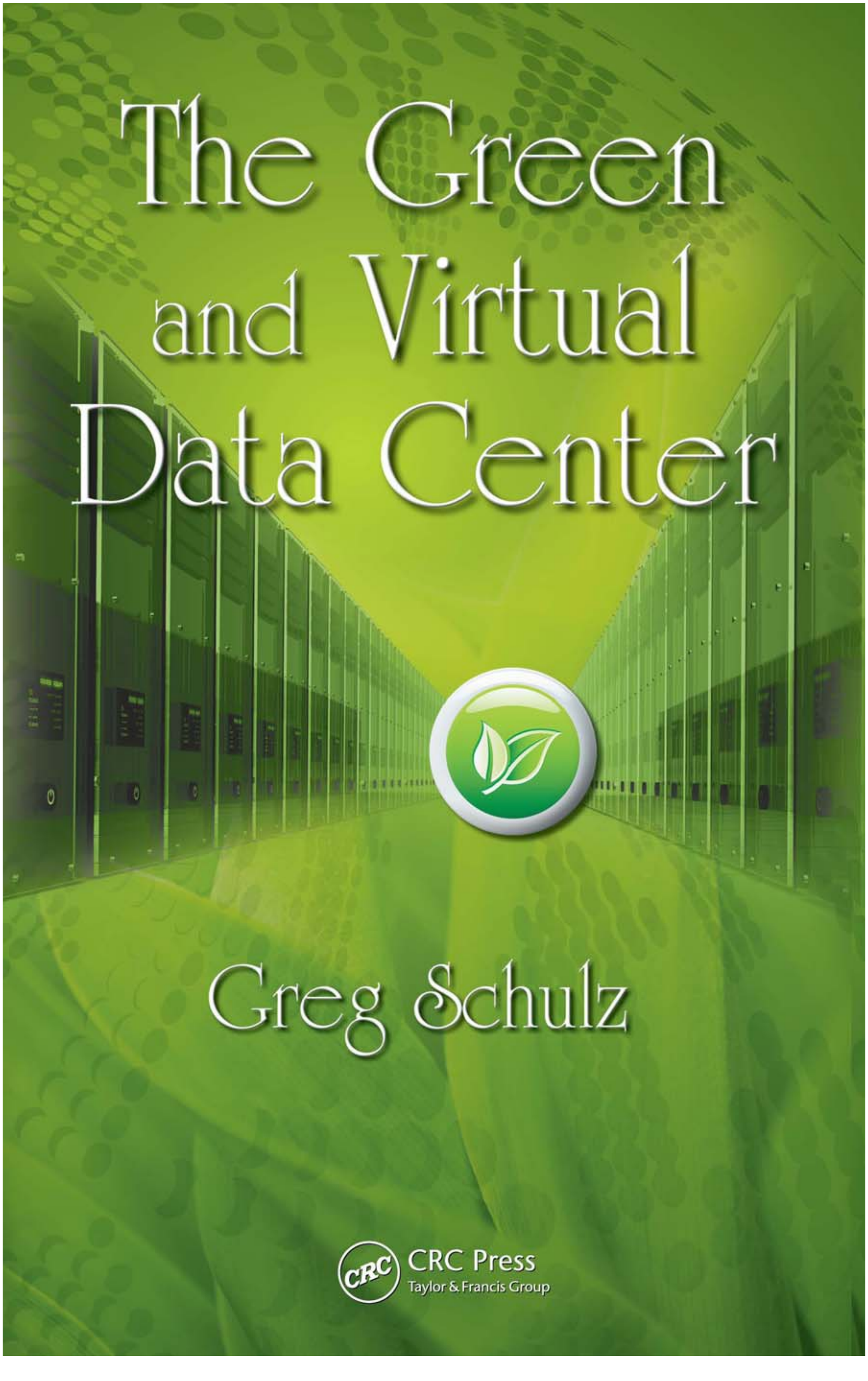
The Green and Virtual Data Center
Greg Schulz
CRC Press
Taylor & Francis Group

Auerbach Publications
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

Printed in the United States of America on acid-free paper
10 9 8 7 6 5 4 3 2 1

International Standard Book Number-13: 978-1-4200-8666-9 (Hardcover)

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the Auerbach Web site at
http://www.auerbach-publications.com

Contents

Chapter 3

What Defines a Next-Generation and Virtual Data Center?

