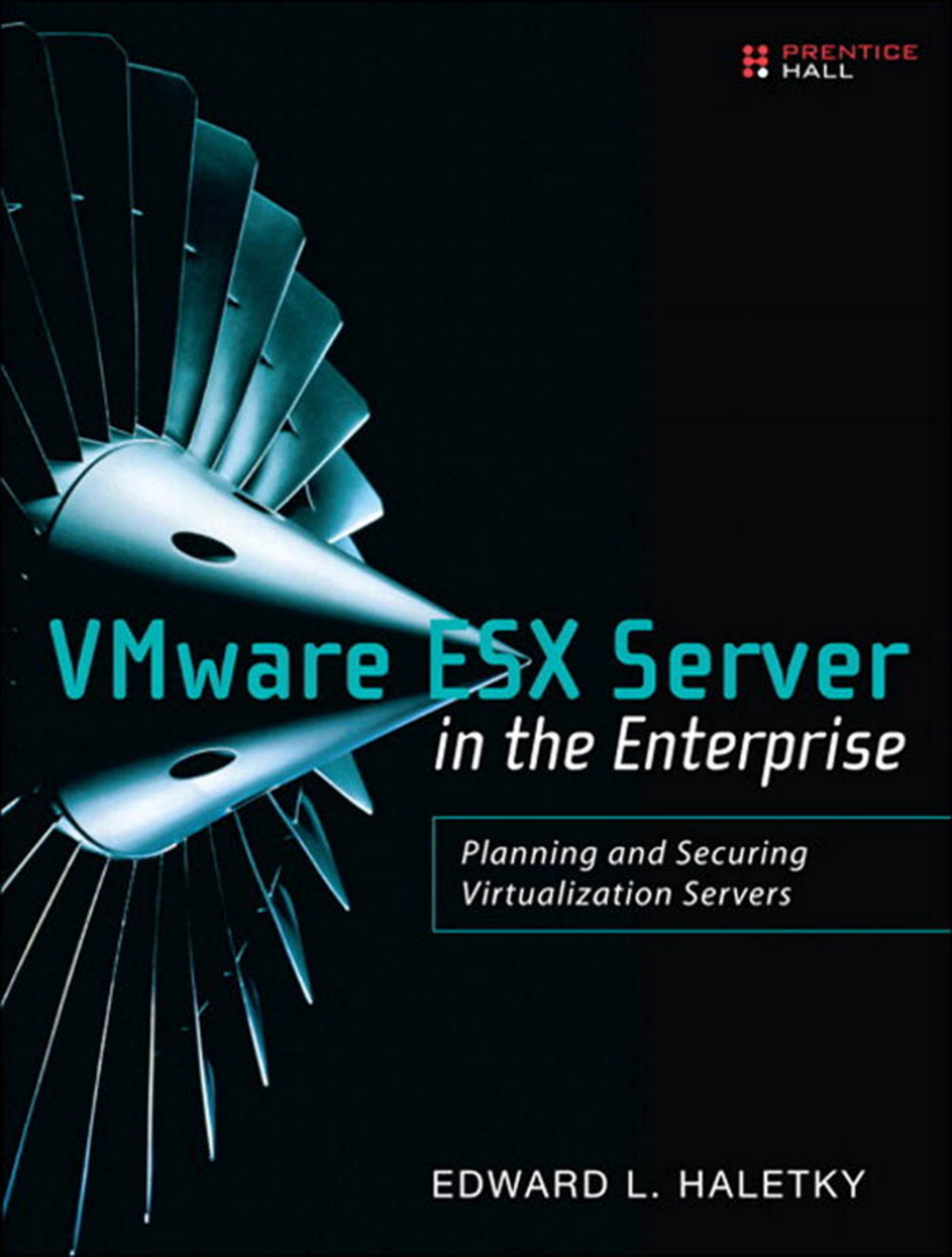
PRENTICE
HALL
VMware ESX Server
in the Enterprise
Planning and Securing
Virtualization Servers
EDWARD L. HALETKY

VMWARE ESX SERVER IN THE ENTERPRISE

This page intentionally left blank

VMWARE ESX SERVER IN THE ENTERPRISE

PLANNING AND SECURING VIRTUALIZATION SERVERS

EDWARD L. HALETKY

Upper Saddle River, NJ • Boston • Indianapolis • San Francisco
New York • Toronto • Montreal • London • Munich • Paris • Madrid
Cape Town • Sydney • Tokyo • Singapore • Mexico City

Many of the designations used by manufacturers and sellers to distinguish their products are claimed as trademarks. Where those designations appear in this book, and the publisher was aware of a trademark claim, the designations have been printed with initial capital letters or in all capitals.

The author and publisher have taken care in the preparation of this book, but make no expressed or implied warranty of any kind and assume no responsibility for errors or omissions. No liability is assumed for incidental or consequential damages in connection with or arising out of the use of the information or programs contained herein.

The publisher offers excellent discounts on this book when ordered in quantity for bulk purchases or special sales, which may include electronic versions and/or custom covers and content particular to your business, training goals, marketing focus, and branding interests. For more information, please contact:

U.S. Corporate and Government Sales
(800) 382-3419
corpsales@pearsontechgroup.com

For sales outside the United States please contact:

International Sales
international@pearsoned.com

Library of Congress Cataloging-in-Publication Data:

Haletky, Edward.
VMware ESX server in the enterprise : designing and securing virtualization servers / Edward Haletky.— 1st ed.
p. cm.
Includes bibliographical references.
ISBN 0-13-230207-1 (pbk. : alk. paper) 1. VMware. 2. Virtual computer systems. 3. Operating systems (Computers) I. Title.
QA76.9.V5H35 2007
005.4'3—dc22

2007044443

ISBN-13: 978-0-13-230207-4
ISBN-10: 0-13-230207-1

Text printed in the United States on recycled paper at R. R. Donnelley in Crawfordsville, Indiana.
First printing December 2007

Editor-in-Chief
Karen Gettman

Acquisitions Editor
Jessica Goldstein

Senior Development Editor
Chris Zahn

Managing Editor
Gina Kanouse

Project Editor
Andy Beaster

Copy Editor
Keith Cline

Indexer
Erika Millen

Publishing Coordinator
Romny French

Cover Designer
Chuti Prasertsith

Composition and Proofreading
Fastpages

This Book Is Safari Enabled

The Safari® Enabled icon on the cover of your favorite technology book means the book is available through Safari Bookshelf. When you buy this book, you get free access to the online edition for 45 days.

Safari Bookshelf is an electronic reference library that lets you easily search thousands of technical books, find code samples, download chapters, and access technical information whenever and wherever you need it.

To gain 45-day Safari Enabled access to this book:

- Go to http://www.prenhallprofessional.com/safarienabled
- Complete the brief registration form
- Enter the coupon code 4BA2-YYDB-7R37-E4BV-6Q29

If you have difficulty registering on Safari Bookshelf or accessing the online edition, please e-mail customer-service@safaribooksonline.com.

Visit us on the Web: www.prenhallprofessional.com

To my wife, whom I thank for her support and understanding.

This page intentionally left blank

Table of Contents

PREFACE xvii

1 SYSTEM CONSIDERATIONS 1

Basic Hardware Considerations 2
- Processor Considerations 2
- Cache Considerations 4
- Memory Considerations 6
- I/O Card Considerations 7
- Disk Drive Space Considerations 10
- Basic Hardware Considerations Summary 12

Specific Hardware Considerations 13
- Blade Server Systems 13
- 1U Server Systems 14
- 2U Server Systems 15
- Large Server-Class Systems 17
- The Effects of External Storage 18

Examples 22
- Example 1: Existing Datacenter 22
- Example 2: Office in a Box 23
- Example 3: The Latest and Greatest 25
- Example 4: The SAN 26
- Example 5: Secure Environment 27
- Example 6: Disaster Recovery 28

Hardware Checklist 29

Extra Hardware Considerations 30

Conclusion 31

2 VERSION COMPARISON 33

VMkernel Differences 34

Console Operating System/Service Console 37
- Version Differences 37
- Tool Differences 40

Virtual Networking 41

Data Stores 43

Virtual Resources 44

Backup Differences 46

Licensing Differences 47

VMware Certified Professional 48

Virtual Hardware 48

Virtual Machine and Server Management 50

Security Differences 51
Installation Differences 52
Conclusion 53

3 INSTALLATION 55

Upgrade Steps 57
Step 1: Back Up ESX 58
Step 2: Read the Release Notes 58
Step 3: Perform a Pre-Upgrade Test 59
Step 4: Prepare Your ESX Server 59
Installation Steps 60
Step 1: Read the Release Notes 60
Step 2: Read All Relevant Documentation 60
Step 3: Is Support Available for the Hardware Configuration? 60
Step 4: Verify the Hardware 61
Step 5: Are the Firmware Levels at Least Minimally Supported? 61
Step 6: Is the System and Peripheral BIOS correctly set? 62
Step 7: Where Do You Want the Boot Disk Located? 63
Step 8: VMware ESX Server License 63
Step 9: VM License and Installation Materials 64
Step 10: Service Console Network Information 64
Step 11: Memory Allocated to the Service Console 64
Step 12: VMkernel Swap Size (ESX Versions Earlier Than Version 3 Only) 65
Step 13: PCI Device Allocations 66
Step 14: File System Layouts 66
Step 15: Configure the Server and the FC HBA to Boot from SAN or Boot from iSCSI 68
Setting Up the Emulex HBAs 71
Setting Up QLogic FC HBAs 74
ESX Version 3 76
Set Up Boot from iSCSI 76
Step 16: Start ESX Server Installations 76
Step 17: Connecting to the Management User Interface for the First Time 83
Step 18: Additional Software Packages to Install 86
Automating Installation 87
Conclusion 92

4 AUDITING, MONITORING, AND SECURING 93

VM Security 95
Security Recipe 98
Step 1 103
Step 2 103
Step 3 104
ESX version 3 104
ESX Version 2.5.2 and Earlier Than Version 3 104
ESX Versions Earlier Than 2.5.2 106
Step 4 106
ESX Version 3 108
ESX Version 2.5.2 and Earlier 109
Step 5 110
ESX Version 3 110
ESX Version 2.5.2 and Earlier 110
VM Use of iptables 111

Step 6 111
Step 6A 112
Step 6B 124
Step 7 130
Auditing Recipe 130
Monitoring Recipe 137
ESX-Specific Security Concerns 152
VMkernel Security Considerations 152
VMotion Security Considerations 153
Other ESX Security Considerations 153
Roles and Permissions 156
What to Do If There Is a Break-In 161
Conclusion 162

5 STORAGE WITH ESX 163

Overview of Storage Technology with ESX 164
SCSI Versus SAS Versus ATA Versus SATA 165
iSCSI (SCSI over IP) 166
NAS (Network-Attached Storage) 167
SANs (Storage Area Networks) 168
Gigabit Interface Converters (GBICs) 168
Fibre Channel Switch 168
Bridge 169
Topologies 169
Point-to-Point Topology 169
Switched Fabric 170
Arbitrated Loop 172
Multipath Fabric 173
Redundant Fabric 174
Other SAN Options and Components 174
Replication 174
Zoning 176
Multipath 177
Overview of Storage Technology with ESX 179
SAN Best Practices 179
iSCSI/NFS Best Practices 180
Virtual Machine File System 180
VMDK and VMFS Manipulation 182
VMFS Types 183
Structure of VMFS 183
VMFS Accessibility Modes 184
Raw Disk Maps 185
SCSI Pass Thru 185
VMFS Error Conditions 185
SCSI: 24/0 185
SCSI: 24/7 186
SCSI: 2/0 186
SCSI: 0/1 186
SCSI: 0/4 186
SCSI: 0/7 186
0xbad0010 186
0xbad0023 186
0xbad0004 189
Storage Checklist 189
Assessing Storage and Space Requirements 191
Example of LUN Sizing 195
Storage-Specific Issues 195
Increasing the Size of a VMDK 196
Increasing the Size of a VMFS 197
Adding Extents 197
Deleting Extents 197
Searching for New LUNs 198
ESX Version 3 198
ESX Versions Earlier Than Version 3 198

VMFS Created on One ESX Server Not Appearing on Another 198
How to Unlock a LUN 200
Boot from SAN or iSCSI 200
Conclusion 201

6 EFFECTS ON OPERATIONS 203

Data Store Performance or Bandwidth Issues 204
SCSI-2 Reservation Issues 204
Performance-Gathering Agents 211
Other Operational Issues 213
Sarbanes-Oxley 214
Conclusion 215

