# Private Clouds: Selecting the Right Hardware for a Scalable Virtual Infrastructure

Greg Shields

sponsored by

Realtime publishers

# Introduction to Realtime Publishers

**by Don Jones, Series Editor**

For several years now, Realtime has produced dozens and dozens of high-quality books that just happen to be delivered in electronic format—at no cost to you, the reader. We've made this unique publishing model work through the generous support and cooperation of our sponsors, who agree to bear each book's production expenses for the benefit of our readers.

Although we've always offered our publications to you for free, don't think for a moment that quality is anything less than our top priority. My job is to make sure that our books are as good as—and in most cases better than—any printed book that would cost you $40 or more. Our electronic publishing model offers several advantages over printed books: You receive chapters literally as fast as our authors produce them (hence the "realtime" aspect of our model), and we can update chapters to reflect the latest changes in technology.

I want to point out that our books are by no means paid advertisements or white papers. We're an independent publishing company, and an important aspect of my job is to make sure that our authors are free to voice their expertise and opinions without reservation or restriction. We maintain complete editorial control of our publications, and I'm proud that we've produced so many quality books over the past years.

I want to extend an invitation to visit us at http://nexus.realtimepublishers.com, especially if you've received this publication from a friend or colleague. We have a wide variety of additional books on a range of topics, and you're sure to find something that's of interest to you—and it won't cost you a thing. We hope you'll continue to come to Realtime for your educational needs far into the future.

Until then, enjoy.

Don Jones

Introduction to Realtime Publishers ..... i

Chapter 1: Deconstructing the Private Cloud: An Application Consumer's Perspective ..... 1

Constructing the Private Cloud, In Four Chapters ..... 2

IT's Primary Mission: Delivering Applications and Data ..... 3

Defining the Private Cloud Black Box ..... 5

Availability ..... 6

Flexibility ..... 7

Scalability ..... 9

Hardware Resource Optimization ..... 10

Resiliency ..... 11

Globalization ..... 12

What Is the Private Cloud User Experience? ..... 13

IT's Mission Is Inside the Box ..... 14

## Copyright Statement

# Chapter 1: Deconstructing the Private Cloud: An Application Consumer's Perspective

In 2009, a close personal friend and former editor of *Virtualization Review Magazine*, Keith Ward, was credited with stating that "virtualization is no longer cool."

"No longer cool?" you might be asking. Why this description of a technology that not that many years ago was dubbed "The Next Big Thing?" Virtualization promised to reduce our power and cooling requirements. It would buy back precious data center square footage through server consolidation. It would eliminate resource waste, enabling a new era of on-demand computing where every processor cycle and RAM kilobyte was discretely assigned to needy workloads.

Virtualization for many of us represented and continues to represent the embodiment of cool. But Ward argues that "lack of coolness" is in fact a good thing because *a cool technology* and *a mature technology that just simply works* are two descriptions often at odds. With this in mind, consider the rest of his thoughts:

> What I mean by that is that it's no longer brand new and something IT administrators are just learning about. They know what it is, they've used it, they're learning about different ways to leverage it. The products are maturing, and we're entering the next phase: Getting the most out of virtualization.
>
> It's no longer: "Whoa, look at what this can do!"
>
> Instead it's: "How does this transform our data center?"

Answering that question is the topic of this book. In its four chapters, you'll learn how businesses are only now seeing past the "coolness" that has been long associated with virtualization's technologies. Replacing virtualization's inner geek is a new focus on how it impacts overall IT service delivery, optimizing how services get delivered to their end consumer. One approach to accomplish this is the new phenomenon called *private clouds*, which create a virtual infrastructure seamlessly capable of scaling to meet business demands.

## Constructing the Private Cloud, In Four Chapters

Yet today's biggest problem with private clouds remains in their definition. What exactly is a private cloud? How do you get there? Once you've got one, what is the end result for the application consumer? And, finally, how can modular hardware make a cloud infrastructure dead simple for even the most complex of enterprise needs?

Getting to that private cloud end goal isn't easy. You first have to understand what it is and what it means for the business computing environment. In an era where the term "cloud" is overused, it is critically important to recognize what a private cloud actually is in comparison with all the other cloud options. You'll get a feeling for the answers to all of these questions in this guide's first chapter.