Virtual data centers enable data mobility, resiliency, and improved IT efficiency.

In this chapter you will learn:

- What defines a green and virtual data center
- How virtualization can be applied to servers, storage, and networks
- The many faces of server, storage, and networking virtualization
- How to leverage virtualization beyond server, storage, and network consolidation
- The various components and capabilities that comprise a virtual data center
- How existing data centers can transform into next-generation virtual data centers

Many approaches and technologies, addressing different issues and requirements, can be used to enable a green and virtual data center. Virtualization is a popular approach to consolidating underutilized IT resources, including servers, storage, and input/output (I/O) networks to free up floor-space, lower energy consumption, and reduce cooling demand, all of which can result in cost savings. However, virtualization—and particularly consolidation—applies to only a small percentage of all IT resources. The importance of this chapter is that there are many facets of virtualization that can be used to enable IT infrastructure resource management to improve service delivery in a more cost-effective and environmentally friendly manner.

In Chapters 1 and 2 we discussed various green and environmental issues. In this chapter we turn to what can be done and how to leverage different technologies to enable a green virtual data center. Here we look at what defines a green and virtual data center, building on the foundation laid out in Part I.

This chapter also looks at how virtualization can be applied to servers, storage, and networks. Virtualization can be applied to consolidate underutilized IT resources. It can also be used to support transparent management, enabling scaling for high growth and performance applications in conjunction with clustering and other technologies.

3.1 Why Virtualize a Data Center?

A virtual data center can, and should, be thought of as an information factory that needs to run 24-7, 365 days a year, to delivery a sustained stream of useful information. For the information factory to operate efficiently, it needs to be taken care of and seen as a key corporate asset. Seen as an asset, the IT factory can be invested in to maintain and enhance productivity and efficiency, rather than being considered a cost center or liability.

The primary focus of enabling virtualization technologies across different IT resources is to boost overall effectiveness while improving application service delivery (performance, availability, responsiveness, security) to sustain business growth in an economic and environmentally friendly manner. That is, most organizations do not have the luxury, time, or budget to deploy virtualization or other green-related technologies and techniques simply for environmental reasons—there has to be a business case or justification.

Virtual data centers (Figure 3.1), regardless of whether new or existing; require physical resources including servers, storage, and networking, and facilities to support a diverse and growing set of application capabilities while sustaining business growth. In addition to sustaining business growth, applications need to be continually enhanced to accommodate changing business rules and enhance service delivery. Application enhancements include ease of use, user interfaces, rich media (graphics and video, audio and intuitive help), along with capturing, storing, and processing more data.

There are a growing number of business cases and justifications for adopting green technologies that reduce costs or maximize use of existing resources while also benefiting the environment. However, it is rare for a business to have surplus budget dollars, personnel, facilities, and management support to deploy new technologies simply for the sake of deploying them.

