Syngress would like to thank Simon Crosby, Ian Pratt, Roger Klorese, Aimee Francioni, Leo Zarkhin, and the rest of the team at XenSource for your support and guidance throughout the development of this project. We could not have done it without you.
Lead Author

David E. Williams is a principal at Williams & Garcia, LLC, a consulting practice based in Atlanta, GA, specializing in effective enterprise infrastructure solutions. He specializes in the delivery of advanced solutions for x86 and x64 environments. Because David focuses on cost containment and reduction of complexity, virtualization technologies have played a key role in his recommended solutions and infrastructure designs. David has held several IT leadership positions in various organizations, and his responsibilities have included the operations and strategy of Windows, open systems, mainframe, storage, database, and data center technologies and services. He has also served as a senior architect and an advisory engineer for Fortune 1000 organizations, providing strategic direction on technology infrastructures for new enterprise-level projects.

David studied Music Engineering Technology at the University of Miami, and he holds MCSE+I, MCDBA, VCP, and CCNA certifications. When not obsessed with corporate infrastructures, he spends his time with his wife and three children.

Contributors

Kris Buytaert is Founder and CTO of X-Tend. He is a longtime Linux, Security, and Open Source consultant. He has consulting and development experience with multiple enterprise-level clients and government agencies. In addition to his high-level technical experience, he is also a team leader who likes to deliver his projects on time. He is a contributor to the Linux Documentation Project and author of various technical publications. Kris is a Red Hat Certified Engineer and is currently the maintainer of the openMosix HOWTO Web site. Kris is also a frequent speaker at Linux and OpenSource conferences. He is currently focusing on Linux clustering (both HA and HPC), virtualization, and large infrastructure management.
Juan R. Garcia is a Principal Consultant at Williams & Garcia, LLC. He provides strategic and technical consulting in legacy systems migrations, enterprise architecture, disaster recover planning, and enterprise IT resource consolidation to Williams & Garcia’s customers. He specializes in open systems (UNIX/Linux), virtualization technologies (VMware, Xen, and AIX 5L), storage solutions, and RDMBS technologies. Juan’s previous positions include Solutions Architect for Bellsouth, Senior Enterprise Architect for John H. Harland Co., and Technical Manager for Sun Professional Services.

Rami Rosen (B.Sc, Computer Science, Technion—Israel High Institute of Technology) is working as a Linux and Open Solaris kernel programmer accompanying advanced networking and security projects. His background includes positions in Ethernet switching and Avionic operating system start-ups. His specialities include virtualization technologies and kernel networking internals. His articles are occasionally published in the Linux Journal and the lwn.net Web site.

Simon Crosby is an industry evangelist for the Xen™ open source hypervisor and CTO of XenSource Inc. In this position, he is responsible for XenEnterprise R&D, technology leadership, and product management. He also maintains a close affiliation to the Xen project run by Ian Pratt, the founder of XenSource. Prior to XenSource, Simon was a principal engineer at Intel, where he led strategic research in distributed autonomic computing, platform security, and trust. Before joining Intel, Simon founded CPlane Inc., a network optimization software vendor, and held a variety of executive roles while there, including president & CEO, chairman, and CTO. Prior to his position at CPlane, Simon was a tenured faculty member at the University of Cambridge, U.K., where he led research on network performance and control, and multimedia operating systems. He is an
author of more than 35 research papers and patents on a number of data center and networking topics, including security, network and server virtualization, and resource optimization and performance. Simon is a frequent speaker on the topic of enterprise-grade virtualization with open source technologies, and he has most recently been a presenter at such well-known industry events as LinuxWorld, Interop, and the Server Blade Summit.
Contents

