



Matthew Portnoy

Virtualization

ESSENTIALS

Second Edition

 **SYBEX**
A Wiley Brand

VIRTUALIZATION

ESSENTIALS

SECOND EDITION

VIRTUALIZATION

ESSENTIALS

SECOND EDITION

Matthew Portnoy



Executive Editor: Jody Lefever
Development Editor: Kelly Talbot
Technical Editor: Van Van Noy
Production Editor: Barath Kumar Rajasekaran
Copy Editor: Kathy Grider-Carlyle
Editorial Manager: Mary Beth Wakefield
Production Manager: Kathleen Wisor
Proofreader: Nancy Bell
Indexer: Johnna VanHoose Dinse
Project Coordinator, Cover: Brent Savage
Cover Designer: Wiley
Cover Image: ©DrHitch/Shutterstock

Copyright © 2016 by John Wiley & Sons, Inc., Indianapolis, Indiana
Published simultaneously in Canada

ISBN: 978-1-119-26772-0
ISBN: 978-1-119-26774-4 (ebk.)
ISBN: 978-1-119-26773-7 (ebk.)

Manufactured in the United States of America

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except as permitted under Sections 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 646-8600. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, or online at <http://www.wiley.com/go/permissions>.

Limit of Liability/Disclaimer of Warranty: The publisher and the author make no representations or warranties with respect to the accuracy or completeness of the contents of this work and specifically disclaim all warranties, including without limitation warranties of fitness for a particular purpose. No warranty may be created or extended by sales or promotional materials. The advice and strategies contained herein may not be suitable for every situation. This work is sold with the understanding that the publisher is not engaged in rendering legal, accounting, or other professional services. If professional assistance is required, the services of a competent professional person should be sought. Neither the publisher nor the author shall be liable for damages arising herefrom. The fact that an organization or Web site is referred to in this work as a citation and/or a potential source of further information does not mean that the author or the publisher endorses the information the organization or Web site may provide or recommendations it may make. Further, readers should be aware that Internet Web sites listed in this work may have changed or disappeared between when this work was written and when it is read.

For general information on our other products and services or to obtain technical support, please contact our Customer Care Department within the U.S. at (877) 762-2974, outside the U.S. at (317) 572-3993 or fax (317) 572-4002.

Wiley publishes in a variety of print and electronic formats and by print-on-demand. Some material included with standard print versions of this book may not be included in e-books or in print-on-demand. If this book refers to media such as a CD or DVD that is not included in the version you purchased, you may download this material at <http://booksupport.wiley.com>. For more information about Wiley products, visit www.wiley.com.

Library of Congress Control Number: 2016944315

TRADEMARKS: Wiley, the Wiley logo, and the Sybex logo are trademarks or registered trademarks of John Wiley & Sons, Inc. and/or its affiliates, in the United States and other countries, and may not be used without written permission. All other trademarks are the property of their respective owners. John Wiley & Sons, Inc. is not associated with any product or vendor mentioned in this book.

10987654321

*To my friends
and family,
near and far.*

ACKNOWLEDGMENTS

A project is rarely a solo affair, and this one depended on a large crew for it to arrive. I need to thank Scott Lowe for shoveling the path and aiming me at the correct door. My deepest gratitude goes to Mark Milow for helping me climb aboard this rocket, to Mike Szfranski for your always open book of knowledge, to Nick Gamache for the insights, and to Tony Damiano for keeping our vehicle in the fast lane.

My heartfelt thanks also go to the virtual team at Sybex: Kelly Talbot, Stephanie McComb, Van Van Noy, Kathy Grider-Carlyle, and Barath Kumar Rajasekaran for their steadfast support, forcing me to improve with each chapter and keeping it all neat and clean. Special thanks go to Agatha Kim for getting this whole adventure rolling.

I need to thank my family beginning with my parents, teachers both, who instilled me with a love of reading and writing and set me on a path that somehow led here. Thank you to my boys, Lucas and Noah, who fill our days with laughter and music. And finally, a huge hug to my wife, Elizabeth, who encouraged me even when she had no idea what I was writing about. I love you.

ABOUT THE AUTHOR



Matthew Portnoy has been an information technology professional for more than 30 years, working in organizations such as NCR, Sperry/Unisys, Stratus Computer, Oracle, and VMware. He has been in the center of many of the core technological trends during this period, including the birth of the PC, client-server computing, fault tolerance and availability, the rise of the Internet, and now virtualization, which is the foundation for cloud computing.