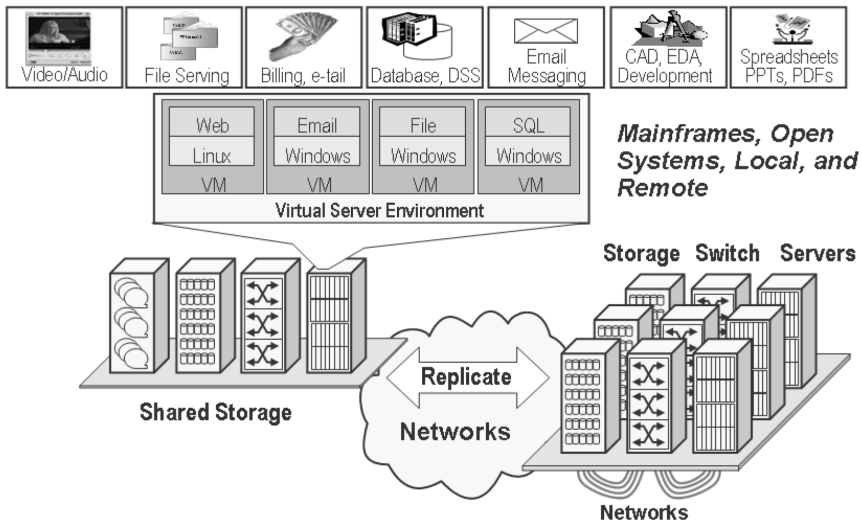

Figure 3.1 Virtual Data Center with Physical Resources

Challenges facing IT data centers and businesses of all sizes include:

- Ensuring reliability, availability, serviceability, and management
- Dealing with performance bottlenecks and changing workloads
- Securing applications and data from diverse internal and external threat risks
- Enabling new application features and extensibility while supporting business growth
- Managing more data for longer periods of time
- Supporting compliance or self-governance while securing intellectual property data
- Maximizing **power, cooling, floor space, and environmental (PCFE)** issues
- Ensuring data and transaction integrity for business continuance and disaster recovery
- Consolidating IT resources to reduce complexity and contain management costs

- Supporting interoperability with existing applications, servers, storage, and networks
- Reducing or abstracting complexities associated with applications interdependencies
- Leveraging existing IT personal skill sets and experience
- Scaling existing and new applications with stability and in cost-effectively
- Moving data transparently from old to new storage during technology replacement
- Enhancing IT service delivery in a cost-effective and environmentally friendly manner

IT data centers are increasingly looking to virtualization technology and techniques to address all these issues and activities. Virtualization can be used to consolidate servers and storage resources in many environments to boost resource utilization and contain costs—for example, using a software based **virtual machine (VM)** to combine applications and operating systems images from underutilized physical servers to virtual machines on a single server. With a growing focus on PCFE and the greening of IT in general, consolidating servers, as well as storage, networking, and facilities, is a popular and easy-to-understand use of virtualization.

Energy efficiency today can sometimes mean simply energy avoidance. In the future, however, emphasis will shift to doing and enabling more work and storing increasing amounts of information for longer periods of time. This will have to be accomplished while consuming less energy and using less floor space. Doing more work in a more productive manner with less energy will result in efficiencies from improved technologies, techniques and best practices.

Consequently, a green and virtual data center should be much more than just an environment that leverages some virtualization for consolidation purposes. A green and virtual data center should enable transparent management of different physical resources to support flexible IT service delivery in an environmentally and economically friendly manner on both local and remote bases.

Scaling with stability means that as performance is increased, application availability or capacity is neither decreased nor is additional management complexity or cost introduced. Similarly, scaling with stability means that as capacity is increased, neither performance nor availability suffer, nor is performance negatively affected by growth or increased workload or application functionality. This also includes eliminating single points of failure and supporting fault isolation and containment, self-healing, supporting mixed performance of small and large I/O operations, and additional functionality or intelligence in technology solutions without adding cost or complexity. In addition, scaling with stability means not introducing downtime or loss of availability or negative performance as a result of growth. High resilience or self-healing with fault isolation and containment will prevent single points of failure from cascading into rolling disasters.

Flexibility and agility will enable virtual data centers to meet changing business and application requirements while quickly and transparently adopting new technology enhancements without disruption.