Foreword .................................. xix

Chapter 1 An Introduction to Virtualization ............ 1
Introduction .................................. 2
What Is Virtualization? .......................... 2
The History of Virtualization ....................... 3
The Atlas Computer ......................... 3
The M44/44X Project ......................... 4
CP/CMS .................................. 4
Other Time-Sharing Projects ..................... 5
Virtualization Explosion of the 1990s and Early 2000s 6
The Answer: Virtualization Is .................... 8
Why Virtualize? ................................ 9
Decentralization versus Centralization ............ 9
True Tangible Benefits .......................... 13
Consolidation .................................. 15
Reliability ................................... 17
Security ..................................... 18
How Does Virtualization Work? .................. 19
OS Relationships with the CPU Architecture .......... 20
The Virtual Machine Monitor and Ring-0 Presentation 22
The VMM Role Explored ........................ 23
The Popek and Goldberg Requirements .......... 23
The Challenge: VMMs for the x86 Architecture ...... 25
Types of Virtualization .......................... 26
Server Virtualization .......................... 26
Storage Virtualization .......................... 29
Network Virtualization .......................... 30
Application Virtualization ........................ 31
Common Use Cases for Virtualization ............... 32
Technology Refresh ........................... 32
Business Continuity and Disaster Recovery .......... 34
Proof of Concept Deployments ..................... 35
Virtual Desktops ............................... 35
Rapid Development, Test Lab, and Software Configuration Management ...........36
Summary .........................................................38
Solutions Fast Track .................................38
Frequently Asked Questions .........................41

Chapter 2 Introducing Xen .........................43
Introduction .................................................44
What Is Xen? .............................................44
    Features of Xen ................................46
    The XenServer Product Family ..........47
Xen’s Virtualization Model Explored ........50
    Architecture Overview ..............50
    Processor Architecture ........51
    Paravirtualization with Xen ........52
    Xen Domains ..................................54
CPU Virtualization ..........................58
    Exceptions ................................59
    CPU Scheduling ......................60
    Time ......................................62
Memory Virtualization ..................63
    Memory Allocation ................64
    Page Tables and Segmentation ....66
    Virtual Address Translation ....69
I/O Virtualization .....................72
    Device I/O Rings ................73
    Event Channels ...................76
    Virtual I/O Devices and Split Device Drivers ....76
        Network I/O ..................78
        Block I/O ...................79
        Trusted Platform Module and Other Devices ..79
    Driver Domains ................79
    Software and Hardware IOMMUs ....81
        SWIOTLB ...................81
        Grant Tables ................82
    The Xenstore ...................82
Summary ...........................................87
Solutions Fast Track .....................87
Frequently Asked Questions ...............92
Chapter 3 Deploying Xen: Demystifying the Installation ........................................... 95
  Introduction ........................................... 96
  Determining Which Xen to Choose .......... 96
  System Requirements ......................... 97
  Thinking Before You Start ................. 99
  Installing Xen on a Free Linux Distribution .... 101
    Fedora Core 6 ........................... 101
    VirtManager .................................. 114
    Installing Windows XP ...................... 121
  Installing the XenServer Product Family ...... 125
    What Is XenServer? .............. 125
    XenServer Requirements .......... 126
  Getting and Installing XenServer ............... 127
    Installing the Host .................. 127
    Client Installation .................. 135
    Installing an Initial Virtual Machine on XenServer 138
  Other Xen Installation Methods ............... 142
    Using the XenSource Binaries and LVM ...... 142
  Configuring Xen .................................. 146
  Getting Xen on Your Network .................. 149
  Summary .......................................... 153
  Solutions FastTrack ............................ 153
  Frequently Asked Questions .................. 156

Chapter 4 The Administrator Console and Other Native Tools ......................... 159
  Introduction ...................................... 160
  Native Xen Command-Line Tools ............... 161
    The xe Command-Line Interface ........... 161
      Installing and Cloning XenVMs .......... 162
      Starting Up, Shutting Down, Rebooting, Suspending, and Resuming XenVMs 162
      Shutting Down and Rebooting XenHosts 163
      Query Options for XenHosts ............. 163
  XenServer Administrator Console .......... 163
    System Requirements for the Administrator Console 164
    Installing the Administrator Console .... 164
Contents

Installing the Administrator Console on Windows
(XP/2000/2003) .............................................. 164
Installing the Administrator Console on Linux ... 169
Using the Administrator Console ..................... 170
Working with Hosts .................................. 171
Deploying and Configuring XenVMs ............... 173
Summary .................................................. 177
Solutions Fast Track ................................ 177
Frequently Asked Questions .......................... 178

Chapter 5 Managing Xen with
Third-Party Management Tools ..................... 179