As both a presales and post-sales analyst, he has worked with all of the disciplines computing offers, including innumerable programming languages, operating systems, application design and development, database operations, networking, security, availability, and virtualization. He has spoken at the industry's largest virtualization conference, VMworld, and is a frequent speaker at user group meetings. He also has been teaching virtualization and database classes as an adjunct professor at Wake Tech Community College in Raleigh, North Carolina, since 2007.

CONTENTS AT A GLANCE

| | | |
|------------|---|-------------|
| | <i>Introduction</i> | <i>xvii</i> |
| CHAPTER 1 | Understanding Virtualization | 1 |
| CHAPTER 2 | Understanding Hypervisors | 21 |
| CHAPTER 3 | Understanding Virtual Machines | 37 |
| CHAPTER 4 | Creating a Virtual Machine | 55 |
| CHAPTER 5 | Installing Windows on a Virtual Machine | 81 |
| CHAPTER 6 | Installing Linux on a Virtual Machine | 107 |
| CHAPTER 7 | Managing CPUs for a Virtual Machine | 133 |
| CHAPTER 8 | Managing Memory for a Virtual Machine | 143 |
| CHAPTER 9 | Managing Storage for a Virtual Machine | 157 |
| CHAPTER 10 | Managing Networking for a Virtual Machine | 177 |
| CHAPTER 11 | Copying a Virtual Machine | 197 |
| CHAPTER 12 | Managing Additional Devices in Virtual Machines | 221 |
| CHAPTER 13 | Understanding Availability | 239 |
| CHAPTER 14 | Understanding Applications in a Virtual Machine | 255 |
| APPENDIX | Answers to Additional Exercises | 275 |
| | <i>Glossary</i> | <i>287</i> |
| | <i>Index</i> | <i>293</i> |

CONTENTS

Introduction

xvii

| | | |
|------------------|--|-----------|
| CHAPTER 1 | Understanding Virtualization | 1 |
| | Describing Virtualization | 1 |
| | Microsoft Windows Drives Server Growth | 3 |
| | Explaining Moore's Law | 6 |
| | Understanding the Importance of Virtualization | 9 |
| | Examining Today's Trends | 12 |
| | Virtualization and Cloud Computing | 14 |
| | Understanding Virtualization Software Operation | 15 |
| | Virtualizing Servers | 16 |
| | Virtualizing Desktops | 17 |
| | Virtualizing Applications | 18 |
| CHAPTER 2 | Understanding Hypervisors | 21 |
| | Describing a Hypervisor | 21 |
| | Exploring the History of Hypervisors | 22 |
| | Understanding Type 1 Hypervisors | 23 |
| | Understanding Type 2 Hypervisors | 25 |
| | Understanding the Role of a Hypervisor | 26 |
| | Holodecks and Traffic Cops | 26 |
| | Resource Allocation | 27 |
| | Comparing Today's Hypervisors | 29 |
| | VMware ESX | 29 |
| | Citrix Xen | 31 |
| | Microsoft Hyper-V | 33 |
| | Other Solutions | 34 |
| CHAPTER 3 | Understanding Virtual Machines | 37 |
| | Describing a Virtual Machine | 37 |
| | Examining CPUs in a Virtual Machine | 40 |
| | Examining Memory in a Virtual Machine | 41 |
| | Examining Network Resources in a Virtual Machine | 42 |

| | |
|---|----|
| Examining Storage in a Virtual Machine | 43 |
| Understanding How a Virtual Machine Works | 45 |
| Working with Virtual Machines | 47 |
| Understanding Virtual Machine Clones | 48 |
| Understanding Templates | 49 |
| Understanding Snapshots | 50 |
| Understanding OVF | 51 |
| Understanding Containers. | 52 |

CHAPTER 4 Creating a Virtual Machine 55

| | |
|---|----|
| Performing P2V Conversions | 55 |
| Investigating the Physical-to-Virtual Process | 56 |
| Hot and Cold Cloning. | 57 |
| Loading Your Environment | 58 |
| Loading VMware Workstation Player | 59 |
| Exploring VMware Workstation Player | 65 |
| Loading VirtualBox | 68 |
| Building a New Virtual Machine | 73 |
| Thinking about VM Configuration | 74 |
| Creating a First VM | 75 |