Tenets of a green and virtualized data center include:

- Flexible, scalable, stable, agile, and highly resilient or self-healing systems
- Quick adaptation and leverage of technology improvements
- Transparency of applications and data from physical resources
- Efficient operation without loss of performance or increased cost complexity
- Environmentally friendly and energy efficient yet economical to maintain
- Highly automated and seen as information factories as opposed to cost centers
- Measureable with metrics and reporting to gauge relative effectiveness
- Secure from various threat risks without impeding productivity

3.2 Virtualization Beyond Consolidation—Enabling Transparency

There are many facets of virtualization. Aggregation is a popular approach to consolidate underutilized IT resources including servers, storage, and networks. The benefits of consolidation include improved efficiency by eliminating underutilized servers or storage to reduce electrical power, cooling, and floor space requirements as well as management time, or to reuse and repurpose servers that have become surplus to enable growth or support new application capabilities.

Another form of virtualization is emulation or transparency providing abstraction to support integration and interoperability with new technologies while preserving existing technology investments and not disrupting software procedures and policies. Virtual tape libraries are a commonly deployed example of storage technology that combines emulation of existing tape drives and tape libraries with disk-based technologies. The value proposition of virtual tape and disk libraries is to coexist with existing backup software and procedures while enabling new technology to be introduced.

Figure 3.2 shows two examples of virtualization being used, with consolidation on the left side and transparency for emulation and abstraction to support scaling on the right. On the consolidation side, the operating systems and applications of multiple underutilized physical servers are consolidated onto a single or, for redundancy, multiple servers in a virtual environment with a separate virtual machine emulating a physical machine. In this example, each of the operating systems and applications that were previously running on their own dedicated servers now run on a virtual server. Consolidation enables multiple underutilized servers to be combined yet let each system think and operate as though it still had its own server.

For a variety of reasons, not all servers or other IT resources lend themselves to consolidation. These reasons may include performance, politics, finances, service-level, or security issues. For example, an application may need to run on a server with low CPU utilization to meet performance and response-time objectives or to support seasonal workload adjustments. Also, certain applications, data, or even users of servers may need to be isolated from each other for security and privacy reasons.

Politics and financial, legal, or regulatory requirements also need to be considered. For example, a server and application may be "owned" by

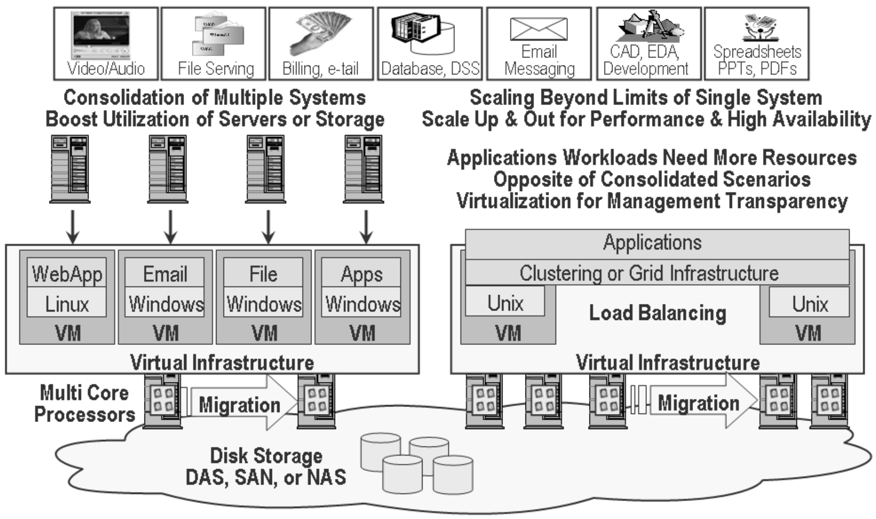

Figure 3.2 Consolidation vs. Transparent Management with Virtualization

different departments or groups and thus managed and maintained separately. Regulatory or legal requirements may dictate that certain systems be kept separate from other general-purpose or mainstream applications, servers, and storage. Separation of applications may also be necessary isolate development, test, quality assurance, back-office, or other functions from production or online applications and systems, as well as to support business continuance, disaster recovery, and security.