Introduction .......................................... 180
Qlusters openQRM ..................................... 180
  Xen Management with openQRM .................. 181
    Overview ....................................... 181
    General Concepts for the Xen/openQRM Mix ... 182
    Plug-ins and Licensing ......................... 184
    Installing openQRM ............................ 187
    System Requirements ........................... 188
    Installing openQRM 3.1.x Server ............ 190
    Installing the openQRM Xen Plug-in ...... 193
    Managing Xen with openQRM ................. 196
    Provisioning with openQRM-Pro ............ 201
Enomalism ............................................. 203
  Overview of Enomalism ........................... 204
  Installing Enomalism ............................. 205
  System Requirements ............................ 205
  Installation Walkthrough ....................... 206
  Using Enomalism to Manage Xen .............. 206
Project ConVirt and XenMan .......................... 210
  Overview of ConVirt ............................ 211
  Installing ConVirt ............................... 212
  System Requirements ........................... 212
  Installation .................................... 212
  Using ConVirt to Manage Xen ................ 214
Summary ............................................. 220
Solutions Fast Track ................................ 220
Frequently Asked Questions ....................... 222
## Chapter 6 Deploying a Virtual Machine in Xen ...... 225

- Introduction .................................................. 226
- Workload Planning and Virtual Machine Placement .... 226
  - Memory ......................................................... 226
  - CPU ............................................................ 227
  - Network .......................................................... 228
- Installing Modified Guests .................................. 229
  - Installing Red Hat Enterprise Linux 4 ................. 229
- Installing Unmodified Guests ............................... 235
  - Installing Red Hat Linux Enterprise 5 ............... 236
- Installing Windows Guests .................................. 240
  - Windows Guest Installation ............................... 240
- Physical-to-Virtual Migrations of Existing Systems .... 245
  - P2V Migration .................................................. 245
- Importing and Exporting Existing Virtual Machines .... 246
  - Exporting XenVMs ............................................. 247
  - Importing XenVMs .......................................... 249
- Summary .......................................................... 252
- Solutions Fast Track ......................................... 252
- Frequently Asked Questions ................................. 255