7 NETWORKING 217

Basic Building Blocks 217
Details of the Building Blocks 219
Physical Switch (pSwitch) 220
Physical NIC (pNIC) 221
Virtual Switch (vSwitch) 221
Virtual NIC (vNIC) 223
Virtual Firewall (vFW) 223
Virtual Router (vRouter) 224
Virtual Gateway (vGateway) 224
Network Definitions 224
Administration Network 224
VMotion Network 228
VMkernel Network 228
VM Network 229
Checklist 230
vSwitch Settings 232
802.1q 232
Load-Balancing pNICs 234
ESX Version 3 234
ESX Version 2.5.x or Earlier 235
Redundant pNICs 235
ESX Version 3 236
ESX Version 2.5.x and Earlier 236
Beacon Monitoring 236
ESX Version 3 237
ESX Version 2.5.x 237
Traffic Shaping 238
Average Bandwidth 238
Peak Bandwidth 238
Burst Size 238
pSwitch Settings 239
Example vNetworks 239
Configuration 243
Conclusion 244

8 CONFIGURING ESX FROM A HOST CONNECTION 245

Configuration Tasks 246
Server-Specific Tasks 246
Administrative User 246
Command Line 247
Local User 247
Non-Active Directory Network User 247
NIS 248
Kerberos 248
LDAP 248
LDAP User Information 248
Active Directory Domain User 249
ESX Version 3 249
ESX Versions Earlier Than Version 3 255
Virtual Infrastructure Client 255
Management User Interface (MUI) 258

Security Configuration 259
Command Line 259
ESX Version 3 259
ESX Versions Earlier Than Version 3 259
VIC20 260
MUI 262
Network Time Protocol (NTP) 263
Service Console Memory 264
Command Line 265
ESX Version 3 265
Earlier Versions of ESX 266
VIC20 266
MUI 268
Conclusion 269

9 Configuring ESX from a Virtual Center or Host 271

Configuration Tasks 271
Add Host to VC 272
Licensing 273
Command Line 273
ESX Version 3 274
Earlier Versions of ESX 274
VIC 274
MUI 276
Virtual Swap 277
Command Line 277
VIC 277
MUI 277
Local VMFS 278
Command Line 279
Renaming a VMFS 279
Creating a VMFS 279
Extending a VMFS 281
Deleting a VMFS Extent 282
Deleting a VMFS 282
VIC 284
Renaming a VMFS 284
Creating a VMFS 286
Extending a VMFS 291
Deleting a VMFS Extent 293
Deleting a VMFS 294
MUI 296
Renaming a VMFS 296
Creating a VMFS 297
Extending a VMFS 298
Deleting a VMFS Extent 298
Deleting a VMFS 299
FC HBA VMFS 299
Finding the WWPN 300
Command Line 300
VIC 300
MUI 301
Masking and Max LUN Manipulations 302
Command Line 302
ESX Version 3 303
ESX Versions Earlier Than Version 3 303
VIC 304
MUI 305
Deleting a VMFS Extent 305
Command Line 306
ESX Versions Earlier Than Version 3 306
VIC 306
MUI 307
Virtual Networking 307
Configuring the Service Console 308
Command Line 308
ESX Version 3 308
ESX Versions Earlier Than Version 3 309

VIC 310
MUI 310
Creating a VM Network
vSwitch 311
Command Line 311
VIC 311
MUI 313
Creating a VMotion vSwitch 315
Command Line 315
VIC 316
MUI 316
Adding an iSCSI Network to the Service Console vSwitch 318
Command Line 318
VIC 318
Adding a NAS vSwitch for Use by NFS 319
Command Line 319
VIC 320
Adding a Private vSwitch 320
Command Line 320
VIC 321
MUI 321
Adding Additional pNICs to a vSwitch 321
Command Line 321
VIC 322
MUI 322
Adding vSwitch Port Groups 322
Command Line 322
VIC 323
MUI 323
Removing vSwitch Port Groups 323
Command Line 323
VIC 323
MUI 324
vSwitch Removal 324
Command Line 324
VIC 324
MUI 324
vSwitch Security 325
Command Line 325
VIC 326
MUI 327
vSwitch Properties 327
Command Line 328
VIC 328
MUI 330
Changing VMkernel Gateways 330
Command Line 330
VIC 330
Changing pNIC Settings 331
Command Line 331
VIC 332
MUI 332
Changing Traffic-Shaping Settings 333
Command Line 333
VIC 334
MUI 335
iSCSI VMFS 335
Command Line 335
VIC 336
Network-Attached Storage 338
Command Line 338
VIC 339
Mapping Information 340
Secure Access to Management Interfaces 342
Advanced Settings 343
Conclusion 344

10 Virtual Machines 345

Overview of Virtual Hardware 345

Creating VMs 349
VM Creation from VIC 351
VM Creation from VC1.X 359
VM Creation from MUI 364
VM Creation from Command Line 368
Installing VMs 374
Using Local to the ESX Server CD-ROMs 374
MUI 374
Command Line 375
VC1.x/VIC 375
Using a Local or Shared ESX Server ISO Image 376
MUI 376
Command Line 377
VC1.x/VIC 377
Using Client Device or ISO 377
Command Line 378
VIC 378
Importance of DVD/CD-ROM Devices 379
Other Installation Options 379
Special Situations 380
Virtual Guest Tagging Driver 380
Virtual Hardware for Nondisk SCSI Devices 380
Command Line 381
VIC 382
MUI 383
Virtual Hardware for Raw Disk Map Access to Remote SCSI 384
Command Line 384
VIC 384
MUI 385
Virtual Hardware for RDM-Like Access to Local SCSI 385
VM Disk Modes and Snapshots 387
Command line 391
OS Installation Peculiarities 392
Cloning, Templates, and Deploying VMs 393
VM Solutions 393
Private Lab 394
Firewalled Private Lab 394
Firewalled Lab Bench 394
Cluster in a Box 395
Cluster between ESX Servers 396
Cluster between Virtual and Physical Servers 396
VC as a VM 396
Virtual Appliances 397
VMware Tools 400
VMX Changes 401
Conclusion 405

11 Dynamic Resource Load Balancing 407

Defining DRLB 407
The Basics 408
The Advanced Features 411
Monitoring 415
Alarms 415
Performance Analysis 419
Shares 426
Resource Pool Addendum 428
Putting It All Together 429
Conclusion 430

12 Disaster Recovery and Backup 431

Disaster Types 432
Recovery Methods 435
Best Practices 437
Backup and Business Continuity 439

Backup 439
Backup Paths 441
Modification to Path 3 442
Additional Hot Site Backup Paths 444
Summary of Backup 446
Business Continuity 446
The Tools 447
Simple Backup Scripts 449
ESX version 3 449
ESX Version 2 452
ESX Version Earlier Than 2.5.0 453
Local Tape Devices 453
Vendor Tools 457
ESXRanger 458
HPSIM VMM 458
Other Tools 459
Conclusion 460

EPILOGUE
THE FUTURE OF VIRTUALIZATION 461

APPENDIX A
SECURITY SCRIPT 465

APPENDIX B
ESX VERSION 3 TEXT INSTALLATION 481

APPENDIX C
ESX VERSION 3 GRAPHICAL INSTALLATION 501

REFERENCES 519

INDEX 521

Acknowledgments

I would like to acknowledge my original coauthors: Paul and Sumithra. Although they were not able to complete the work, I am very thankful for their early assistance and insights into the world of virtualization. I would also like to thank my reviewers; they provided great feedback. I would like to also thank Bob, once a manager, who was the person who started me on this journey by asking one day 'Have you ever heard of this VMware stuff?' I had. This book is the result of many a discussion I had with customers and my extended team members, who took a little extra work so that I could concentrate on virtualization. I want to thank Greg from the Hewlett-Packard Storage team for his help and insights into SCSI Reservation Conflicts. Last but not least, I would like to acknowledge my editors. Thank you one and all.

About the Author

Edward L. Haletky graduated from Purdue University with a degree in aeronautical and astronautical engineering. Since then, he has worked programming graphics and other lower-level libraries on various UNIX platforms. Edward recently left Hewlett-Packard, where he worked on the Virtualization, Linux, and High-Performance Technical Computing teams. He owns AstroArch Consulting, Inc., providing virtualization, security, and network consulting and development. Edward is very active (rated Virtuoso by his peers) on the VMware discussion forums providing answers to security and configuration questions.

Preface

How often have you heard this kind of marketing hype around the use of ESX Server and its compatriots, GSX Server and VMware Workstation?

ESX Server from VMware is hot, Hot, HOT!

The latest version of ESX Server does everything for you!

A Virtualization Utopia!

VMware is the Bomb!

VMware ESX, specifically its latest incarnation, Virtual Infrastructure 3, does offer amazing functionality with virtualization, dynamic resource load balancing, and failover. However, you still need to hire a consultant to come in to share the mysteries of choosing hardware, good candidates for virtualization, choosing installation methods, installing, configuring, using, and even migrating machines. It is time for a reference that goes over all this information in simple language and detail so that readers with different backgrounds can begin to use this extremely powerful tool.

Therefore, this book explains and comments on VMware ESX Server versions 2.5.x and 3.0. I have endeavored to put together a "soup to nuts" description of the best practices for ESX Server that can also be applied in general to the other tools available in the Virtual Infrastructure family inside and outside of VMware. To this end, I use real-world examples wherever possible and do not limit the discussions to just those products developed by VMware, but instead expand the discussion to virtualization tools developed by Vizioncore, Hewlett-Packard (HP), and other third parties. Given that I worked for HP, I use HP hardware and tools to show the functionality we are discussing, yet everything herein translates to other hardware

just as easily. Although things are named differently between HP and their competitors, the functionality of the hardware and hardware tools is roughly the same. I have endeavored to present all the methods available to achieve best practices, including the use of graphical and command-line tools.

As you read, keep in mind the big picture that virtualization provides: better utilization of hardware and resource sharing. In many ways, virtualization takes us back to the days of yore when developers had to do more with a lot less than we have available now. Remember the Commodore 64 and its predecessors, where we thought 64KB of memory was huge? Now we are back in a realm where we have to make do with fewer resources than perhaps desired. By keeping the big picture in mind, we can make the necessary choices that create a strong and viable virtual environment. Because we are doing more with less, this thought must be in the back of our mind as we move forward and helps to explain many of the concerns raised within this tome.

As you will discover, I believe there is quite a bit of knowledge to acquire and numerous decisions to make before you even insert a CD-ROM to begin the installation. How these questions are answered will guide the installation, as you need to first understand the capabilities and limitations of the ESX environment, and the application mix to be placed in the environment. Keeping in mind the big picture and your application mix is a good idea as you read through each chapter of this book.

Who Should Read this Book?

This book delves into many aspects of virtualization and is designed for the beginning administrator as well as the advanced administrator.

How Is this Book Organized?

Here is, in brief, a listing of what each chapter brings to the table.

Chapter 1: System Considerations

By endeavoring to bring you "soup to nuts" coverage, we start at the beginning of all projects: the requirements. These requirements will quickly move into discussions of hardware and capabilities of hardware required by ESX Server, as is often the case when I talk to customers. This section is critical, because understanding your hardware limitations and capabilities will point you to a direction that you

can take to design your virtual datacenter and infrastructure. As a simple example, picture the idea of whether you will need to run 23 servers on a set of blades. Understanding hardware capabilities will let you pick and choose the appropriate blades for your use and how many blades should make up the set. In addition, understanding your storage and virtual machine (VM) requirements can lead you down different paths for management, configuration, and installation. Checklists that lead to each chapter come out of this discussion. In particular, look for discussions on cache capabilities, the best practice for networking, mutual exclusiveness when dealing with storage area networks (SANs), hardware requirements for backup and disaster recovery, and a checklist when comparing hardware. This chapter is a good place to start when you need to find out where else in the book to go look for coverage of an issue.

Chapter 2: Version Comparison

Before we launch down the installation paths and further discussion, best practices, and explorations into ESX, it is time to take time out and discuss the differences between ESX version 3.x.x and ESX version 2.5.x. This chapter opens with a broad stroke of the brush and clearly states that they *are* different. Okay, everyone knows that, but the chapter then delves into the major and minor differences that are highlighted in further chapters of the book. This chapter creates another guide to the book similar to the hardware guide that will lead you down different paths as you review the differences. The chapter covers installation differences, VM differences, and management differences. Once these are clearly laid out and explained, the details are left to the individual chapters that follow. Why is this not before the hardware chapter? Because hardware may change, but the software running on it definitely has with ESX 3, so this chapter treats the hardware as relatively static when compared to the major differences between ESX version 3 and ESX version 2.