For applications and data that do not lend themselves to consolidation, a different way to use virtualization is to enable transparency of physical resources to support interoperability and coexistence between new and existing software tools, servers, storage, and networking technologies, such as enabling new, more energy-efficient servers or storage with improved performance to coexist with existing resources and applications.

On the right side in Figure 3.2 are examples of applications that don't lend themselves to consolidation because they need to be isolated from other applications or clients for performance or other reasons. However, these applications can still benefit from transparency and abstraction when combined with clustering technology to enable scaling beyond the limits of a single server, storage, or networking device. Included on the right side of Figure 3.2 are applications that need more resources for performance or availability than are available from a single large server and hence need to scale horizontally (also known as scale up and out).

Another use of virtualization transparency is to enable new technologies to be moved into and out of running or active production environments to facilitate technology upgrades and replacements. Still another use is to adjust physical resources to changing application demands, such as seasonal planned or unplanned workload increases. Transparency via virtualization also enables routine planned and unplanned maintenance functions to be performed on IT resources without disrupting applications and users of IT services.

Virtualization in the form of transparency or abstraction of physical resources to applications can also be used to help achieve energy savings and address other green issues by enabling newer, more efficient technologies to be adopted more quickly. Transparency can also be used to implement tiered servers and storage to leverage the right technology and resource for the task at hand as of a particular point in time.

Business continuity (BC) and **disaster recovery (DR)** are other areas in which transparency via virtualization can be applied to in a timely and cost-efficient manner in-house, via a managed service provider, or using some combination. For example, traditionally, a BC or DR plan requires the availability of similar server hardware at a secondary site. Some challenges with this kind of redundancy are that the service and servers must be available to those who need them when they need them. For planned testing, this may not be a problem; however, in the event of a real disaster, a first-come, first-served situation could arise, with too many subscribers to the same finite set of physical servers, storage, and networking facilities.

If dedicated and guaranteed servers and storage resources are available for BC and DR, competition for resources is eliminated. This means, however, that additional servers and storage need to be powered, cooled, and given floor space and management. In addition, these operating systems and applications may require identical or very similar hardware and configurations.

3.3 Components of a Virtual Data Center

For some organizations, the opportunity to start from scratch with a new green data center may exist. However, for most, enabling a virtualized data center relies on transforming existing facilities, servers, storage, networking, and software tools along with processes and procedures to adopt

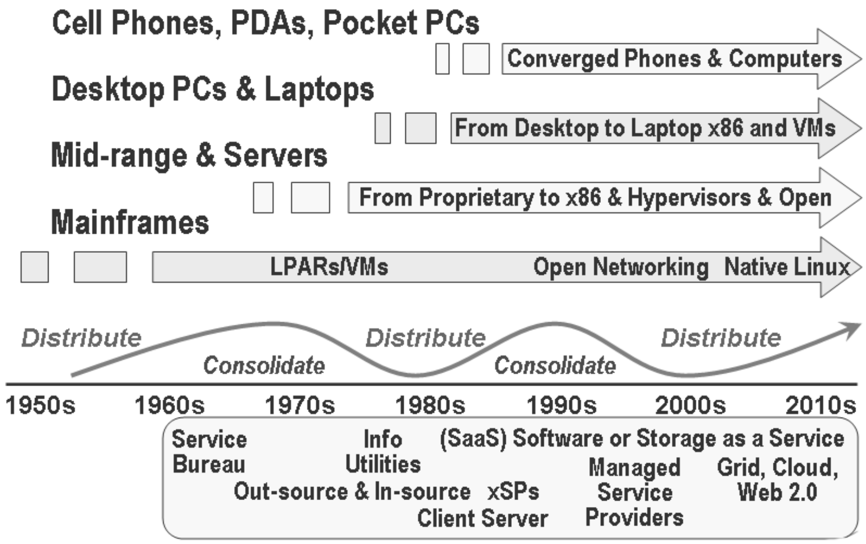

Figure 3.3 IT Resource Computing and Consolidation Continuum

virtualization technologies and techniques. Thus, virtualization technologies should enable existing IT resources to be transformed and transition to a next-generation virtualization data center environment. The benefit is support of growth and enhancement of service delivery in a cost-effective and timely manner.