## Chapter 7 Advanced Xen Concepts ....................... 257

- Introduction .................................................. 258
- The Virtual Split Devices Model ........................... 258
- Advanced Storage Concepts ................................ 259
  - High-Performance Solutions for Xen .................... 259
  - iSCSI Integration with Xen ................................ 260
  - Copy-on-Write ............................................... 263
    - DmUserspace ............................................... 264
    - UnionFS ................................................... 265
- Advanced Networking Concepts ............................ 266
  - Bridging VLANs ............................................. 266
  - Creating Interface Bonds for
    - High Availability and Link Aggregation ............... 267
    - Routing, Forwarding, and Other Network Tricks .... 269
- Building a Xen Cluster ...................................... 270
- XenVM Migration ............................................. 279
- XenVM Backup and Recovery Solutions ................... 282
<table>
<thead>
<tr>
<th>Contents</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Options for Backing Up Your XenVM</td>
<td>282</td>
</tr>
<tr>
<td>Making Xen Part of Your Disaster Recovery Plan</td>
<td>283</td>
</tr>
<tr>
<td>Full Virtualization in Xen</td>
<td>283</td>
</tr>
<tr>
<td>The New Processors with Virtual Extensions (VT-x and AMD-V)</td>
<td>284</td>
</tr>
<tr>
<td>Summary</td>
<td>288</td>
</tr>
<tr>
<td>Solutions Fast Track</td>
<td>288</td>
</tr>
<tr>
<td>Frequently Asked Questions</td>
<td>290</td>
</tr>
<tr>
<td>Chapter 8 The Future of Virtualization</td>
<td>293</td>
</tr>
<tr>
<td>Introduction</td>
<td>294</td>
</tr>
<tr>
<td>The Unofficial Xen Road Map</td>
<td>294</td>
</tr>
<tr>
<td>Performance and Scalability</td>
<td>295</td>
</tr>
<tr>
<td>NUMA-Aware Architecture</td>
<td>296</td>
</tr>
<tr>
<td>Multicore Processors</td>
<td>298</td>
</tr>
<tr>
<td>Smart I/O</td>
<td>299</td>
</tr>
<tr>
<td>Operating System Support</td>
<td>301</td>
</tr>
<tr>
<td>Support in Linux Distributions</td>
<td>301</td>
</tr>
<tr>
<td>Xen and Microsoft</td>
<td>302</td>
</tr>
<tr>
<td>Other HVM Guests</td>
<td>303</td>
</tr>
<tr>
<td>Beyond the x86 CPU Architecture</td>
<td>303</td>
</tr>
<tr>
<td>IA-64 Feature Sync with x86</td>
<td>304</td>
</tr>
<tr>
<td>Porting to PowerPC</td>
<td>304</td>
</tr>
<tr>
<td>Porting to the UltraSPARC Architecture</td>
<td>306</td>
</tr>
<tr>
<td>Architecture Enhancements</td>
<td>307</td>
</tr>
<tr>
<td>Control Tools</td>
<td>307</td>
</tr>
<tr>
<td>Virtual Hard Disk Images and XenFS</td>
<td>308</td>
</tr>
<tr>
<td>Virtual Device Enhancements</td>
<td>309</td>
</tr>
<tr>
<td>Virtual Infrastructure in Tomorrow’s Data Center</td>
<td>311</td>
</tr>
<tr>
<td>Technology Trends Driving</td>
<td>311</td>
</tr>
<tr>
<td>Improvements in Virtualization</td>
<td>311</td>
</tr>
<tr>
<td>Hardware Economies of Scale</td>
<td>311</td>
</tr>
<tr>
<td>Multicore and Multithreaded Computing</td>
<td>312</td>
</tr>
<tr>
<td>Solutions for Small and Medium-Sized Businesses</td>
<td>316</td>
</tr>
<tr>
<td>Integrated Computing</td>
<td>316</td>
</tr>
<tr>
<td>Data Center in a Box</td>
<td>317</td>
</tr>
<tr>
<td>Large Enterprises</td>
<td>318</td>
</tr>
<tr>
<td>Reliability and Availability</td>
<td>318</td>
</tr>
<tr>
<td>Security</td>
<td>320</td>
</tr>
</tbody>
</table>
The open source Xen project, led by Ian Pratt of the University of Cambridge and XenSource Inc., will arguably have a greater impact on the enterprise software industry than Linux has had. The Xen hypervisor is a lightweight, high-performance, secure virtualization platform that is now collaboratively developed by more than 20 of the industry’s largest enterprise IT vendors as an open industry standard for virtualization. Its architecture has tremendous advantages over existing virtualization technologies: It has broad hardware support through its reuse of existing operating systems such as Linux to safely virtualize I/O for other guests; it offers superb resource partitioning for performance isolation and security; and it can be implemented either as a virtualization platform or as an integrated component within an operating system. Xen has been ported to a wide range of hardware architectures, including x86, x86_64, the Intel Itanium, IBM’s PowerPC, the SGI Altix, and the ARM 9. It also is ideally suited to hardware-accelerated virtualization. For these reasons, and because Xen is freely available in source code form, every major OS vendor has either adopted Xen or the Xen architecture as a core component of the next major OS release for the x86 platform.

As a catalyst of change in the IT industry, the Xen project needs to scale its knowledge and skill base so that a competent IT pro can easily get a handle on the technology and how to deploy and use it. Cool technology can be adopted only as fast as human users can acquire the understanding and skills to use it. Perhaps more important, however, is the fact that the Xen project relies on the innovation of the community to continue the development of its feature set. It is therefore a great
pleasure and also a great relief to introduce this timely, thorough, and highly accessible guide to the art of Xen virtualization. This book demystifies Xen by placing it in a practical context that any IT pro who wants to get something working will immediately understand, while also providing a thorough grounding in the architecture and implementation of the hypervisor. It also takes an important step beyond the basics of Xen by offering a detailed tutorial on how to use the definitive platform implementation of Xen, XenEnterprise from XenSource. Although enthusiasts may want to dig into Xen by building it from source, most readers will be delighted to find included with this book a CD containing the powerful free XenExpress bare-metal hypervisor from XenSource ready to install for production virtualization of Linux and Windows guests.