Chapter 3: Installation

After delving into hardware considerations and ESX version differences, we head down the installation path, but before this happens, there is another checklist that helps us to best plan the installation. Just doing an install will get ESX running for perhaps a test environment, but the best practices will fall out from planning your installation. You would not take off in a plane without running down the preflight checklist. ESX server is very similar, and it is easy to get into trouble. As an example, we had one customer that decided on an installation without first understand-

ing the functionality required for clustering VMs together. This need to cluster the machines led to a major change and resulted in the reinstallation of all ESX servers in many different locations. A little planning would have alleviated all the rework. The goal is to make the readers aware of these gotchas before they bite. After a review of planning, the chapter moves on to various installations and discusses where paths diverge and why they would. As an example, installing boot from SAN is quite a bit different from a simple installation, at least in the setup, and due to this there is a discussion of the setup of the hardware prior to installation for each installation path. When the installations are completed, there is post-configuration and special considerations when using different SANs or multiple SANs. Limitations on VMFS with respect to sizing a LUN, spanning a LUN, and even the choice of a standard disk size could be a pretty major concern. This chapter even delves into the vendor and Linux software that could be added after ESX Server is fully installed and why you would or would not want to do add it. Also this chapter suggests noting the divergent paths so that you can better install and configure ESX Server. When it comes to the additional software, this chapter leads you to other chapters that discuss the usage details in depth. And last, this chapter covers some of the aspects of automated deployment of ESX Servers and the tools needed to accomplish this task.

Chapter 4: Auditing, Monitoring, and Securing

Because the preceding chapter discussed additional software, it is now time to discuss even more software to install that aids in the auditing, monitoring, and securing of ESX server. This chapter approaches ESX from the perspective of security, and out of that will come better tools for monitoring and auditing your server for failures and possible issues. There is nothing like having to read through several thousands of lines of errors just to determine a problem started. Using good monitoring tools will simplify this task and even enable better software support. That is indeed a bonus! Yet knowing when a problem occurred is only part of monitoring and auditing; you also need to know who did the deed and where they did it, and hopefully why. This leads to auditing. More and more government intervention (Sarbanes-Oxley) requires better auditing of what is happening and even when. This chapter launches into automating this as much as possible. Why would I need to sit and read log files when the simple application can e-mail me when there is a problem? How do I get these tools to page me or even self-repair? I suggest you take special note of how these concepts, tools, and implementations fit with your overall auditing, monitoring, and security requirements. Also, note

how security works inside ESX, because this is an extremely different view of the world, generally distinct from normal systems.

Chapter 5: Storage with ESX

There are many issues dealing with SANs within ESX. There are simple ones from "is my SAN supported" and "why not" to more complex ones such as "will this SAN, switch, Fibre Channel host bus adapter provide the functionality I desire?" Because SANs are generally required to share VMs between ESX Servers, we discuss them in depth. This chapter lets you in on the not-so-good and the good things about each SAN and what the best practices are for use, support, and configuration. Also discussed in this chapter are network-attached storage (NAS) and iSCSI support within ESX. With SANs, there is good, bad, and the downright ugly. For example, if you do not have the proper firmware version on some SANs, things can get downright ugly very quickly! Although the chapter does not discuss the configuration of your SAN for use outside of ESX, it does discuss presentation in general terms and how to get the most out of hardware and, to a certain extent, software multipath capabilities. This chapter suggests you pay close attention to how SAN and NAS interoperate with ESX Server.

Chapter 6: Effects on Operation

Before proceeding to the other aspects of ESX, including the creation of a VM, it is important to review some operational constraints associated with the management of ESX and the running of VMs. Operation issues directly affect VMs. These issues are as basic as maintaining lists of IPs and netmasks, when to schedule services to run through the complexities imposed when using remote storage devices, and its impact on how and when certain virtualization tasks can take place.

Chapter 7: Networking

This chapter discusses the networking possibilities within ESX Server and the requirements placed upon the external environment if any. A good example is mentioned under the hardware discussion, where we discuss hardware redundancy with respect to networking. In ESX Server terms, this discussion is all about network interface card (NIC) teaming, or in more general terms, the bonding of multiple NICs into one bigger pipe for the purpose of increasing bandwidth and failover. However, the checklist is not limited to just the hardware but also

includes the application of best practices for the creation of various virtual switches (vSwitches) within ESX Server, what network interfaces are virtualized, and when to use one over the other, as well as any network lag considerations. It also includes the pitfalls related to debugging problems related to the network. The flexibility of networking inside ESX server implies that the system and network administrators also have to be flexible, as the best practices dictated by a network switch company may lead to major performance problems when applied to ESX Server. Out of this chapter comes a list of changes that may need to be applied to the networking infrastructure, with the necessary data to back up these practices so that discussions with network administrators do not lead toward one-sided conversations. Using real-world examples, this chapter runs through a series of procedures that can be applied to common problems in setting up networking within ESX Server.

Chapters 8 and 9: Configuring ESX from a Host Connection, and Configuring ESX from a Virtual Center or Host

These chapters tie it all together; we have installed, configured, and attached storage to our ESX Server. Now what? Well, we need to manage our ESX Server. There are three primary ways to manage an ESX Server: the use of the management user interface (MUI), which is a web-based client; the use of Virtual Center (VC), which is a .NET client; and the use of the command-line interface (CLI); as well as variations on these provided by HP and third parties. These chapters delve into configuration and use of these interfaces. Out of these chapters will come tools that can be used as part of a scripted installation of an ESX server, as mentioned in Chapter 3.

Chapter 10: Virtual Machines

This chapter goes into the usage of the management interfaces as I address real-world examples of planning installations. This chapter discusses making and storing images of your installation media, where to place VM configuration files, choosing your installation size, and how dynamic disks, logical volume manager, and minimized installs affect that size. Also, this chapter launches into a discussion of the various swap files available to ESX and when each is used and why. In essence, the chapter discusses everything you need to know before you start installing VMs. Once that is discussed, it is possible to launch into installation of VMs using all the standard interfaces. We install Windows, Linux, and NetWare VMs, pointing out where things diverge on the creation of a VM and what has to

be done post install. This chapter looks at specific solutions to VM problems posed to us by customers: the use of eDirectory, private labs, firewalls, clusters, growing Virtual Machine File Systems (VMFSs), and other customer issues. This chapter is an opportunity to see how VMs are created and how VMs differ from one another and why. Also, the solutions shown are those from real-world customers, and they should guide you down your installation paths.

Chapter 11: Dynamic Resource Load Balancing

Because monitoring is so important, it is covered once more with an eye toward dynamic resource load balancing (DRLB) and utilization goals. The chapter discusses the use of various performance-monitoring tools that will add to your understanding of how to balance resources across multiple ESX Servers. Whereas some tools perform DRLB, still others report the important aspects for the administrator to apply changes by hand. With the advent of DRLB, there needs to be a clear understanding of what is looked at by such tools. This chapter gives you that understanding by reviewing hardware-utilization data, usage data, and performance data presented by various tools. Then, after this data is understood, the chapter shows you the best practices for the application of the data using VMotion, ESX Server clustering techniques, and how to apply alarms to various monitoring tools to give you a heads up when something needs to happen either by hand or has happened dynamically. I suggest paying close attention to the makeup of DLRB to understand the limitations of all the tools.

Chapter 12: Disaster Recovery and Backup

A subset of DLRB can apply to disaster recovery (DR). DR is a huge subject, so it is limited to just the ESX Server and its environment that lends itself well to redundancy and in so doing aids in DR planning. But, before you plan, you need to understand the limitations of the technology and tools. DR planning on ESX is not more difficult than a plan for a single physical machine. The use of a VM actually makes things easier if the VM is set up properly. A key component of DR is the making of safe, secure, and proper backups of the VMs and system. What to backup and when is a critical concern that fits into your current backup directives, which may not apply directly to ESX Server and which could be made faster. The chapter presents several real-world examples around backup and DR, including the use of redundant systems, how this is affected by ESX and VM clusters, the use of locally attached tape, the use of network storage, and some helpful scripts to make it all work. In addition, this chapter discusses some third-party tools avail-

able to make your backup and restoration tasks simpler. The key to DR is a good plan, and the checklist in this chapter will aid in developing a plan that encompasses ESX Server and can be applied to all the Virtual Infrastructure products. Some solutions require more hardware (spare disks, perhaps other SANS), more software (Vizioncore's ESXRanger, Power Management, and so on), and almost all of them require time to complete.

Epilogue: The Future of Virtualization

After all this, the book concludes with a discussion of the future of ESX Server.

Appendixes

Appendix A, "Security Scripts," presents a shell script that can be used to increase the security of an ESX Server. Appendix B, "ESX Version 3 Text Installation," presents the ESX text installation and Appendix C, "ESX Version 3 Graphical Installation," presents the ESX installation through the graphical interface.

References

This element suggests possible further reading.

Reading...

Please sit down in your favorite comfy chair, with a cup of your favorite hot drink, and prepare to enjoy the chapters in this book. Read it from cover to cover, or use as it a reference. The best practices of ESX Server sprinkled throughout the book will entice and enlighten, and spark further conversation and possibly well-considered changes to your current environments.

Chapter 1

System Considerations

The design and architecture of a VMware ESX Server environment depends on a few different considerations, ranging from the types of applications and operating systems to virtualize, to how many physical machines are desired to virtualize, to upon what hardware to place the virtual environments. Quite quickly, any discussion about the virtual infrastructure soon evolves to a discussion of the hardware to use in the environment. Experience shows that, before designing a virtual datacenter, it's important to understand what makes a good virtual machine host and the limitations of current hardware platforms. In this chapter, customer examples illustrate various architectures based on limitations and desired results. These examples are not exhaustive, just a good introduction to understand the impact of various hardware choices on the design of the virtual infrastructure. An understanding of potential hardware use will increase the chance of virtualization success. The architecture potentially derived from this understanding will benefit not just a single ESX Server, but also the tens or hundred that may be deployed throughout a single or multiple datacenters. Therefore, the goal here is to develop a basis for enterprisewide ESX Server deployment. The first step is to understand the hardware involved.

As an example, a customer wanted a 20:1 compression ratio for virtualization of their low-utilization machines. However, they also had networking goals to compress their network requirements at the same time. The other limiting factor was the hardware they could choose, because they were limited to a certain set, with the adapters precisely limited. The specifications stated that with the hardware they could do what they wanted to do, so they proceeded down that path. However, what the hardware specification states is not necessarily the best practice for ESX, and this led to quite a bit of hardship as they worked through the issues with their chosen environment. They could have alleviated certain hardships early on with a better understanding of the impact of ESX on the various pieces of

hardware and that hardware's impact on ESX. (Whereas most, if not all, of the diagrams and notes use Hewlett-Packard hardware, these are just examples; similar hardware is available from Dell, IBM, Sun, and many other vendors.)

Basic Hardware Considerations

An understanding of basic hardware aspects and their impact on ESX can greatly increase your chances of virtualization success. To begin, let's look at the components that make up modern systems.

When designing for the enterprise, one of the key considerations is the processor to use, specifically the type, cache available, and memory configurations; all these factors affect how ESX works in major ways. The wrong choices may make the system seem sluggish and will reduce the number of virtual machines (VMs) that can run, so it is best to pay close attention to the processor and system architecture when designing the virtual environment.

Before picking any hardware, always refer to the VMware Hardware Compatibility Lists (HCLs), which you can find as four volumes at www.vmware.com/support/pubs/vi_pubs.html:

- ESX Server 3.x Systems Compatibility Guide
- ESX Server 3.x I/O Compatibility Guide
- ESX Server 3.x Storage/SAN Compatibility Guide
- ESX Server 3.x Backup Software Compatibility Guide

Processor Considerations

Processor family, which is not a huge consideration in the scheme of things, is a consideration when picking multiple machines for the enterprise because the different types of processor architectures impact the availability of ESX features. Specifically, mismatched processor types will prevent the use of VMotion. VMotion allows for the movement of a running VM from host to host by using a specialized network connection. VMotion momentarily freezes a VM while it copies the memory and register footprint of the VM from host to host. Afterward, the VM on the old host is shut down cleanly, and the new one will start. If everything works appropriately, the VM does not notice anything but a slight hiccup that can be absorbed with no issues. However, because VMotion copies the register and memory footprint from host to host, the processor architecture and chipset in use needs to match. It is not possible without proper masking of processor features to

VMotion from a Xeon to an AMD processor or from a single-core processor to a dual-core processor, even if it is the same family of processor that was introduced in ESX version 2.5.2. If the Virtual Machine to be moved is a 64 bit VM, then the processors must match exactly as there is no method available to mask processor features. Therefore the processor architecture and chipset (or the instruction set) is extremely important, and because this can change from generation to generation of the machines, it is best to introduce two machines into the virtual enterprise at the same time to ensure VMotion actually works. When introducing new hardware into the mix of ESX hosts, test to confirm that VMotion will work.