Once you understand what a private cloud is, you then need the technical details required to build one. A private cloud encompasses servers, storage, networking, and management, along with a host of supporting components; thus, *creating your own private cloud is no small task*. Chapter 2 discusses hardware—along with critical software pieces—that you'll need.

The problem is that your own private cloud isn't something that's best constructed out of individual pieces and parts. Remember back to IT's "old days" when you realized that the preconfigured servers bought from a major manufacturer gave you a superior experience than the "white boxes" you built yourself? Today's technologies are swiftly creating that same preconfigured virtual infrastructure for your private cloud.

Chapter 3 will confirm that virtualization might indeed not be cool any more. But as its cool edge wanes, it gains much-needed stability, modularity, and compatibility. Incorporating modular hardware into your private cloud enables you *to lean on the expertise of someone else* to ensure that you get it absolutely right from moment one.

Lastly, is the need for business value in even going down the private cloud road. Smart businesses don't invest in something unless they can see the value in doing so. *Private clouds by themselves don't guarantee a positive return*, in part due to their intrinsic complexity. They've got a lot of moving parts. Constructing and monitoring all those parts comprehensively isn't a task that's easily accomplished, particularly if you've never done it before. However, an effective set of private cloud–oriented tools that keeps tabs on the entire environment will get you the metrics you need. Chapter 4 shows you how to quantify that return while reaping the biggest possible benefit through the infrastructure convergence that a private cloud enables.

## IT's Primary Mission: Delivering Applications and Data

You really can't start this conversation by talking about the technologies that create a private cloud. That's because truly answering Ward's question—"How do we get this to work in our data center?"—first requires a look at how virtualization aligns with IT's core charter. If you look at the processes and tasks that define an IT service delivery infrastructure, how does virtualization improve the fulfillment of IT's mission? More specifically, when you look at the actions that go on inside the IT universe, which of them does the business care about and which are only important to IT itself. Understanding that breakdown of responsibility is key to recognizing a private cloud's true value.

So what is the primary mission of IT? Let me offer a simple statement that I believe defines IT's job at its core:

> IT is responsible for the assured and secured delivery of applications and data with an acceptable user experience.

A very direct statement, one that I think you'll agree with. But also one that too often gets mangled as businesses with good intentions try too hard to define what they want out of their own IT organizations. This mission statement identifies three responsibilities. Each responsibility is targeted towards the two core needs by the business: data and those applications that work with that data. Everything else in IT is just supporting activities.

Let's take a look at each of these three responsibilities in turn. In a minute, I'll show you why they're important as you consider what you want out of your own private cloud infrastructure.

- *Assured delivery.* IT's first responsibility is to ensure that applications and data are available to users when they need access. This means that users are guaranteed access to their client applications through every available delivery system (local installation, remote access, application virtualization, and so on). It also means that the server components of such systems are always available to service needs as well. This means incorporating high-availability functions into data storage, network connections, and even server instances.
- *Assured security.* IT's second responsibility is in securing that data against attack, inappropriate access, and inadvertent or malicious deletion. Security is critical for protecting the company and its assets against all outside attackers.
- *Acceptable user experience.* IT's third responsibility is in maintaining an acceptable user experience during the entire period that users are working with applications and data. Although sometimes superseded by the requirements of the other two, this third responsibility is critical for ensuring that user interfaces are created in ways that users will actually want to use them.

How we deliver on these responsibilities is really up to IT. When you wave away the "everything else," all IT's users really want are their applications and that data. Figure 1.1 shows what our users experience out of our IT infrastructure: They see a desktop OS with one or more applications, along with the data they need to run inside those applications.