CHAPTER 5 Installing Windows on a Virtual Machine 81

| | |
|--|-----|
| Loading Windows into a Virtual Machine | 81 |
| Installing Windows 10 | 82 |
| Installing VMware Tools | 93 |
| Understanding Configuration Options. | 97 |
| Optimizing a New Virtual Machine | 103 |

CHAPTER 6 Installing Linux on a Virtual Machine 107

| | |
|---|-----|
| Loading Linux into a Virtual Machine. | 107 |
| Exploring Oracle VM VirtualBox. | 108 |
| Installing Linux into a Virtual Machine | 111 |
| Installing VirtualBox Guest Additions | 123 |
| Understanding Configuration Options. | 125 |
| Optimizing a New Linux Virtual Machine | 129 |

| | | |
|-------------------|--|------------|
| CHAPTER 7 | Managing CPUs for a Virtual Machine | 133 |
| | Understanding CPU Virtualization | 133 |
| | Configuring VM CPU Options | 137 |
| | Tuning Practices for VM CPUs | 138 |
| | Choosing Multiple vCPUs vs. a Single vCPU | 139 |
| | Hyper-Threading | 140 |
| | Working with Intel and AMD Servers | 141 |
| CHAPTER 8 | Managing Memory for a Virtual Machine | 143 |
| | Understanding Memory Virtualization | 143 |
| | Configuring VM Memory Options | 146 |
| | Tuning Practices for VM Memory | 148 |
| | Calculating Memory Overhead | 148 |
| | Memory Optimizations | 150 |
| CHAPTER 9 | Managing Storage for a Virtual Machine | 157 |
| | Understanding Storage Virtualization | 157 |
| | Configuring VM Storage Options | 162 |
| | Tuning Practices for VM Storage | 169 |
| CHAPTER 10 | Managing Networking for a Virtual Machine | 177 |
| | Understanding Network Virtualization | 177 |
| | Configuring VM Network Options | 187 |
| | Tuning Practices for Virtual Networks | 192 |
| CHAPTER 11 | Copying a Virtual Machine | 197 |
| | Cloning a Virtual Machine | 197 |
| | Working with Templates | 207 |
| | Saving a Virtual Machine State | 212 |
| | Creating a Snapshot | 214 |
| | Merging Snapshots | 218 |
| CHAPTER 12 | Managing Additional Devices in Virtual Machines | 221 |
| | Using Virtual Machine Tools | 222 |
| | Understanding Virtual Devices | 224 |

| | | |
|-------------------|---|------------|
| | Configuring a CD/DVD Drive | 224 |
| | Configuring a Floppy Disk Drive | 225 |
| | Configuring a Sound Card | 229 |
| | Configuring USB Devices | 230 |
| | Configuring Graphic Displays | 232 |
| | Configuring Other Devices. | 233 |
| CHAPTER 13 | Understanding Availability | 239 |
| | Increasing Availability | 239 |
| | Protecting a Virtual Machine | 242 |
| | Protecting Multiple Virtual Machines | 246 |
| | Protecting Data Centers | 251 |
| CHAPTER 14 | Understanding Applications in a Virtual Machine | 255 |
| | Examining Virtual Infrastructure Performance Capabilities | 255 |
| | Deploying Applications in a Virtual Environment | 260 |
| | Understanding Virtual Appliances and vApps | 269 |
| | Open Stack and Containers | 271 |
| APPENDIX | Answers to Additional Exercises | 275 |
| | <i>Glossary</i> | 287 |
| | <i>Index</i> | 293 |

INTRODUCTION

We live in an exciting time. The information age is exploding around us, giving us access to dizzying amounts of data the instant it becomes available. Smart phones and tablets provide an untethered experience that offers streaming video, audio, and other media formats to just about any place on the planet. Even people who are not “computer literate” use Facebook to catch up with friends and family, use Google to research a new restaurant choice and print directions to get there, or Tweet their reactions once they have sampled the fare. The budding Internet-of-things will only catalyze this data eruption. The infrastructure supporting these services is also growing exponentially, and the technology that facilitates this rapid growth is virtualization.

On one hand, virtualization is nothing more than an increasingly efficient use of existing resources that delivers huge cost savings in a brief amount of time. On the other, virtualization also offers organizations new models of application deployment for greater uptime to meet user expectations, modular packages to provide new services in minutes instead of weeks, and advanced features that bring automatic load balancing, scalability without downtime, self-healing, self-service provisioning, and many other capabilities to support business-critical applications that improve on traditional architecture. Large companies have been using this technology for 10 to 15 years, while smaller and medium-sized businesses are just getting there now. Some of them might miss the movement altogether and jump directly to cloud computing, the next evolution of application deployment. Virtualization is the foundation for cloud computing as well.