Best Practice

Standardize on a single processor and chipset architecture. If this is not possible because of the age of existing machines, test to ensure VMotion still works, or introduce hosts in pairs to guarantee successful VMotion. Different firmware revisions can also affect VMotion functionality.

Ensure that all the processor speed or stepping parameters in a system match, too.

Note that many companies support mismatched processor speeds or stepping in a system. ESX would really rather have all the processors at the same speed and stepping. In the case where the stepping for a processor is different, each vendor provides different instructions for processor placement. For example, Hewlett-Packard (HP) will require that the slowest processor be in the first processor slot and all the others in any remaining slots. To alleviate any type of issue, it is a best practice that the processor speeds or stepping match within the system.

Before proceeding to the next phase, a brief comment on dual-core (DC) versus single-core (SC) processors is warranted. ESX Server does not differentiate in its licensing scheme between DC and SC processors, so the difference between them becomes a matter of cost versus performance gain of the processors. The DC processor will handle more VMs than an SC but also cost more and has support only in the later releases of ESX. In some cases, it is possible to start with SC processors and make the first upgrade of the ESX Servers to be DC processors in their effort to protect the hardware investment. If performance is the issue, DC is the way to go. Nevertheless, for now, the choice is a balance of cost versus performance. Due to current shared-cached mechanisms for DC, an eight-core or four-processor server has the same processing power as if there were seven

physical processors, and once shared cache goes away there is a good chance the efficiency of the DC will match that of a true eight-way machine.

Cache Considerations

Unlike matching processor architectures and chipsets, it is not important to match the L2 Cache between multiple hosts. A mismatch will not prevent VMotion from working. However, L2 Cache is most likely to be more important when it comes to performance because it controls how often main memory is accessed. The larger the L2 Cache, the better an ESX Server will run. Consider Figure 1.1 in terms of VMs being a complete process and the access path of memory. Although ESX tries to limit memory usage as much as possible, with 40 VMs this is just not possible, so the L2 Cache plays a significant part in how VMs perform.

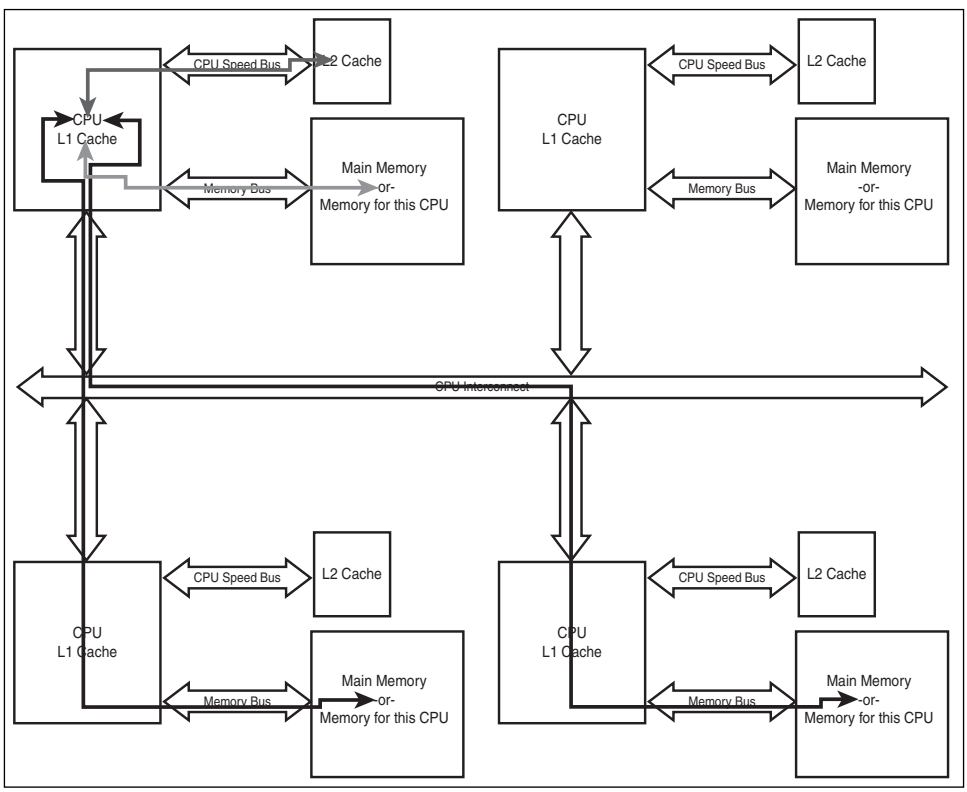

Figure 1.1 Memory access paths

As more VMs are added to a host of the same operating system (OS) type and version, ESX will start to share code segments between VMs. Code segments are the instructions that make up the OS within the VM and *not* the data segments that contain the VM's memory. Code-segment sharing between VMs does not violate any VM's security, because the code never changes, and if it does, code-segment sharing is no longer available for the VMs. That aside, let's look at Figure 1.1 again. When a processor needs to ask the system for memory, it first goes to the L1 Cache (up to a megabyte usually) and sees whether the memory region requested is already on the processor die. This action is extremely fast, and although different for most processors, we can assume it is an instruction or two (measured in nanoseconds). However, if the memory region is not in the L1 Cache, the next step is to go to the L2 Cache, which is generally off the die, over an extremely fast channel (green arrow) usually running at processor speeds. However, this takes even more time and instructions than L1 Cache access, and adds to the overall time to access memory. If the memory region you desire is not in L2 Cache, it is in main memory (yellow arrow) somewhere, and must be accessed and loaded into L2 Cache so that the processor can access the memory, which takes another order of magnitude of time to access. Usually, a cache line is copied from main memory, which is the desired memory region and some of the adjacent data, to speed up future memory access. When we are dealing with non-uniform memory access (NUMA) architecture, because is the case with AMD processors, there is yet another step to memory access if the memory necessary is sitting on a processor board elsewhere in the system. The farther away it is, the slower the access time (red and black arrows), and this access over the CPU interconnect will add another order of magnitude to the memory access time, which in processor time can be rather slow.

Okay, but what does this mean in real times? Assuming that we are using a 3.06GHz processor, the times could be as follows:

- L1 Cache, one cycle (~0.33ns)
- L2 Cache, two cycles, the first one to get a cache miss from L1 Cache and another to access L2 Cache (~0.66ns), which runs at CPU speeds (green arrow)
- Main memory is running at 333MHz, which is an order of magnitude slower than L2 Cache (~3.0ns access time) (yellow arrow)
- Access to main memory on another processor board (NUMA) is an order of magnitude slower than accessing main memory on the same processor board (~30–45ns access time, depending on distance) (red or black arrow)

This implies that large L2 Cache sizes will benefit the system more than small L2 Cache sizes; so, the larger the better, so that the processor has access to larger chunks of contiguous memory, because the memory to be swapped in will be on the larger size and this will benefit the performance of the VMs. This discussion does not state that NUMA-based architectures are inherently slower than regular-style architectures, because most NUMA-based architectures running ESX Server do not need to go out to other processor boards very often to gain access to their memory.

Best Practice
Invest in the largest amount of L2 Cache available for your chosen architecture.

Memory Considerations

After L2 Cache comes the speed of the memory, as the preceding bulleted list suggests. Higher-speed memory is suggested, and lots of it! The quantity of memory and the number of processors govern how many VMs can run simultaneously without overcommitting this vital resource. In many cases, the highest-speed memory often comes with a lower memory penalty. An example of this is the HP DL585, which can host 32GB of the highest-speed memory, yet it can host 64GB of the lower-speed memory. So, obviously, there are trade-offs in the number of VMs and how you populate memory, but generally the best practice is high-speed and a high quantity. Consider that the maximum number of vCPUs per core is eight. On a 4-processor box, that could be 32 VMs. If each of these VMs is 1GB, we need 33GB of memory to run the VMs. Why 33GB? Because 33GB gives both the console OS (COS, the service console) and the VMkernel up to 1GB of memory to run the VMs. Because 33GB of memory is a weird number for most computers these days, we would need to overcommit memory. When we start overcommitting memory in this way, the performance of ESX can degrade. In this case, it might be better to move to 64GB of memory instead. However, that same box with DC processors can, theoretically, run up to 64 VMs, which implies that we take the VM load to the logical conclusion, and we are once more overcommitting memory. However, eight VMs per processor is a theoretical limit, and it's hard to achieve. (It is not possible to run VMs with more vCPUs than available physical cores, but there is still a theoretical limit of eight vCPUs per core.) There are rumors that it has been done. Unfortunately, that pushes the machine to its limits and is not recommended. Recommended memory utilization differs significantly for each configuration.

Best Practice
High-speed memory and lots of it! However, be aware of the possible trade-offs involved in choosing the highest-speed memory. More VMs may necessitate the use of slightly slower memory.

What is the recommended memory configuration? This subject is covered when we cover VMs in detail, because it really pertains to this question; but, the strong recommendation is to put in the maximum memory the hardware will support that is not above the 64GB limit set by ESX (because overcommitting memory creates too much of a performance hit and should only be done in extreme circumstances). However, this is a pretty major cost-benefit solution because redundancy needs to be considered with any implementation of ESX; it is therefore beneficial to cut down on the per-machine memory requirements to afford redundant systems.

I/O Card Considerations

The next consideration is which I/O cards are supported. Unlike other operating systems, there is a finite list of supported I/O cards. There are limitations on the redundant array of inexpensive drives (RAID) arrays, Small Computer System Interface (SCSI) adapters for external devices including tape libraries, network interface cards (NICs), and Fibre Channel host bus adapters. Although the list changes frequently, it boils down to a few types of supported devices limited by the set of device drivers that are a part of ESX. Table 1.1 covers the devices and the associated drivers.

Table 1.1

Devices and Drivers

Device Type	Device Driver Vendor	Device Driver Name	Notes
Network	Broadcom	bcm5700	
	Broadcom	bcm5721	
	Intel	e1000	Quad-port MT is supported on ESX >= 2.5.2
	Intel	e100	
	Nvidia	forcedeth	ESX >= 3.0.2 only
	3Com	3c90x	ESX <= 2.5.x only
	AceNIC	Acenic	ESX <= 2.5.x only

continues...

Table 1.1 continued

Devices and Drivers			
Device Type	**Device Driver Vendor**	**Device Driver Name**	**Notes**
Fibre Channel	Emulex	Lpfcdd	Dual/single ports
	Qlogic	qla2x00	Dual/single ports
SCSI	Adaptec	aic7xxx	Supported for external devices
	Adaptec	aic79xx	Supported for external devices
	Adaptec	adp94xx	Supported for external devices
	LSI Logic	ncr53c8xx	ESX <= 2.5.x only
	LSI Logic	sym53c8xx	ESX <= 2.5.x only
	LSI Logic	mptscsi	
RAID array	Adaptec	dpt_i2o	ESX <= 2.5.x only
	HP	cpqarray	External SCSI is for disk arrays only. ESX <= 2.5.x only
	HP	cciss	External SCSI for disk arrays only
	Dell	aacraid	
	Dell	megaraid	
	IBM/Adaptec	ips	
	IBM/Adaptec	aacraid	
	Intel	gdth	ESX <= v2.5.x only
	LSI	megaraid	
	Mylex	DAC960	
iSCSI	Qlogic 4010	qla4010	ESX v3 only

If the driver in question supports a device, in most cases it will work in ESX. However, if the device requires a modern device driver, do not expect it to be part of ESX, because ESX by its very nature does not support the most current devices. ESX is designed to be stable, and that often precludes modern devices. For example, Serial Advanced Technology Attachment (SATA) devices are not a part of ESX version 2.5, yet are a part of ESX version 3.5 (soon to be available). Another missing device that is commonly requested is the TCP Offload Engine NIC (TOE cards), and the jury is still out on the benefit given the network sharing design of ESX. As noted in the table, various SCSI adapters have limitations. A key limitation is that an Adaptec card is required for external tape drives or libraries and that any other type of card is usable with external disk arrays.