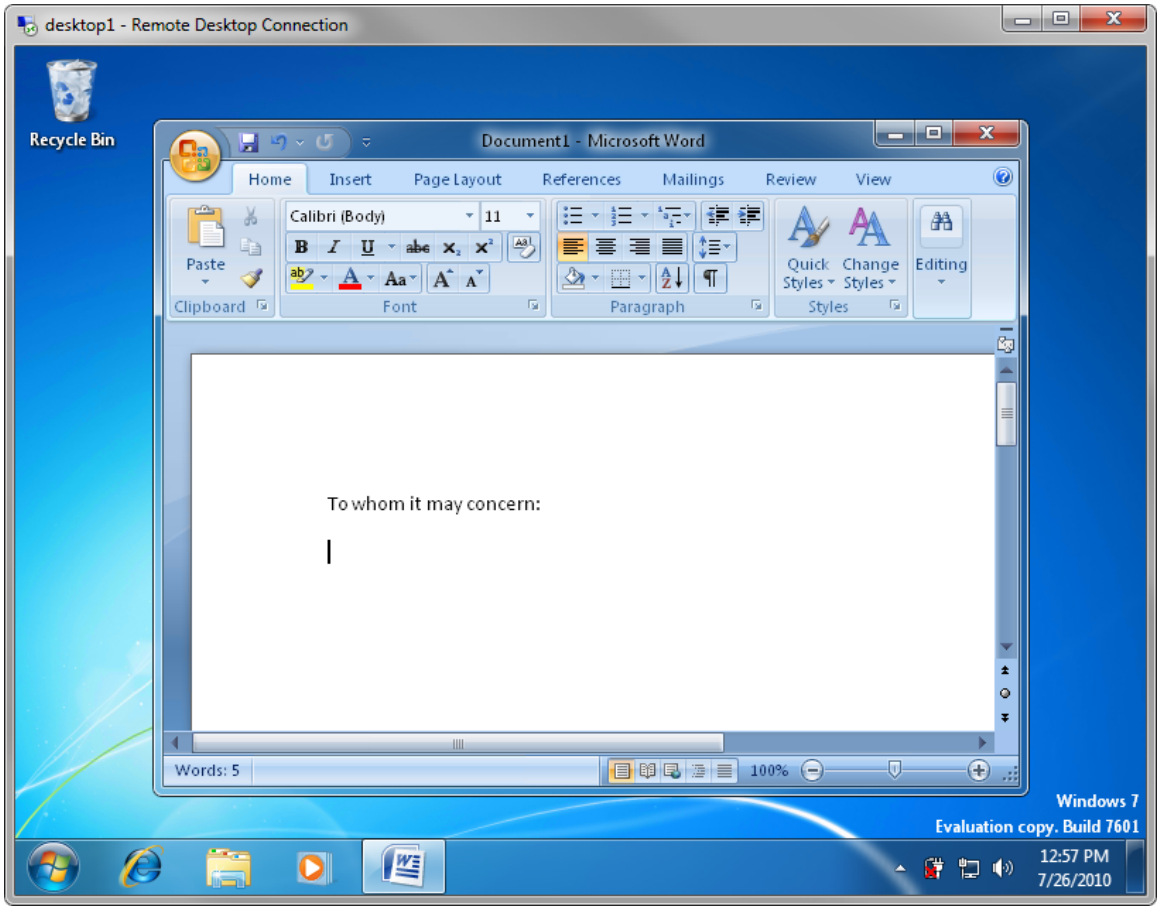


**Figure 1.1: IT service delivery, from the user's perspective.**

The data here represents the business' lifeblood. It is the information that creates value for customers and ultimately business profit. The applications are the instruments for viewing and manipulating that data. Without applications, data all by itself simply isn't useful. This is all fairly basic stuff, but what's fundamentally important in this image is the recognition that *we can meet our users' needs in whatever fashion makes sense.* We're given the charter to accomplish this as long as we provide an available, secure, and acceptable user experience.

## Defining the Private Cloud Black Box

Although every user today is comfortable with Figure 1.1's experience, many don't consciously realize the requirements to achieve it. In fact, most don't even recognize how the world of work is updating those requirements on a constant basis. Consider how the current generation of service delivery technologies is increasing the ways in which applications and data can be provisioned to users:

- Getting to that Windows 7 instance requires a desktop computer, a virtual desktop instance, or even a remote OS infrastructure. Access to that instance might be local or it might be global with access needs anywhere with an Internet connection.
- Microsoft Word might be installed locally or streamed down via an application virtualization solution. It could even be hosted on a server and delivered through a remote desktop protocol.
- In extreme cases, an individual virtual desktop might have been spawned on demand, with the application streamed to its desktop and then its single instance delivered to the user via a remote desktop protocol.
- Once within Word, finding and loading the right document requires yet another infrastructure for file services. That infrastructure can be local, on a server somewhere, or even globally distributed to replicated servers in the largest of environments.