It is my hope that this book will achieve two aims: encourage a new generation of contributors to the Xen project and foster broad adoption of the Xen hypervisor as a ubiquitous open standard for virtualization.

—Simon Crosby
CTO and Founder
XenSource Inc.

About the CDs

This book includes two CDs to help you consolidate Windows, Linux, or mixed deployment with XenExpress 3.2. This free, production-ready virtualization platform enables you to quickly get started with Xen virtualization. With XenExpress you can host up to four virtual servers running Linux or Windows (Intel VT or AMD-V hardware-assist virtualization technology is required for Windows) on a broad range of standard server hardware. XenExpress supports dual-socket servers with up to 4 GB of RAM, offering you all of the base performance, tools, and easy-to-use features of XenEnterprise. Easily installed and seamlessly upgradable to XenEnterprise or XenServer, XenExpress is the ideal on-ramp to Xen.
Chapter 1

An Introduction to Virtualization

Solutions in this chapter:

- What Is Virtualization?
- Why Virtualize?
- How Does Virtualization Work?
- Types of Virtualization
- Common Use Cases for Virtualization

☑ Summary
☑ Solutions Fast Track
☑ Frequently Asked Questions
Introduction

Virtualization is one of those buzz words that has been gaining immense popularity with IT professionals and executives alike. Promising to reduce the ever-growing infrastructure inside current data center implementations, virtualization technologies have cropped up from dozens of software and hardware companies. But what exactly is it? Is it right for everyone? And how can it benefit your organization?

Virtualization has actually been around more than three decades. Once only accessible by the large, rich, and prosperous enterprise, virtualization technologies are now available in every aspect of computing, including hardware, software, and communications, for a nominal cost. In many cases, the technology is freely available (thanks to open-source initiatives) or included for the price of products such as operating system software or storage hardware.

Well suited for most inline business applications, virtualization technologies have gained in popularity and are in widespread use for all but the most demanding workloads. Understanding the technology and the workloads to be run in a virtualized environment is key to every administrator and systems architect who wishes to deliver the benefits of virtualization to their organization or customers.

This chapter will introduce you to the core concepts of server, storage, and network virtualization as a foundation for learning more about Xen. This chapter will also illustrate the potential benefits of virtualization to any organization.

What Is Virtualization?

So what exactly is virtualization? Today, that question has many answers. Different manufacturers and independent software vendors coined that phrase to categorize their products as tools to help companies establish virtualized infrastructures. Those claims are not false, as long as their products accomplish some of the following key points (which are the objectives of any virtualization technology):

- Add a layer of abstraction between the applications and the hardware
- Enable a reduction in costs and complexity
- Provide the isolation of computer resources for improved reliability and security
- Improve service levels and the quality of service
Better align IT processes with business goals
Eliminate redundancy in, and maximize the utilization of, IT infrastructures

While the most common form of virtualization is focused on server hardware platforms, these goals and supporting technologies have also found their way into other critical—and expensive—components of modern data centers, including storage and network infrastructures.

But to answer the question “What is virtualization?” we must first discuss the history and origins of virtualization, as clearly as we understand it.

The History of Virtualization

In its conceived form, virtualization was better known in the 1960s as time sharing. Christopher Strachey, the first Professor of Computation at Oxford University and leader of the Programming Research Group, brought this term to life in his paper *Time Sharing in Large Fast Computers*. Strachey, who was a staunch advocate of maintaining a balance between practical and theoretical work in computing, was referring to what he called multiprogramming. This technique would allow one programmer to develop a program on his console while another programmer was debugging his, thus avoiding the usual wait for peripherals. Multiprogramming, as well as several other groundbreaking ideas, began to drive innovation, resulting in a series of computers that burst onto the scene. Two are considered part of the evolutionary lineage of virtualization as we currently know it—the Atlas and IBM’s M44/44X.

The Atlas Computer

The first of the supercomputers of the early 1960s took advantage of concepts such as time sharing, multiprogramming, and shared peripheral control, and was dubbed the Atlas computer. A project run by the Department of Electrical Engineering at Manchester University and funded by Ferranti Limited, the Atlas was the fastest computer of its time. The speed it enjoyed was partially due to a separation of operating system processes in a component called the supervisor and the component responsible for executing user programs. The supervisor managed key resources, such as the computer’s processing time, and was passed special instructions, or extracodes, to help it provision and manage the computing environment for the user program’s instructions. In essence, this was the birth of the hypervisor, or virtual machine monitor.