Best Practice Regarding I/O Cards
If the card you desire to use is *not* on the HCL, do not use it. The HCL is definitive from a support perspective. Although a vendor may produce a card and self-check it, if it is not on the HCL VMware will not support the configuration.

Table 1.1 refers particularly to those devices that the VMkernel can access, and not necessarily the devices that the COS installs for ESX versions earlier than 3.0. There are quite a few devices for which the COS has a driver, but the VMs cannot use them. Two examples of this come to mind, the first are NICs not listed in Table 1.1 but that actually have a COS driver; Kingston or old Digital NICs fall into this category. The second example is the IDE driver. It is possible to install the COS onto an Intelligent Drive Electronics (IDE) drive for versions of ESX earlier than version 3, or SATA/IDE drives for ESX version 3. However, these devices cannot host a Virtual Machine File System (VMFS), so a storage area network (SAN) or external storage is necessary to hold the VM disk files and any VMkernel swap files for each VM.

For ESX to run, it needs at a minimum two NICs (yes, it is possible to use one NIC, but this is never a recommendation for production servers) and one SCSI storage device. One NIC is for the service console and the other for the VMs. Although it is possible to share these so that only one NIC is required, VMware does not recommend this except in extreme cases (and it leads to possible performance and security issues). The best practice for ESX is to provide redundancy for everything so that all your VMs stay running even if network or a Fibre Channel path is lost. To do this, there needs to be some considerations around network and Fibre configurations and perhaps more I/O devices. The minimum best practice for network card configuration is four ports, the first for the SC, the second and third teamed together for the VMs (to provide redundancy), and the fourth for VMotion via the VMkernel interface on its own private network. For full redundancy and performance, six NIC ports are recommended with the extra NICs being assigned to the service console and VMotion. If another network is available to the VMs, either use 802.1q virtual LAN (VLAN) tagging or add a pair of NIC ports for redundancy. Add in a pair of Fibre Channel adapters and you gain failover for your SAN fabric. If there is a need for a tape library, pick an Adaptec SCSI adapter to gain access to this all-important backup device.

Best Practice

Four NIC ports for performance, security, and redundancy and two Fibre Channel ports for redundancy are the best practice for ESX versions earlier than version 3. For ESX version 3, six NIC ports are recommended for performance, security, and redundancy.

If adding more networks for use by the VMs, either use 802.1q VLAN tagging to run over the existing pair of NICs associated with the VMs or add a new pair of NICs for the VMs.

When using iSCSI with ESX version 3, add another NIC port to the service console for performance, security, and redundancy.

When using Network File System (NFS) via network-attached storage (NAS) with ESX version 3, add another pair of NIC ports to give performance and redundancy.

If you are using locally attached tape drives or libraries, use an Adaptec SCSI adapter. No other adapter will work properly. However, the best practice for tape drives or libraries is to use a remote archive server.

For ESX version 3, iSCSI and NAS support is available, and this differs distinctly from the method by which it is set up for ESX version 2.5.x and earlier. iSCSI and NFS-based NAS are accessed using their own network connection assigned to the VMkernel similar to the way VMotion works or how a standard VMFS-3 is accessed via Fibre. Although NAS and iSCSI access can share bandwidth with other networks, keeping them separate could be better for performance. The iSCSI VMkernel device must share the subnet as the COS for authentication reasons, regardless of whether Challenge Handshake Authentication Protocol (CHAP) is enabled, although an NFS-based NAS would be on its own network. Before ESX version 3, an NFS-based NAS was available only via the COS, and iSCSI was not available when those earlier versions were released. Chapter 8, "Configuring ESX from a Host Connection," discusses this new networking possibility in detail.

Disk Drive Space Considerations

The next item to discuss is what is required for drive space. In essence, the disk subsystem assigned to the system needs to be big enough to contain the COS and

ESX. The swap file for the COS, storage space for the virtual swap file (used to overcommit memory in ESX), VM disk files, local ISO images, and backups of the Virtual Machine Disk Format (VMDK) files for disaster-recovery reasons. If Fibre Channel or iSCSI is available, it is obvious that you should offload the VM disk files to these systems. When we are booting from a SAN we have to share the Fibre Channel adapter between the service console and ESX for ESX earlier than version 3.0. The sharing of the Fibre Channel adapter ports is not a best practice and is offered as a matter of convenience and not really suggested for use. (Boot from a SAN is covered fully in Chapter 3, "Installation"). Putting temporary storage (COS swap) onto expensive SAN or iSCSI storage is also not a best practice; the recommendation is that there be some form of local disk space to host the OS and the COS swap files. It is a requirement for VMotion in ESX version 3 that the per-VM VMkernel swap live on the remote storage device. The general recommendation is roughly 72GB in a RAID 1 or mirrored configuration for the operating system and its necessary file systems, and for local storage of ISO files and other items as necessary.

For ESX versions earlier than version 3, the VMkernel swap file space should be twice the amount of memory in the machine. However, if twice the amount of memory in the machine is greater than 64GB, another VMkernel swap file should be used. Each VMkernel swap file should live on its own VMFS. VMs could live on a VMFS created larger than 64GB, and then a few VMs could live with the virtual swap files. However, if there will be no VMs on these VMFS partitions, the partitions could be exactly 64GB and use RAID 0 or unprotected RAID storage. The caveat in this case is if you lose a drive for this RAID device, it's possible the ESX Server will no longer be able to overcommit memory and those VMs currently overcommitted will fail. Use the fastest RAID level and place the virtual swap file on a VMFS on its own RAID set. It is also possible to place the VMkernel swap with the operating system on the recommended RAID 1 device. RAID 5 is really a waste for the VMkernel swap. RAID 1 or the VMFS partition containing the VMkernel swap file for ESX versions earlier than version 3 is the best choice.

For ESX version 3, there is no need to have a single VMkernel swap file. These are now included independently with each VM.

Any VMFS that contains VMs should use a RAID 5 configuration for the best protection of data. Chapter 12, "Disaster Recovery and Backup," covers the disk configuration in much more detail as it investigates the needs of the local disk from a disaster-recovery (DR) point of view. The general DR point of view is to have enough local space to run critical VMs from the host without the need for a SAN or iSCSI device.

Best Practice for Disk

Have as much local disk as possible to hold VMkernel swap files (twice memory for low-memory systems and equal to memory for the larger-memory systems) for ESX versions earlier than version 3.

Have as much local disk necessary to hold the OS, local ISO images, local backups of critical VMs, and perhaps some local VMs.

Basic Hardware Considerations Summary

Table 1.2 conveniently summarizes the hardware considerations discussed in this section.

Table 1.2

Best Practices for Hardware

Item	ESX Version 3	ESX Versions Earlier Than Version 3	Chapter to Visit for More Information
Fibre Ports	Two 2GB	Two 2GB	Chapter 5
Network Ports	Six 1GB Two for COS Two for VMs Two for VMotion	Four 1GB One for COS Two for VMs One for VMotion	Chapter 8
Local disks	SCSI RAID Enough to keep a copy of the most important VMs	SCSI RAID Enough to keep a copy of the most important VMs and local vSwap file	
iSCSI	Two 1GB network ports via VMkernel or iSCSI HBA	N/A	Chapter 8
SAN	Enterprise class	Enterprise class	Chapter 5
Tape	Remote	Remote	Chapter 11
NFS-based NAS	Two 1GB network ports via VMkernel	Via COS	Chapter 8
Memory	Up to 64GB	Up to 64GB	
Networks	Three or four Admin/iSCSI network VM network VMotion network VMkernel network	Three Admin network VM network VMotion network	Chapter 8

Specific Hardware Considerations

Now we need to look at the hardware currently available and decide how to best use it to meet the best practices listed previously. All hardware will have some issues to consider, and applying the comments from the first section of this chapter will help show the good, bad, and ugly about the possible hardware currently used as a virtual infrastructure node. The primary goal is to help the reader understand the necessary design choices when choosing various forms of hardware for an enterprise-level ESX Server farm. Note that the number of VMs mentioned are based on an average machine that does not do very much network, disk, or other I/O and has average processor utilization. This number varies too much based on the utilization of the current infrastructure, and these numbers are a measure of what each server is capable of and are not intended as maximums or minimums. A proper analysis will yield the best use of your ESX Servers and is part of the design for any virtual infrastructure.

Blade Server Systems

Because blade systems (see Figure 1.2) virtualize hardware, it is a logical choice for ESX, which further virtualizes a blade investment by running more servers on each blade. However, there are some serious design considerations when choosing blades. The majority of these considerations are in the realm of port density and availability of storage. Keep in mind our desire to have at least four NICs, two Fibre Channel ports, and local disk: Many blades do not have these basic requirements. Take, for example, the IBM HS20. This blade has two on-board NICs and two Fibre Channel ports. Although there is plenty of Fibre Channel, there is a dearth of NICs in this configuration. That is not to say that the HS20 is not used, but the trade-off in its use is either lack of redundancy, or security, and performance trade-offs. Other blades have similar trade-offs, too. Another example is the HP BL3 p blade. Although it has enough NIC ports, the two Fibre Channel ports share the same port on the fabric, which in essence removes Fibre redundancy from the picture. On top of that restriction, the BL3 p uses an IDE/ATA drive and not a SCSI drive, which implies that a SAN or iSCSI server is also required to run VMs. There are also no Peripheral Component Interconnect (PCI) slots in most blades, which makes it impossible to add in an additional NIC, Fibre, or SCSI adapter. In addition to the possible redundancy issue, there is a limitation on the amount of memory that you can put into a blade. With a blade, there is no PCI card redundancy because all NIC and Fibre ports are part of the system or some form of dual-port mezzanine card. If more than one network will be available to

the VMs, 802.1q VLAN tagging would be the recommendation, because there is no way to add more NIC ports and splitting the NIC team for the VMs would remove redundancy. Even with these trade-offs, blades make very nice commonly used ESX Servers. It is common for two processor blades to run between four and ten VMs. This limitation depends on the amount of memory available. On four-processor blades, where you can add quite bit more memory, the loads can approach those of comparable nonblade systems.

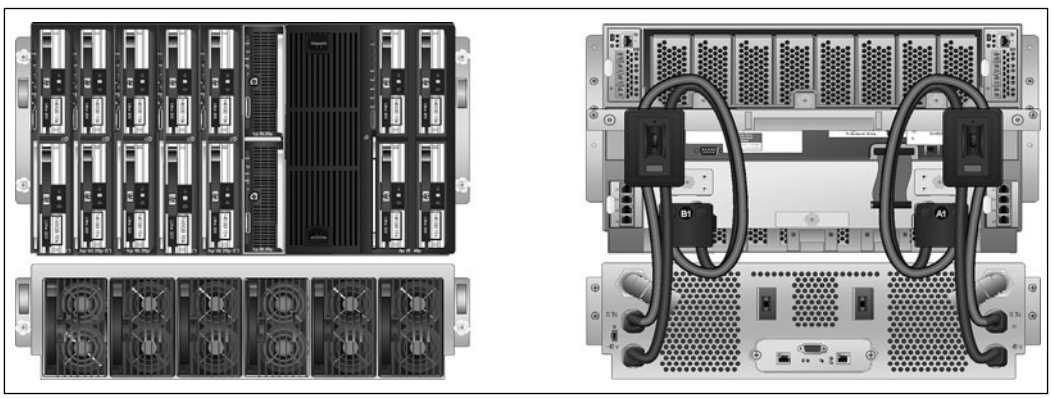

Figure 1.2 Front and back of blade enclosure
Visio templates for image courtesy of Hewlett-Packard.

Best Practice with Blades
Pick blades that offer full NIC and Fibre redundancy.

1U Server Systems

The next device of interest is the 1U server (see Figure 1.3), which offers in most cases two on-board NICs, generally no on-board Fibre, perhaps two PCI slots, and perhaps two to four SCSI/SAS disks. This is perfect for adding a quad-port NIC and a dual-port Fibre controller; but if you need a SCSI card for a local tape device, which is sometimes necessary but never recommended, there is no chance to put one in unless there is a way to get more on-board NIC or Fibre ports. In addition to the need to add more hardware into these units, there is a chance that PCI card redundancy would be lost, too. Consider the HP DL360 as a possible ESX Server, which is a 1U device with two SCSI or SATA drives, two on-board NICs, and possibly a mezzanine Fibre Channel adapter. In this case, if we were using

ESX version 2.5.x or earlier, we would need to only choose SCSI drives, and for any version, we would want to add at least a quad-port NIC card to get to the six NICs that make up the best practice and gain more redundancy for ESX version 3. In some cases, there is a SCSI port on the back of the device, so access to a disk array will increase space dramatically, yet often driver deficiencies affect its usage with tape devices.