A green and virtual data center is more than an environment that leverages server, storage, or network virtualization to improve resource usage. IT resource consolidation has been a recurring theme over the past several decades with the shift from centralized to distributed computing. The cycle shown in Figure 3.3 went from distributed resources to consolidation, followed by client–server systems, followed by reconsolidation, followed by Internet dispersion of resources to the current consolidation phase.

Some reasons for the cycle of distribute and consolidate include changing business and IT models, technology trends, and financing models. Another factor is the decades-old issue of addressing the server-to-storage I/O performance gap, where the cost of hardware continues to decrease and servers becoming faster and smaller. Meanwhile, storage capacity and availability continue to increase while physical footprint and price decrease. However, there is a gap between storage capacity, server, and storage performance (see Figure 3.4). The result is that for some

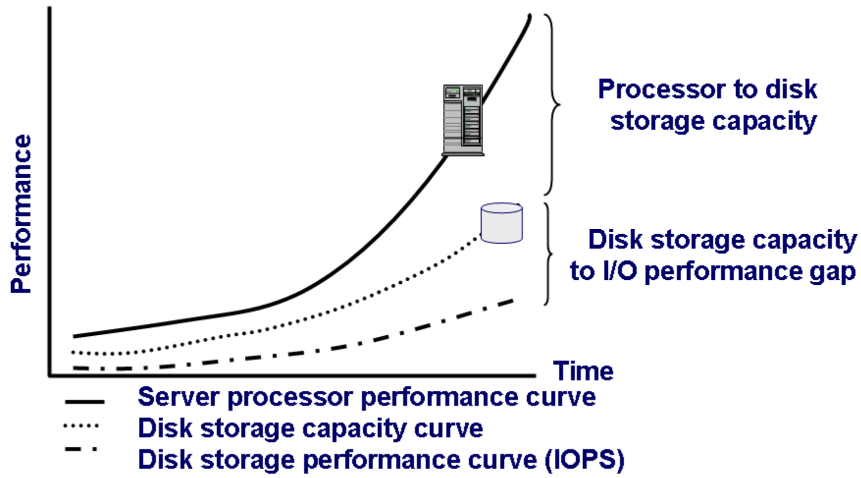

Figure 3.4 Server and Storage I/O performance Gap

applications, to achieve a given level of performance, more resources (disks, controllers, adapters, and processors) are needed, resulting in a surplus of storage capacity.

In an attempt to reduce the excess storage capacity, consolidation sometimes happens without an eye on performance, looking only at the floor space, power, and cooling benefits of highly utilized storage. Then, to address storage performance bottlenecks, the storage gets reallocated across more storage systems, and the cycle starts again. Another driver for underutilized hardware is a result of servers and storage being bought by individual departments or for specific applications, with politics and financial constraints limiting their shared usage.

Another variation is the notion that hardware is inexpensive, so buy more. With higher energy prices, simply throwing more hardware at application performance or other issues is no longer an option for environments with PCFE constraints.

An opportunity enabled by virtualization transparency and abstraction is to address performance and related issues with tiered servers, storage, and networks using high-performance, energy-efficient devices balanced with high-capacity energy-efficient devices. For example, instead of using several low-cost disk drives and associated adapters for performance-intensive applications, use faster storage measured on an activity-per-watt (IOPS/

watt, bandwidth/watt, transaction or message/watt) basis instead of on a cost-per-gigabyte or energy-per-gigabyte basis.