This quantum change in our world echoes similar trends from our recent history as electrical power and telephony capabilities spread and then changed our day-to-day lives. During those periods, whole industries sprang up out of nothing, providing employment and opportunity to people who had the foresight and chutzpah to seize the moment. That same spirit and opportunity is available today as this area is still being defined and created right before our eyes. If not virtualization vendors, there are hardware partners who provide servers, networking vendors for connectivity, storage partners for data storage, and everyone provides services. Software vendors are designing and deploying new applications specifically for these new architectures. Third parties are creating tools to monitor and manage these applications and infrastructure areas. As cloud computing begins to become the de facto model for development, deployment, and maintaining application services, this area will expand even further.

The first generation of virtualization specialists acquired their knowledge out of necessity: They were server administrators who needed to understand the

new infrastructure being deployed in their data centers. Along the way, they picked up some networking knowledge to manage the virtual networks, storage knowledge to connect to storage arrays, and application information to better interface with the application teams. Few people have experience in all of those areas. Whether you have some virtualization experience or none at all, this text will give you the foundation to understand what virtualization is, why it is a crucial portion of today's and tomorrow's information technology infrastructure, and the opportunity to explore and experience one of the most exciting and fastest growing topics in technology today.

Good reading and happy virtualizing!

Who Should Read This Book

This text is designed to provide the basics of virtualization technology to someone who has little or no prior knowledge of the subject. This book will be of interest to you if you are an IT student looking for information about virtualization or if you are an IT manager who needs a better understanding of virtualization fundamentals as part of your role. This book might also be of interest if you are an IT professional who specializes in a particular discipline (such as server administration, networking, or storage) and are looking for an introduction into virtualization or cloud computing as a way to advance inside your organization.

The expectation is that you have:

- ▶ Some basic PC experience
- ▶ An understanding of what an operating system is and does
- ▶ Conceptual knowledge of computing resources (CPU, memory, storage, and network)
- ▶ A high-level understanding of how programs use resources

This text would not be of interest if you are already a virtualization professional and you are looking for a guidebook or reference.

What You Need

The exercises and illustrations used in this text were created on a system with Windows 10 as the operating system. VMware Workstation Player version 12 is used as the virtualization platform. It is available as a free download from <http://downloads.vmware.com/d/>. It is recommended that you have at least 2 GB of memory, though more will be better. The installation requires 150 MB

of disk storage. Also used is Oracle VirtualBox version 5. It is available as a free download from <http://www.virtualbox.org>. It is recommended that you have at least 2GB of memory. VirtualBox itself requires only about 30 MB of disk storage, but virtual machines will require more.

The examples demonstrate the creation and use of two virtual machines: one running Windows 10, the other running Ubuntu Linux. You will need the installation media for those as well. Each of the virtual machines requires about 30 GB of disk space.

What Is Covered in This Book

Here's a glance at what is in each chapter.

Chapter 1: Understanding Virtualization Introduces the basic concepts of computer virtualization beginning with mainframes and continues with the computing trends that have led to current technologies.

Chapter 2: Understanding Hypervisors Focuses on hypervisors, the software that provides the virtualization layer, and compares some of the current offerings in today's marketplace.

Chapter 3: Understanding Virtual Machines Describes what a virtual machine is composed of, explains how it interacts with the hypervisor that supports its existence, and provides an overview of managing virtual machine resources.

Chapter 4: Creating a Virtual Machine Begins with the topic of converting existing physical servers into virtual machines and provides a walkthrough of installing VMware Workstation Player and Oracle VirtualBox, the virtualization platforms used in this text, and a walkthrough of the creation of a virtual machine.

Chapter 5: Installing Windows on a Virtual Machine Provides a guide for loading Microsoft Windows in the created virtual machine and then describes configuration and tuning options.

Chapter 6: Installing Linux on a Virtual Machine Provides a guide for loading Ubuntu Linux in a virtual machine and then walks through a number of configuration and optimization options.

Chapter 7: Managing CPUs for a Virtual Machine Discusses how CPU resources are virtualized and then describes various tuning options and optimizations. Included topics are hyper-threading and Intel versus AMD.