Figure 1.3 1U server front and back
Visio templates for image courtesy of Hewlett-Packard.

In the case of SAN redundancy, if there were no mezzanine Fibre Channel adapter, the second PCI slot would host a dual-port Fibre Channel adapter, which would round out and fill all available slots. With the advent of quad-port NIC support, adding an additional pair of NIC ports for another network requires the replacement of the additional dual-port NIC with the new PCI card. There are, once again, a fair number of trade-offs when choosing this platform, and its low quantity of memory implies fewer VMs per server, perhaps in the four to ten range of VMs, depending on the quantity of memory and size of disk in the box. With slightly more capability than blades, the 1U box makes a good backup server, but can be a workhorse when needed.

Best Practice for 1U Boxes
Pick a box that has on-board Fibre Channel adapters so that there are free slots for more network and any other necessary I/O cards. Also, choose large disk drives when possible. There should be at least two on-board network ports. Add quad-port network and dual-port Fibre Channel cards as necessary to get port density.

2U Server Systems

The next server considered is the 2U server (see Figure 1.4), similar to the HP DL380. This type of server usually has two on-board Ethernet ports, perhaps one on-board Fibre Channel port, and usually an external SCSI port for use with external drive arrays. In addition to all this, there are at least three PCI slots, up to six

SCSI drives, and at least twice as much memory than a 1U machine. The extra PCI slot adds quite a bit of functionality, because it either can host an Adaptec SCSI card to support a local tape drive or library, which is sometimes necessary but never recommended, or it can host more network capability. At the bare minimum, at least two more NIC ports are required and perhaps a dual-port Fibre Channel adapter if there is not a pair of ports already in the server. Because this class of server can host six SCSI disks, they can be loaded up with more than 1TB of space, which makes the 2U server an excellent stand-alone ESX Server. Introduce dual-core processors and this box has the power to run many VMs. The major limitation on this class of server is the possible lack of network card space and the memory constraint. Even with these limitations, it is a superb class of server and provides all the necessary components to make an excellent ESX Server.

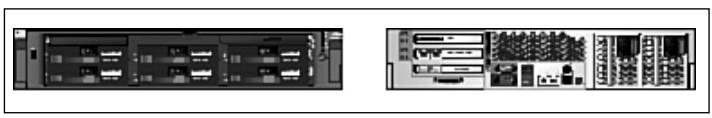

Figure 1.4 Front and back of 2U server
Visio templates for image courtesy of Hewlett-Packard.

Pairing a 2U server with a small tape library to become an office in a box that ships to a remote location does not require a SAN or another form of remote storage because it has plenty of local disk space, to which another disk array connects easily. Nevertheless, the 2U has the same characteristics as a 1U box in many cases. Is the extra memory and PCI slot very important? It can be, and depending on the type of server, there might be a need for a dual or quad-port NIC, dual-port host bus adapter (HBA), and a SCSI adapter for a tape library. The extra slot, extra memory, and lots of local disk make this class of server an extremely good workhorse for ESX. It is possible to run between 6 and 24 VMs on these types of servers depending on available memory and whether DC processors are in use.

Best Practice for 2U Servers
Pick a server that has at least two on-board NIC ports, two on-board Fibre Channel ports, plenty of disk, and as much memory as possible. Add a quad-port network card to gain port density and, if necessary, two single-port Fibre Channel adapters add more redundancy

Large Server-Class Systems

The next discussion combines multiple classes of servers (see Figure 1.5). The class combines the 4, 8, and 16 processor machines. Independent of the processor count, all these servers have many of the same hardware features. Generally, they have four SCSI drives, at least six PCI slots, two on-board NICs, RAID memory, and very large memory footprints ranging from 32GB to 128GB. The RAID memory is just one technology that allows for the replacement of various components while the machine is still running, which can alleviate hardware-based downtime unless it's one of the critical components. RAID memory is extremely nice to have, but it is just a fraction of the total memory in the server and does not count as available memory to the server. For example, it is possible to put a full 80GB of memory into an HP DL760, but the OS will only see 64GB of memory. The missing 16GB becomes the RAID memory pool, which comes into use only if there is a bad memory stick discovered by the hardware. Generally, the larger machines have fewer disks than the 2U servers do, but it makes up for that by having an abundance of PCI buses and slots enabling multiple Fibre Channel adapters and dual-port NICs for the highest level of redundancy. In these servers, the multiple Fibre Channel ports suggested by the general best practice would each be placed on different PCI buses, as would the NIC cards to get better performance and redundancy in PCI cards, SAN fabric, and networking. These types of servers can host a huge number of VMs. The minimum number of VMs is usually in the range of 20, but it can grow to as high as 50 depending on processor count, utilization, and load.

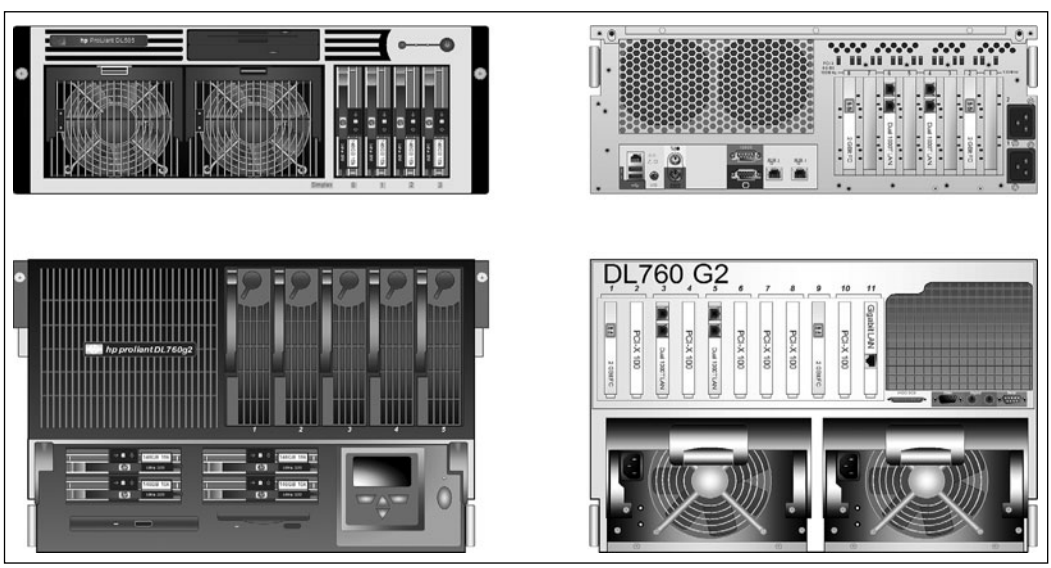

Figure 1.5 Back and front of large server-class machines
Visio templates for image courtesy of Hewlett-Packard.

The Effects of External Storage

There are many different external storage devices, ranging from simple external drives, to disk arrays, shared disk arrays, active/passive SAN, active/active SAN, SCSI tape drives, to libraries, Fibre-attached tape libraries.... The list is endless actually, but we will be looking at the most common devices in use today and those most likely to be used in the future. We shall start with the simplest device and move on to the more complex devices. As we did with servers, this discussion points out the limitations or benefits in the technology so that all the facts are available when starting or modifying virtual infrastructure architecture.

For local disks, it is strongly recommended that you use SCSI/SAS RAID devices; although IDE is supported for running ESX, it does not have the capability to host a VMFS, so some form of external storage will be required. ESX version 3 supports local SATA devices, but they share the same limitations as IDE. In addition, if you are running any form of shared disk cluster, such as Microsoft Cluster servers, a local VMFS is required for the boot drives, yet remote storage is required for all shared volumes using raw disk maps. If one is not available, the shared disk cluster will fail with major locking issues.

Best Practice for Local Disks
Use SCSI or SAS disks.

Outside of local disks, the external disk tray or disk array (see Figure 1.6) is a common attachment and usually does not require more hardware outside of the disk array and the proper SCSI cable. However, like stand-alone servers, the local disk array does not enable the use of VMotion to hot migrate a VM. However, when VMotion is not required, this is a simple way to get more storage attached to a server. If the disk array is a SATA array, it is probably better to go to SCSI instead, because although you can add more space into SATA, SCSI is much faster and is supported on all versions of ESX.

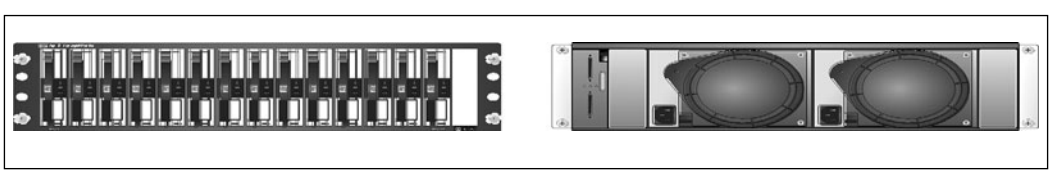

Figure 1.6 Front and back of an external disk array
Visio templates for image courtesy of Hewlett-Packard.

The next type of device is the shared disk array (see Figure 1.7), which has its own controllers and can be attached to a pair of servers instead of only one. The on-board controller allows logical unit numbers (LUNs) to be carved out and to be presented to the appropriate server or shared among the servers. It is possible to use this type of device to share only VMFS-formatted LUNs between at most four ESX hosts because that is generally the limit on how many SCSI interfaces that are available on each shared disk array. It is a very inexpensive way to create multi-machine redundancy. However, using this method limits the cluster of ESX Servers to exactly the number of SCSI ports that are available, and limits the methods for accessing raw LUNs from within VMs.

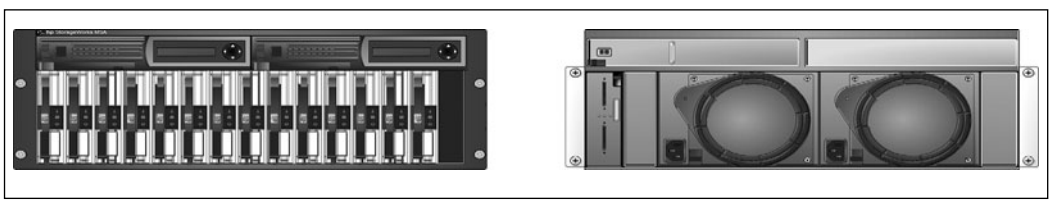

Figure 1.7 Front and back of a shared SCSI array
Visio templates for image courtesy of Hewlett-Packard.

Best Practice for Local Storage
Use local or locally attached SCSI-based storage systems.

A SAN is one of the devices that will allow VMotion to be used and generally comes in an entry-level (see Figure 1.8) and enterprise-level (see Figure 1.9) styles. Each has its uses with ESX and all allow the sharing of data between multiple ESX hosts, which is the prime ingredient for the use of VMotion. SAN information is covered in detail in Chapter 5, "Storage with ESX."

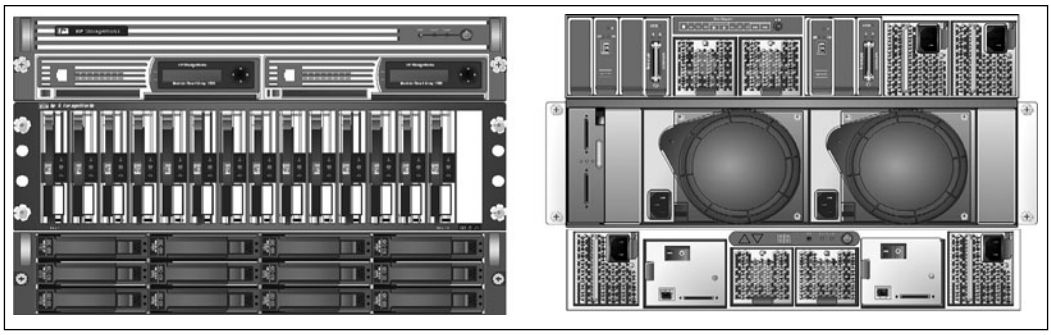

Figure 1.8 Front and back of an entry-level SAN with SATA drives
Visio templates for image courtesy of Hewlett-Packard.