3.3.1 Infrastructure Resource Management Software Tools

Falling under the umbrella of **infrastructure resource management (IRM)** are various activities, tools, and processes for managing IT resources across different technology domains (servers, storage, networks, facilities, and software) with diverse interdependencies to enable IT application service delivery.

Aspects of IRM include:

- Logical and physical security, including rights management and encryption
- Asset management, including configuration management databases
- Change control and management, along with configuration validation management tools
- Data protection management, including business continuity and disaster recovery
- Performance and capacity planning and management tools
- Data search and classification tools for structured and unstructured data
- High-availability and automated self-healing infrastructures
- Data footprint reduction, including archiving, compression, and de-duplication
- Planning and analysis, event correlation, and diagnostics
- Provisioning and allocation of resources across technology domains
- Policies and procedures, including best practices and usage template models

3.3.2 Measurements and Management Insight

Various metrics and measurements are needed in order to provide insight into how data centers and applications are running as well as using resources. Metrics and measurements are also important for timely and proactive problem resolution and isolation, as well as event correlation to support planning and reconfiguration for improved service delivery and growth. Metrics and management insight are also needed to ensure compliance and other requirements are being met, including security or activity logs, as well as that data is being protected as it is intended and required to be.

Examples of metrics, measurements, and reporting include:

- Energy consumption and effectiveness of work being performed
- Server, storage, and network performance and capacity usage information
- Availability of IT resources, including planned and unplanned downtime
- Effectiveness of IT resources to meet application service-level objectives
- Data protection management status and activity
- Error, activity and events logs, data protection status, and alarms
- Metrics for recycling, carbon disclosure, and environmental health and safety reporting

Measurements and monitoring of IT resources are key to achieving increased efficiency so that the right decisions can be made for the right reasons while addressing and fixing problems rather than simply moving them around. For example, if applications require overallocation of server, storage, and networking resources to meet performance and application service objectives, consider options such as leveraging faster technologies that consume less power to accomplish the necessary work.

3.3.3 Facilities and Habitats for Technology

One potentially confusing aspect of next-generation data centers is the implication that they must be built from scratch, as new facilities with all new technology, IT equipment, and software. For some environments, that may be the case. For most environments, however, even if a new physical facility is being built or an existing one expanded or remodeled, integration with existing technologies and management tools is required. Consequently, the road to a virtual or next-generation and green data center is an evolution from a current environment to a new and enhanced way of operating and managing IT resources in an efficient and flexible manner. This means, among other things, that existing legacy mainframes may need to coexist with current-generation blade servers or other modern servers. Similarly, magnetic tape devices may need to coexist with newer disk-based systems, and LANs may need to coexist with networks using copper, optical, and even wireless communications.

All of these IT resources need to be housed in a technology-friendly environment that is or will become more energy efficient, including in how cooling is handled as well as primary power distribution and provisions for standby power. (Chapter 6 looks more closely at various options for improving the energy efficiency of facilities.)

3.3.4 Tiered Servers and Software

Servers have received a lot of attention as prime consumers of electrical power and producers of heat. Consequently, virtualization in the form of server consolidation to combine multiple lower-utilized servers onto a single or fewer physical servers running virtual machines is a popular topic. Having fewer servers means that less electrical power is required for both the servers and the necessary cooling.

Many servers are underutilized at various times, and some are always underutilized as a result of how or when they were acquired and deployed. Through much of the late 1990s, the notion was that hardware was inexpensive, so the easiest solution seemed to be to throw hardware at various challenges as they arose. Or, as new applications came online, it may have been faster to acquire a new server than to try and find space on another server.

Given today's rising energy costs, concerns about PCFE issues, and the need to boost IT resource efficiency, many smaller or underutilized servers are being consolidated onto either larger servers or blade servers running virtual infrastructure software and virtual machines such as those from VMware, Virtual Iron, or Microsoft, among others. Not all servers and applications, however, lend themselves to consolidation, for the reasons we have discussed. As a result, some applications need to scale beyond the limits of a single device, where virtualization, as discussed earlier, enables transparency for maintenance, upgrades, load balance, and other activities.