In addition, Atlas introduced the concept of virtual memory, called one-level store, and paging techniques for the system memory. This core store was also logically
separated from the store used by user programs, although the two were integrated. In many ways, this was the first step towards creating a layer of abstraction that all virtualization technologies have in common.

The M44/44X Project

Determined to maintain its title as the supreme innovator of computers, and motivated by the competitive atmosphere that existed, IBM answered back with the M44/44X Project. Nested at the IBM Thomas J. Watson Research Center in Yorktown, New York, the project created a similar architecture to that of the Atlas computer. This architecture was first to coin the term *virtual machines* and became IBM’s contribution to the emerging time-sharing system concepts. The main machine was an IBM 7044 (M44) scientific computer and several simulated 7044 virtual machines, or 44Xs, using both hardware and software, virtual memory, and multiprogramming, respectively.

Unlike later implementations of time-sharing systems, M44/44X virtual machines did not implement a complete simulation of the underlying hardware. Instead, it fostered the notion that virtual machines were as efficient as more conventional approaches. To nail that notion, IBM successfully released successors of the M44/44X project that showed this idea was not only true, but could lead to a successful approach to computing.

CP/CMS

A later design, the IBM 7094, was finalized by MIT researchers and IBM engineers and introduced Compatible Time Sharing System (CTSS). The term “compatible” refers to the compatibility with the standard batch processing operating system used on the machine, the Fortran Monitor System (FMS). CTSS not only ran FMS in the main 7094 as the primary facility for the standard batch stream, but also ran an unmodified copy of FMS in each virtual machine in a background facility. The background jobs could access all peripherals, such as tapes, printers, punch card readers, and graphic displays, in the same fashion as the foreground FMS jobs as long as they did not interfere with foreground time-sharing processors or any supporting resources.

MIT continued to value the prospects of time sharing, and developed Project MAC as an effort to develop the next generation of advances in time-sharing technology, pressuring hardware manufacturers to deliver improved platforms for their work. IBM’s response was a modified and customized version of its System/360
(S/360) that would include virtual memory and time-sharing concepts not previously released by IBM. This proposal to Project MAC was rejected by MIT, a crushing blow to the team at the Cambridge Scientific Center (CSC), whose only purpose was to support the MIT/IBM relationship through technical guidance and lab activities.

The fallout between the two, however, led to one of the most pivotal points in IBM’s history. The CSC team, lead by Norm Rasmussen and Bob Creasy, a defect from Project MAC, contributed to the development of CP/CMS. In the late 1960s, the CSC developed the first successful virtual machine operating system based on fully virtualized hardware, the CP-40. The CP-67 was released as a reimplementation of the CP-40, as was later converted and implemented as the S/360-67 and later as the S/370. The success of this platform won back IBM’s credibility at MIT as well as several of IBM’s largest customers. It also led to the evolution of the platform and the virtual machine operating systems that ran on them, the most popular being VM/370. The VM/370 was capable of running many virtual machines, with larger virtual memory running on virtual copies of the hardware, all managed by a component called the virtual machine monitor (VMM) running on the real hardware. Each virtual machine was able to run a unique installation of IBM’s operating system stably and with great performance.

Other Time-Sharing Projects

IBM’s CTSS and CP/CMS efforts were not alone, although they were the most influential in the history of virtualization. As time sharing became widely accepted and recognized as an effective way to make early mainframes more affordable, other companies joined the time-sharing fray. Like IBM, those companies needed plenty of capital to fund the research and hardware investment needed to aggressively pursue time-sharing operating systems as the platform for running their programs and computations. Some other projects that jumped onto the bandwagon included:

- Livermore Time-Sharing System (LTSS) Developed by the Lawrence Livermore Laboratory in the late 1960s as the operating system for the Control Data CDC 7600 supercomputers. The CDC 7600 running LTSS took over the title of the world’s fastest computer, trumping on the Atlas computer, which suffered from a form of trashing due to inefficiencies in its implementation of virtual memory.