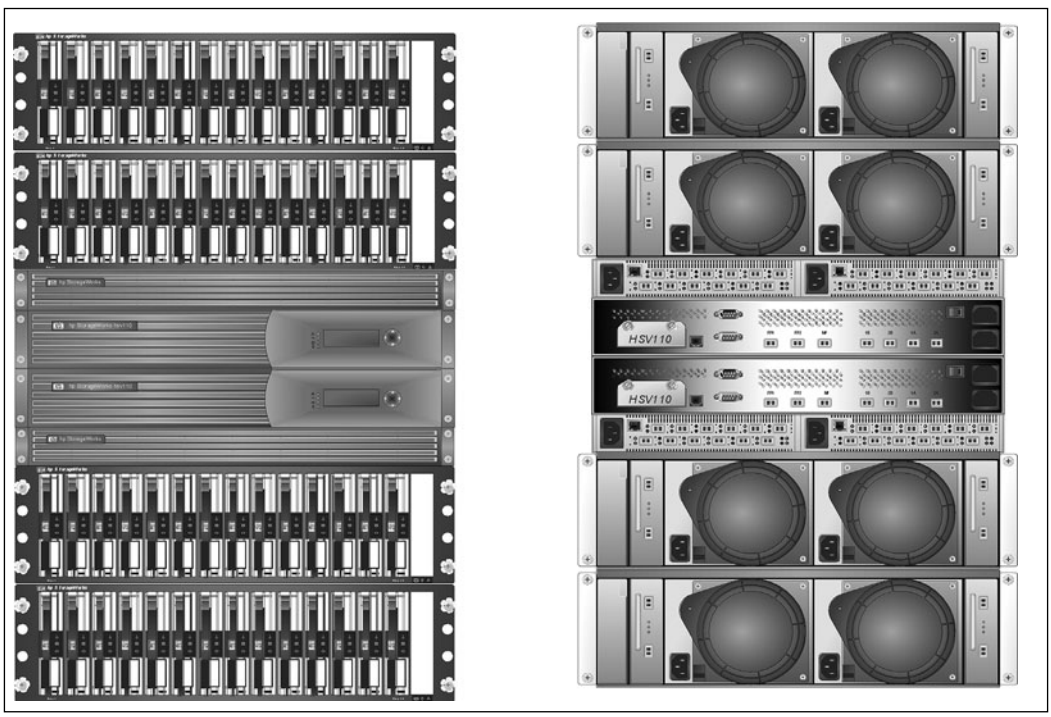

Figure 1.9 Front and back of an enterprise-level SAN
Visio templates for image courtesy of Hewlett-Packard.

Although SATA drives are not supported for ESX earlier than version 3.5, when directly attached to a host unless a SCSI to SATA bridge adapter is in use, they are supported if part of a SAN (refer to Figure 1.8). However, they are slower than using SCSI drives, so they may not be a good choice for primary VMDK

storage, but would make a good temporary backup location; the best solution is to avoid non-SCSI drives as much as possible. Although the entry-level SAN is very good for small installations, enterprise-class installations really require an enterprise-level SAN (refer to Figure 1.9). The enterprise-level SAN provides a higher degree of redundancy, storage, and flexibility for ESX than an entry-level version. Both have their place in possible architectures. For example, if you are deploying ESX to a small office with a pair of servers, it is less expensive to deploy using an entry-level SAN than a full-sized enterprise-class SAN.

Best Practice for SAN Storage
Use SCSI-based SAN storage systems. For small installations, entry-level systems may be best; however, for anything else, it is best to use enterprise SAN systems for increased redundancy.

The last entry in the storage realm is that of NAS devices (see Figure 1.10), which present file systems using various protocols including Network File System (NFS), Internet SCSI (iSCSI), and Common Internet File System (CIFS). Of particular interest is the iSCSI protocol, which is SCSI over Internet Protocol (IP). This protocol is not supported as a storage location for virtual machine disk files in ESX versions earlier than 3.0, but support is available for later versions. With NAS, there is no need for Fibre Channel adapters, only more NICs to support the iSCSI and NFS protocols while providing redundancy. In general, iSCSI and NAS run slightly more slowly than Fibre Channel when looking at the raw speeds networking currently available.

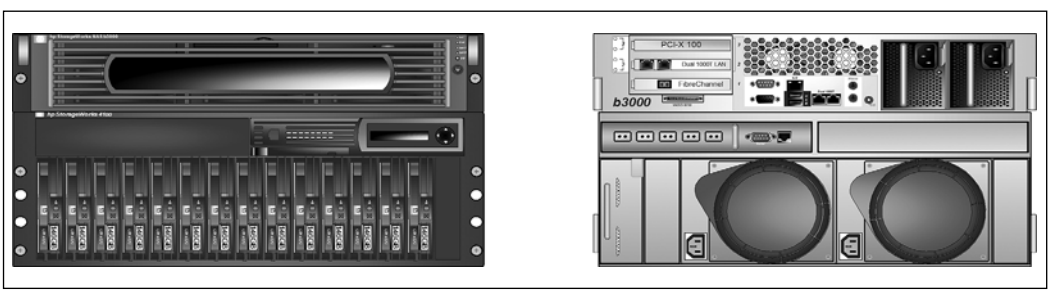

Figure 1.10 NAS device
Visio templates for image courtesy of Hewlett-Packard.

Best Practice for iSCSI
NAS or iSCSI are not supported on versions earlier than ESX version 3.0; do not use this device until an upgrade is available. Also, have enough COS NIC ports to provide redundancy and bandwidth.

Examples

Now it is time to review what customers have done in relation to the comments in the previous sections. The following six examples are from real customers, not from our imagination. The solutions proposed use the best practices previously discussed and a little imagination.

Example 1: Existing Datacenter

A customer was in the midst of a hardware-upgrade cycle and decided to pursue alternatives to purchasing quite a bit of hardware; the customer wanted to avoid buying 300+ systems at a high cost. They decided to pursue ESX Server. Furthermore, the customer conducted an exhaustive internal process to determine the need to upgrade the 300+ systems and believes all of them could be migrated to ESX, because they meet or exceed the documented constraints. Their existing machine mix includes several newer machines from the last machine refresh (around 20), but is primarily made up of machines that are at least 2 to 3 generations old, running on processors no faster than 900MHz. The new ones range from 1.4GHz to 3.06GHz 2U machines (see Figure 1.4). The customer would also like to either make use of their existing hardware somehow or purchase very few machines to make up the necessary difference, because the price for ESX to run 300+ machines approaches their complete hardware budget. In addition, a last bit of information was also provided, and it really throws a monkey wrench into a good solution: They have five datacenters with their own SAN infrastructure.

Following best practices, we could immediately state that we could use the 3.06GHz hosts. Then we could determine whether there were enough to run everything. However, this example shows the need for something even more fundamental than just hardware to run 300+ virtual machines. It shows the need for an appropriate analysis of the running environment to first determine whether the 300+ servers are good candidates for migration, followed by a determination of which servers are best fit to be the hosts of the 300+ VMs. The tool used most often to perform this analysis is the AOG Capacity Planner. This tool will gather

up various utilization and performance numbers for each server over a one- to two-month period. This information is then used to determine which servers make good candidates to run as VMs.

Best Practice
Use a capacity planner or something similar to get utilization and performance information about servers.

When the assessment is finished, you can better judge which machines could be migrated and which could not be. Luckily, the customer had a strict "one application per machine" rule, which was enforced, and which removes possible application conflicts and migration concerns. With the details released about their current infrastructure, it was possible to determine that the necessary hardware was already in use and could be reused with minor hardware upgrades. Each machine would require dual-port NIC and Fibre Channel cards and an increase in memory and local disk space. To run the number of VMs required and to enable the use of VMotion, all machines were paired up at each site at the very least, with a further recommendation to purchase another machine per site (because there were no more hosts to reuse) at the earliest convenience so that they could alleviate possible machine failures in the future. To perform the first migrations, some seed units would be borrowed from the manufacturer and LUNs carved from their own SANs allowing migration from physical to virtual using the seed units. Then the physical host would be converted to an ESX Server and the just-migrated VM VMotioned off the borrowed seed host. This host would be sent to the other sites as their seed unit when the time came to migrate the hosts at the next datacenter. This initial plan would be revised once the capacity planner was run and analyzed.

Example 2: Office in a Box

One of the author's earliest questions was from a company that wanted to use ESX to condense hundreds of remote locations into one easy-to-use and -administer package of a single host running ESX with the remote office servers running as VMs. Because the remote offices currently used outdated hardware, this customer also felt that he should use ESX because it would provide better remote management capability. The customer also believed that the hardware should be upgraded at these remote offices all over the world. Their goal was to ship a box to the

remote location, have it plugged in, powered up, and then remotely manage the server. If there were a machine failure of some sort, they would ship out a new box. The concern the customer had was the initial configuration of the box and how to perform backups appropriately.

One of the very first questions we ask the customer is whether they will be using Microsoft Clusters now or in the future of their ESX deployment. When we first started the discussions, they claimed this was never going to be the case. Just in case, we made sure that they set up their six-drive dual-processor machines with a full complement of memory and disks, an extra dual-port Ethernet card, an external tape device via an Adaptec card (see Figure 1.11), and enough file system space for a possible shared system. We discussed a SAN and the use of VMotion, but the customer thought that this would be overkill for their remote offices. For their datacenter, this was a necessity, but not for a remote office.

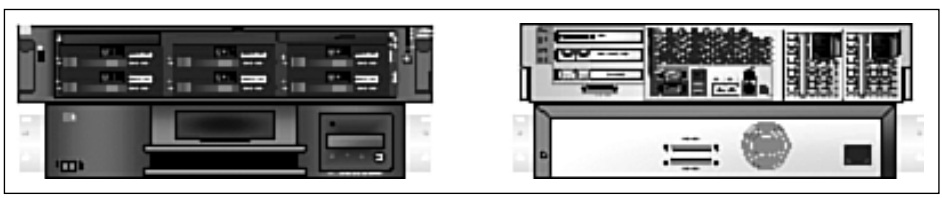

Figure 1.11 Office in a box server with tape library
Visio templates for image courtesy of Hewlett-Packard.

However, the best-laid plan was implemented incorrectly, and a year after the initial confirmation of the customer's design, they needed to implement Microsoft Clustering as a cluster in a box. Because of this oversight, the customer had to reinstall all the ESX Servers to allocate a small shared-mode VMFS. They had to reinstall their machines, but first they set up their operating system disk as a RAID 1, making using of hardware mirroring between disks 1 and 2, leaving the last four disks to make a RAID 5 + 1 spare configuration of 146GB disks. The smaller disks met their VM load quite nicely. On the RAID 5 LUN, they created two file systems, one for the VMFS for the public (nonclustered) VMs and a smaller partition for the shared data drives for the cluster.

Although using a single partition for the two distinct VMFSs is not generally recommended because of LUN-locking considerations, it can be and has been done in a single host environment, as we are discussing. If an entry-level SAN (refer to Figure 1.8) were used, another host would have been added, and the multiple partition approach would *not* be a best practice due to the nature of SCSI reservations, which are further discussed in Chapter 5. However, in a single-host

configuration, SCSI reservations are less of a concern, so use of multiple partitions on the same LUN is not going against any best practices. Ideally, it would be proper to have three LUNs: RAID 1 for the OS and RAID 5 for both necessary VMFSs. However, three LUNs would require at least eight disks, and a disk array would have been necessary, increasing the expense for not much gain, because the VMs in question are small in number and size.

Example 3: The Latest and Greatest

One of our opportunities dealt with the need for the customer to use the latest and greatest hardware with ESX Server and in doing so to plan for the next release of the OS at the same time. The customer decided to go with a full blade enclosure using dual CPU blades with no disk, and many TOE cards so that they could boot their ESX Servers via iSCSI from a NAS (see Figure 1.12). The customer also required an easier and automated way to deploy their ESX Servers.

This presented several challenges up front. The first challenge was that the next release of the OS was not ready at the time, and the HCL for the current release *and* the first release of the next version of ESX showed that some of their desired options would *not* be implemented. So, to use ESX, the hardware mix needed to be changed for ESX version 2.5 and for version 3.0. The customer therefore traded in the TOE cards for Fibre cards or blanks. They also realized that iSCSI and NAS receive limited support in the first release of ESX version 3.0. Therefore, they also needed to get access to local disks to implement their desired virtualization.