3.3.5 Tiered Storage and Storage Management

There are many different ways of implementing storage virtualization, including solutions that aggregate heterogeneous or different vendors' storage to enable pooling of resources for consolidated management. Although it is popular to talk about, storage aggregation has trailed in actual customer deployments to other forms of virtualization such as emulation.

Virtual tape libraries (VTLs) or virtual tape systems (VTS) or disk libraries and de-duplication appliances that emulate the functionality of previous-generation storage solutions are the most common forms of storage virtualization in use today. The benefit of emulation is that it enables abstraction and transparency as well as interoperability between old processes, procedures, software, or hardware and newer, perhaps more energy-efficient or performance-enhancing, technologies.

The next wave of storage virtualization looks to be in step with the next server virtualization wave—virtualization not for consolidation or emulation but to support transparent movement of data and applications over tiered storage for both routine and non-routine IRM functions. These systems will facilitate faster and less disruptive technology upgrades and expansion so that storage resources can be used more effectively and transparently. Moving forward, there should be a blurring of the lines between transparency and abstraction vs. consolidation and aggregation of resources.

Another form of virtualization is partitioning or isolation of consolidated and shared resources. For example, some storage devices enable logical unit numbers (LUNs) to be mapped into partitions or groups to abstract LUN and volume mapping for coexistence with different servers. Another variation enables a storage server to be divided up into multiple logical virtual filers to isolate data from different applications and customers.

3.3.6 Tiered Networks and I/O Virtualization

There are many different types of networks, and convergence may include virtual connect infrastructures inside blade center servers, top-of-rack and end-of-rack solutions, modular switches and routers, as well as core and backbone directors for traditional networks as well as converged virtual I/O and I/O virtualization networks. There are also many kinds of networks and storage interfaces for connecting physical and virtual servers and storage.

3.3.7 Virtual Offices, Desktops, and Workstations

Another component of a virtual data center environment is the remote and virtual or home office. Desktop and workstation virtualization is a natural extension of what is taking place with servers, storage, and networks to boost utilization and effectiveness as well as to address complexities in configuring and deploying large numbers of workstations and desktops while enabling virtual offices to access and use data when and where needed in a secure and flexible manner.

3.4 Summary

Green and next-generation virtual data centers should be highly efficient, flexible, resilient, and environmentally friendly while economical to operate. Current focus is on virtualization from a consolidation perspective, but in the future there will be even more opportunities for IT environments to adapt their processes, techniques, and technologies to sustain business growth and enhance application service delivery experience while reducing costs without compromising performance, availability, or ability to store and process more information. There are many aspects of data storage virtualization that address routine IT management and support tasks, including data protection, maintenance, and load-balancing for seasonal and transient project-oriented application workloads.

The vendor who controls and manages the virtualization software, whether on a server, storage, or in the network, controls the vendor lock-in or "stickiness." If you are looking to virtualization to eliminate vendor lock-in, it is important to make sure you understand what lock-in is being left and what lock-in will be in its place. Vendor lock-in is not a bad thing if the capabilities, efficiency, economics, and stability offered by a solution outweigh any real or perceived risks or issues. Bad technology and tools in the wrong hands for the wrong tasks make for a bad solution, while good technology and tools in the wrong hands for the wrong tasks make for a not-so-good solution. The goal is to put good tools and techniques in the right hands for the right tasks to make an enabling and good solution.

The idea is to leverage virtualization technologies in the form of abstraction and transparency or emulation combined with tiered servers, tiered storage, and tiered networks to align the right technology to the task at hand at a particular time. Start to fix the problems instead of moving them around or bouncing from distributed to consolidated and moving the distributed problems back to a main site. You can get management of an increasing amount of data and resources under control, and you can do more work with less energy while supporting growth.