Chapter 8: Managing Memory for a Virtual Machine Covers how memory is managed in a virtual environment and the configuration options available. It concludes with a discussion of various memory optimization technologies that are available and how they work.

Chapter 9: Managing Storage for a Virtual Machine Examines how virtual machines access storage arrays and the different connection options they can utilize. Included are virtual machine storage options and storage optimization technologies such as deduplication.

Chapter 10: Managing Networking for a Virtual Machine Begins with a discussion of virtual networking and how virtual machines use virtual switches to communicate with each other and the outside world. It concludes with virtual network configuration options and optimization practices.

Chapter 11: Copying a Virtual Machine Discusses how virtual machines are backed up and provisioned through techniques such as cloning and using templates. It finishes with a powerful feature called snapshots that can preserve a virtual machine state.

Chapter 12: Managing Additional Devices in Virtual Machines Begins by discussing virtual machine tools, vendor-provided application packages that optimize a virtual machine's performance, and concludes with individual discussions of virtual support for other peripheral devices like CD/DVD drives and USB devices.

Chapter 13: Understanding Availability Positions the importance of availability in the virtual environment and then discusses various availability technologies that protect individual virtual machines, virtualization servers, and entire data centers from planned and unplanned downtime.

Chapter 14: Understanding Applications in a Virtual Machine Focuses on the methodology and practices for deploying applications in a virtual environment. Topics include application performance, using resource pools, and deploying virtual appliances.

Appendix: Answers to Additional Exercises Contains all of the answers to the additional exercises found at the end of every chapter.

Glossary Lists the most commonly used terms throughout the book.

How to Contact the Author

I welcome feedback from you about this book or about books you'd like to see from me in the future. You can reach me by writing to mportnoyvm@gmail.com.

Sybex strives to keep you supplied with the latest tools and information you need for your work. Please check their website at www.wiley.com/go/virtualizationess2e, where we'll post additional content and updates that supplement this book if the need arises.

Understanding Virtualization

We are in the midst of a substantial change in the way computing services are provided. As a consumer, you surf the Web on your cell phone, get directions from a GPS device, and stream movies and music from the cloud. At the heart of these services is *virtualization*—the ability to abstract a physical server into a virtual machine.

In this chapter, you will explore some of the basic concepts of virtualization, review how the need for virtualization came about, and learn why virtualization is a key building block to the future of computing.

- ▶ **Describing virtualization**
- ▶ **Understanding the importance of virtualization**
- ▶ **Understanding virtualization software operation**

Describing Virtualization

Over the last 50 years, certain key trends created fundamental changes in how computing services are provided. Mainframe processing drove the sixties and seventies. Personal computers, the digitization of the physical desktop, and client/server technology headlined the eighties and nineties. The Internet, boom and bubble, spanned the last and current centuries and continues today. We are, though, in the midst of another of those model-changing trends: virtualization.

Virtualization is a disruptive technology, shattering the status quo of how physical computers are handled, services are delivered, and budgets are allocated. To understand why virtualization has had such a profound effect on today's computing environment, you need to have a better understanding of what has gone on in the past.

The word *virtual* has undergone a change in recent years. Not the word itself, of course, but its usage has been expanded in conjunction with the expansion of computing, especially with the widespread use of the Internet and smart phones. Online applications have allowed us to shop in virtual stores, examine potential vacation spots through virtual tours, and even

Some examples of virtual reality in popular culture are the file retrieval interface in Michael Crichton's *Disclosure*, *The Matrix*, *Tron*, and *Star Trek: The Next Generation's* holodeck.

keep our virtual books in virtual libraries. Many people invest considerable time and actual dollars as they explore and adventure through entire worlds that exist only in someone's imagination and on a gaming server.

Virtualization in computing often refers to the abstraction of some physical component into a logical object. By virtualizing an object, you can obtain some greater measure of utility from the resource the object provides. For example, virtual LANs (local area networks), or VLANs, provide greater network performance and improved manageability by being separated from the physical hardware. Likewise, storage area networks (SANs) provide greater flexibility, improved availability, and more efficient use of storage resources by abstracting the physical devices into logical objects that can be quickly and easily manipulated. Our focus, however, will be on the virtualization of entire computers.