The main concern here is that the customer wanting the latest and greatest instead got a mixed bag of goodies that were not compatible with the current release, and the prelist of the HCL for the next release did not list their desired hardware either. In essence, if it is not on the HCL now, most likely it will not be on the list in the future; if you can get a prerelease HCL, this can be verified. In essence, this customer had to change their plans based on the release schedules, and it made for quite a few headaches for the customer and required a redesign to get started, including the use of on-board SCSI drives and the use of a SAN. In essence, always check the HCL on the VMware website before purchasing anything.

As for the deployment of ESX, the on-board remote management cards and the multiple methods to deploy ESX made life much easier. Because these concepts are covered elsewhere, we do not go into a lot of detail. ESX provides its own method for scripted installations just for blades. Many vendors also provide mechanisms to script the installations of operating systems onto their blades. The key

to scripted installations is adding in all the extra bits often required that are outside of ESX, including hardware agents and other necessary software.

Example 4: The SAN

Our fourth example is a customer who brought in consulting to do a bake-off between competing products using vendor-supplied small SANs. Eventually, the customer made a choice and implemented the results of the bake-off in their production environment that used a completely different SAN that had some significant differences in functionality. Although this information was available during the bake-off, it was pretty much a footnote. This in turn led to issues with how they were implementing ESX in production that had to be reengineered. What made this customer unique is that they wanted to get ESX 3.0 style functionality while using ESX 2.5. Although a noble goal, it leads to setting up 2.5 in a mode that does not follow best practices but that is supportable. The customer wanted to store all VM data on the SAN, including the VM configuration and log files. The customer wrote up their desire and wanted confirmation that this was a supportable option.

The architecture decided upon called for each ESX Server to mount a home directory from the SAN so that VM configuration files could be stored on the SAN, and because the VMFS was already on the SAN, everything related to a VM would be stored on the SAN using two distinctly different file systems. To enable the multiple SAN-based file systems, it is necessary to share the Fibre Channel Adapters between the COS and the VMs for ESX versions before 3.0. The sharing of the Fibre Channel adapters is not a best practice and often causes problems. To limit issues, it is best to have one file system per LUN. Because the customer wanted to have the configuration files available to each possible server, the customer created multiple Linux ext3 file systems sharing the same LUN. This also does not follow the best practice of one file system per LUN. However, they did not mix file system types, so there are no Linux file systems sharing a portion of a LUN with VMFS. This is a good thing because both the VMkernel and the Linux kernel can lock a LUN separately when Fibre Channel adapters are shared, and this will cause SCSI reservations and other SCSI issues. We discuss these issues in Chapter 5.

Even though this customer uses several functions that do not follow best practices, this example is here to point out that although best practices exist, they do not define what is supported or even capable with ESX. We confirmed their architecture was supportable, but also pointed out the best practices and possible problems. Many of the items that were not best practices with ESX versions earlier than 3.0 are now a part of ESX version 3.0. From this example, ESX version 3.0 incorpo-

rates the storage of VM configuration and disk files on a VMFS, instead of needing to use multiple file systems and possibly problematic configurations. Understanding the limitations of ESX will aid in the use of ESX with various hardware.

Example 5: Secure Environment

It is increasingly common for ESX to be placed into secure environments as long as the security specialist understands how ESX works and why it is safe to do so. However, in this case, the security specialist assumed that because the VMs share the same air they are therefore at risk. Although we could prove it was not the case, the design of the secure environment had to work within this limitation. The initial hardware was two dual-CPU machines and a small SAN that would later be removed when they proved everything worked and their large corporate SANs took over. The customer also wanted secure data not to be visible to anyone but the people in the teams using the information.

This presented several concerns. The first is that the administrators of the ESX box must also be part of the secure teams, have the proper corporate clearances, or be given an exception, because anyone with administrator access to an ESX Server also has access to all the VMDKs available on the ESX Server. Chapter 4, "Auditing, Monitoring, and Securing," goes into securing your ESX environment in quite a bit of detail, but suffice to say, virtualization has its own issues. Because the customer wanted to secure their data completely, it is important to keep the service console, VMotion, and the VM networks all on their own secure networks, too. Why should we secure VMotion and everything? Because VMotion will pass the memory footprint of the server across an Ethernet cable and, combined with access to the service console, will give a hacker everything a VM is doing. If not properly secured, this is quite a frightening situation.

Whereas the company had a rule governing use of SANs to present secure data LUNs, they had no such policy concerning ESX. In essence, it was important to create an architecture that kept all the secure VMs to their own set of ESX Servers and place on another set of ESX Servers those things not belonging to the secure environment. This kept all the networking separated by external firewalls and kept the data from being accessed by those not part of the secure team. If a new secure environment were necessary, another pair of ESX Servers (so we can VMotion VMs) would be added with their own firewall.

The preceding could have easily been performed on a single ESX Server, yet require the administrators to have the proper corporate clearances to be allowed to manipulate secured files. Given this and the appropriate network configuration inside ESX, it is possible to create many different secure environments within a

single ESX host, including access to other secure machines external to ESX. However, this customer did not choose this option.

Example 6: Disaster Recovery

We were asked to do a DR plan for a customer that had two datacenters in close proximity to each other. The customer wanted a duplicate set of everything at each site so that they could run remotely if necessary. This is not an uncommon desire, because they in effect wanted a hot site implementation. Their current ESX Server load was two dual-CPU hosts at each location, two distinctly different SANs, and some slightly different operational procedures. The currently light load on each ESX Server would eventually grow until new machines were placed in the environment.

Due to the disparate SAN environments, it was impossible to create a SAN copy of the data because the SANs spoke different languages. Therefore, a hardware solution to the problem was out of the question. This in turn led to political issues that had to be ironed out. Once allowed to proceed, the decision was made to create backups using some other mechanism and physically copy the VMs from site to site using some form of automated script. Although there are plenty of tools that already do this, ESX comes equipped with the necessary script to make backups of VMs while they are still running, so in essence a hot copy can be made by ESX with a bit of scripting. Tie this to a local tape drive (which the customer also wanted to place into the mix) and a powerful local and remote backup solution emerges.

Various other approaches were discussed, but unfortunately, they would not work. A key idea was to use VMotion, but the distances involved implied the VMs would be shipped over a very long yet dedicated wire from site to site, which would put the memory footprints of the VMs at risk. Earlier versions of ESX solve this issue by not allowing VMotion to work through a gateway and router. ESX version 3 on the other hand allows VMotion to work through a router and gateway. Another possibility was the use of an offsite backup repository, but that would make restoration slower.

A plan was devised that made the best use of the resources, including remote backups, backup to tape, and storage of tapes offsite. In essence, everything was thought about, including the requirement for a third site in case the impossible regional disaster hit. Little did we know....

The DR plan that was implemented made restoration much easier when the natural disaster hit. What could have taken weeks to restore took just days

because the customer had DR backups of the virtual disk files for every VM on the system. These types of backups happen through the COS and should be considered as part of any deployment of ESX. A backup through the VMs, which is the traditional method to back up servers, requires other data-restoration techniques that take much longer than a backup and restore of a single file.

Hardware Checklist

Now that we have been through a few of the concepts related to the hardware and the individual limitations of various machines listed, we can devise a simple hardware checklist (see Table 1.3) that, if followed, will create a system that follows best practices.

Table 1.3

Hardware Checklist

Hardware	Best Practice	Comments
Network adapters (discussed further in Chapter 8)	Two gigabit ports for service console	Two gigabit ports could be used for ESX version 3.0 with load balancing and failover, but for ESX version 2.5.x or earlier a watchdog is necessary.
	Two gigabit ports for VMotion	ESX 2.5.x: Two gigabit ports could be used, but the second port is purely for failover.
	Two gigabit ports per network available to the VMs	More than two gigabit ports in a team can cause switching issues. 802.1q VLAN tagging is also available.
	Two gigabit or more ports for NAS	ESX version 3.0 only. Two gigabit ports provide failover and bandwidth.
		ESX version 3.0 only. NFS is the only supported NAS protocol. CIFS is not supported.
iSCSI	Two gigabit ports for iSCSI either in the form of gigabit NICs or an iSCSI HBA	ESX version 3.0 only. Support for boot from iSCSI required an iSCSI HBA. An iSCSI HBA is a specialized TCP Offload Engine NIC.

continues...

Table 1.3 continued

Hardware Checklist		
Hardware	**Best Practice**	**Comments**
Fibre Channel adapters (discussed further in Chapter 5)	Two 2GbE (Gigabit Ethernet) ports	This will provide failover and some multipath functionality with active-active style of SANs.
	Two 4GbE ports	In the future, 4GbE Fibre Channel ports will be supported.
Tape drives or libraries	Adaptec SCSI card	Internal and external tape drives or libraries require an Adaptec SCSI card to be of use.
CPU	Match CPUs within a host	
	Match CPUs between hosts	Required for VMotion.
Disk (discussed further in Chapter 12)	Minimum a 72GB RAID 1 for OS	
	Minimum a 2xMemory RAID 0 for virtual swap	If 2xMemory is 64Gb or less, only one RAID 0 is necessary. If 2xMemory is 128GB, two 64GB RAID 0 disk is necessary. ESX <= 2.5.x only.
	RAID 5 for local VMFS	This is mainly for DR purposes, or if you do not have SAN or iSCSI storage available.

Extra Hardware Considerations

All versions of ESX support connections from the VirtualCenter Management Server, and for ESX version 3 there is the license server and the VMware Consolidated Backup (VCB) proxy server. Because these tools are used to manage or interact with the ESX datacenter it might be necessary to consider the need for specialized hardware to run them and the databases to which they connect. Although many administrators run VirtualCenter within a VM, others never run it from a VM.

Best Practices for Virtual Infrastructure non-ESX Servers

VCB proxy server must run from a physical server because the LUNs attached to the ESX Servers must be presented to the VCB proxy server.

VirtualCenter Management Server can run from a VM, but the best practice is to use a physical server.

VMware License Server should always run on a physical server. It does not need to be a large machine. It is a good idea to keep it with the VirtualCenter Management Server.

Database Server, used by VirtualCenter, should reside on a SQL clustered set of servers. One node of the cluster could be a VM for backup functionality.

Conclusion

There is quite a bit to consider from the hardware perspective when considering a virtualization server farm. Although we touch on networking, storage, and disaster recovery in this chapter, it should be noted that how the hardware plays out depends on the load, utilization goals, compression ratios desired, and the performance gains of new hardware (which were not discussed). The recommendations in this chapter are suggestions of places to start the hardware design of a virtualization server farm. Chapter 2, "Version Comparison," delves into the details of and differences between ESX version 3.0 and earlier versions to help you better understand the impact of hardware on ESX. Understanding these differences will aid you in coming up with a successful design of a virtual environment.

This page intentionally left blank

Chapter 2

Version Comparison

VMware started with a "please try this, it is cool, and tell us what to fix" version of VMware Workstation. Soon after that, VMware Workstation version 2 came out, and the world of computing changed. When version 4 of VMware Workstation came out, more and more people started to use the product, and soon after came the server versions GSX and ESX. With ESX, another change to computing took place, and *virtualization* has become the buzzword and driving force behind many datacenter choices.

VMware produces four major products with varying capabilities and functionality. The products form a triangle where VMware Workstation is at the bottom with the broadest ranges of functionality and capability. It is here that VMware tries out new ideas and concepts, making it the leading edge of virtualization technology. The second tier is VMware ACE and VMware Player, which play VMs that already exist and provide many of the capabilities of VMware Workstation but without the capability to make changes. The third tier of the triangle is GSX and VMware Server, which could be said to be VMware Workstation on steroids as it is a middle ground between VMware Workstation and ESX, providing VM Server-style functionality while running upon another operating system: Windows or Linux. The pinnacle tier is ESX, which is its own operating system and the version comparison covered within this chapter.

The VMware Infrastructure product consists of Workstation, ACE, GSX, and ESX. The VMware Administration product is composed of VMware Virtual Center Server (VC), VMware High Availability (HA), Distributed Resource Scheduling (DRS), SAN, iSCSI, and NAS, and VMotion. The last product suite is VMware Tools, which is composed of the VMware Converter the new Physical to Virtual (P2V) and VMware Consolidated Backup (VCB).

ESX version 3 and ESX version 2.5.x differ in many ways, and some of them revolve around what Administration products are available to each, and others around how the