If you are not yet familiar with the idea of computer virtualization, your initial thoughts might be along the lines of *virtual reality*—the technology that, through the use of sophisticated visual projection and sensory feedback, can give a person the experience of actually being in that created environment. At a fundamental level, this is exactly what computer virtualization is all about: it is how a computer application experiences its created environment.

The first mainstream virtualization was done on IBM mainframes in the 1960s, but Gerald J. Popek and Robert P. Goldberg codified the framework that describes the requirements for a computer system to support virtualization. Their 1974 article “Formal Requirements for Virtualizable Third Generation Architectures” describes the roles and properties of virtual machines and virtual machine monitors that we still use today. The article is available for purchase or rent at <http://dl.acm.org/citation.cfm?doi=361011.361073>. By their definition, a virtual machine (VM) can virtualize all of the hardware resources, including processors, memory, storage, and network connectivity. A virtual machine monitor (VMM), which today is commonly called a *hypervisor*, is the software that provides the environment in which the VMs operate. Figure 1.1 shows a simple illustration of a VMM.

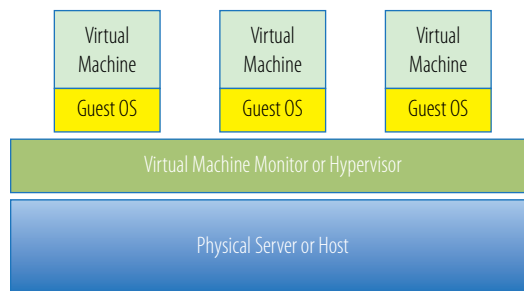


FIGURE 1.1 A basic virtual machine monitor (VMM)

According to Popek and Goldberg, a VMM needs to exhibit three properties in order to correctly satisfy their definition:

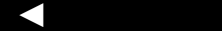
Fidelity The environment it creates for the VM is essentially identical to the original (hardware) physical machine.

Isolation or Safety The VMM must have complete control of the system resources.

Performance There should be little or no difference in performance between the VM and a physical equivalent.

Because most VMMs have the first two properties, VMMs that also meet the final criterion are considered *efficient* VMMs. We will go into these properties in much more depth as we examine hypervisors in Chapter 2, “Understanding Hypervisors,” and virtual machines in Chapter 3, “Understanding Virtual Machines.”

Let’s go back to the virtual reality analogy. Why would you want to give a computer program a virtual world to work in, anyway? It turns out that it was very necessary. To help explain that necessity, let’s review a little history. It would be outside the scope of this text to cover all the details about how server-based computing evolved, but for our purposes, we can compress it to a number of key occurrences.



Between the late 1970s and mid-1980s, there were more than 70 different personal computer operating systems.

Microsoft Windows Drives Server Growth

Microsoft Windows was developed during the 1980s primarily as a personal computer operating system. Others existed, CP/M and OS/2 for example, but as you know Windows eventually dominated the market and today it is still the primary operating system deployed on PCs. During that same time frame, businesses were depending more and more on computers for their operations. Companies moved from paper-based records to running their accounting, human resources, and many other industry-specific and custom-built applications on mainframes or minicomputers. These computers usually ran vendor-specific operating systems, making it difficult, if not impossible, for companies and IT professionals to easily transfer information among incompatible systems. This led to the need for *standards*, agreed upon methods for exchanging information, but also the idea that the same, or similar, operating systems and programs should be able to run on many different vendors’ hardware. The first of these was Bell Laboratories’ commercially available UNIX operating systems.

Companies had both Windows-based PCs and other operating systems in-house, managed and maintained by their IT staffs, but it wasn't cost effective to train IT staffs on multiple platforms. With increasing amounts of memory, faster processors, and larger and faster storage subsystems, the hardware that Windows could run on became capable of hosting more powerful applications that had in the past primarily run on minicomputers and mainframes. These applications were being migrated to, or being designed to run on, Windows servers. This worked well for companies because they already had Windows expertise in house and no longer required multiple teams to support their IT infrastructure. This move, however, also led to a number of challenges. Because Windows was originally designed to be a single-user operating system, a single application on a single Windows server ran fine, but often when a second program was introduced, the requirements of each program caused various types of resource contention and even out and out operating system failures. This behavior drove many companies, application designers, developers, IT professionals, and vendors to adopt a "one server, one application" best practice; so for every application that was deployed, one or more servers needed to be acquired, provisioned, and managed.

Current versions of Microsoft Windows run concurrent applications much more efficiently than their predecessors.

Another factor that drove the growing server population was corporate politics. The various organizations within a single company did not want any common infrastructure. Human Resource and Payroll departments declared their data was too sensitive to allow the potential of another group using their systems. Marketing, Finance, and Sales all believed the same thing to protect their fiscal information. Research and Development also had dedicated servers to ensure the safety of their corporate intellectual property. Sometimes companies had redundant applications, four or more email systems, maybe from different vendors, due to this proprietary ownership attitude. By demanding solitary control of their application infrastructure, departments felt that they could control their data, but this type of control also increased their capital costs.

Aiding the effects of these politics was the fact that business demand, competition, Moore's Law, and improvements in server and storage technologies all drastically drove down the cost of hardware. This made the entry point for a department to build and manage its own IT infrastructure much more affordable. The processing power and storage that in the past had cost hundreds of thousands of dollars could be had for a fraction of that cost in the form of even more Windows servers.

Business computers initially had specialized rooms in which to operate. These computer rooms were anything from oversized closets to specially constructed areas for housing a company's technology infrastructure. They typically had raised floors under which the cables and sometimes air conditioning conduits were run. They held the computers, network equipment, and often telecomm equipment. They needed to be outfitted with enough power to service all of that equipment. Because all of those electronics in a contained space generated considerable heat, commensurate cooling through huge air-conditioning handlers was mandatory as well. Cables to interconnect all of these devices, fire-suppression systems in case of emergency, and separate security systems to protect the room itself, all added to the considerable and ever-rising costs of doing business in a modern corporation. As companies depended more and more on technology to drive their business, they added many more servers to support that need. Eventually, this expansion created data centers. A *data center* could be anything from a larger computer room, to an entire floor in a building, to a separate building constructed and dedicated to the health and well-being of a company's computing infrastructure. Entire buildings existed solely to support servers, and then at the end of twentieth century, the Internet blossomed into existence.

"E-business or out of business" was the cry that went up as businesses tried to stake out their territories in this new online world. To keep up with their competition, existing companies deployed even more servers as they web-enabled old applications to be more customer facing and customer serving. Innovative companies, such as Amazon and Google, appeared from nowhere, creating disruptive business models that depended on large farms of servers to rapidly deliver millions of web pages populated with petabytes of information (see Table 1.1). IT infrastructure was mushrooming at an alarming rate, and it was only going to get worse. New consumer-based services were delivered not just through traditional online channels, but newer devices such as mobile phones compounded data centers' growth. Between 2000 and 2006, the Environmental Protection Agency (EPA) reported that energy use by United States data centers doubled, and that over the next five years they expected it to double again. Not only that, but servers were consuming about 2 percent of the total electricity produced in the country, and the energy used to cool them consumed about the same amount. Recent studies show that energy use by data centers continues to increase with no sign of decreasing any time soon.

TABLE 1.1 Byte Sizes

| Name | Abbreviation | Size |
|-----------|--------------|-----------------------------|
| Byte | B | 8-bits (a single character) |
| Kilobyte | KB | 1,024 B |
| Megabyte | MB | 1,024 KB |
| Gigabyte | GB | 1,024 MB |
| Terabyte | TB | 1,024 GB |
| Petabyte | PB | 1,024 TB |
| Exabyte | EB | 1,024 PB |
| Zettabyte | ZB | 1,024 EB |
| Yottabyte | YB | 1,024 ZB |

Let's take a closer look at these data centers. Many were reaching their physical limits on many levels. They were running out of actual square footage for the servers they needed to contain, and companies were searching for alternatives. Often the building that housed a data center could not get more electrical power or additional cooling capacity. Building larger or additional data centers was and still is an expensive proposition. In addition to running out of room, the data centers often had grown faster than the people managing them could maintain them. It was common to hear tales of lost servers. (*A lost server* is a server that is running, but no one actually knows which line of business owns it or what it is doing.) These lost servers couldn't be interrupted for fear of inadvertently disrupting some crucial part of the business. In some data centers, cabling was so thick and intertwined that when nonfunctioning cables needed to be replaced, or old cables were no longer needed, it was easier to just leave them where they were, rather than try to unthread them from the mass. Of course, these are the more extreme examples, but most data centers had challenges to some degree in one or more of these areas.

Explaining Moore's Law

So far you have seen how a combination of events—the rise of Windows, corporations increasing their reliance on server technology, and the appearance and mushrooming of the Internet and other content-driven channels—all contributed