Today's technologies enable all of these options for application delivery. The one that is selected depends on the needs of the user, the requirements of the application, and the capabilities of the hardware and network in-between.

It is perhaps easiest then to explain those user requirements by abstracting that delivery infrastructure a bit. Figure 1.2 shows a graphical representation of how IT's users experience their computing infrastructure. They sit at the bottom. Drilling up from there, they see a black box of IT services they receive. Both the users and that black box interact in some way with the rest of the Internet. Hosted on the Internet can be additional services "in the cloud" that have little or nothing to do with the private cloud but are still a part of the IT services delivery infrastructure.

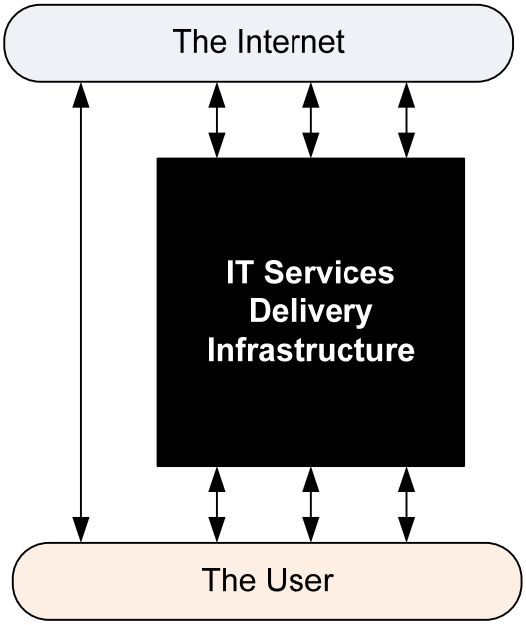


**Figure 1.2: IT services' black box.**

So users have requirements for how they interact with that black box that they might not consciously recognize. But what exactly are those requirements? When you're considering evolving your IT services delivery infrastructure into a private cloud, what does your business expect to gain from its construction? Cloud computing at its core is all about optimizing the delivery if IT services, so consider the following as a set of high-level design goals. With these goals established, you'll see why a private cloud presents the most-optimized way to construct your business IT services delivery infrastructure.

## Availability

Virtualization has been around long enough that you already know its availability story. Virtual machines, unlike their physical counterparts, are logically abstracted from the physical hardware they run on. Thus, a virtual machine can run atop any hardware instance. That same virtual machine can be seamlessly transferred to new hardware prior to a failure. In some situations, particularly so with current-generation technologies, virtual machines can even survive after the complete loss of their host.

Virtualization's availability features create the situation where IT no longer needs to think of service availability by individual server. Rather, IT can trivially incorporate high-availability features into services to protect them against many forms of downtime. Figure 1.3 shows a graphical representation of this situation. There, users require no change in process; they don't need to be aware that a failure has occurred. The private cloud's black box is constructed with the necessary resources *to maintain the availability of the service* even as its individual components experience problems, require attention, or fail over to new hardware. This is obviously a significant advantage to the traditional physical server approach, where the loss of a server tends to equate to the loss of a service.

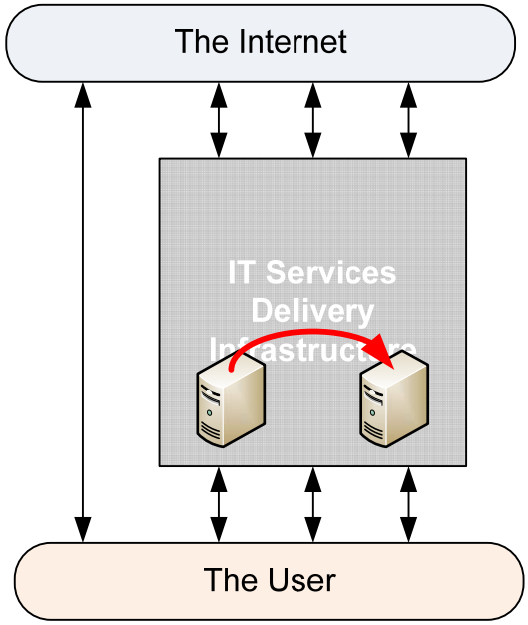


**Figure 1.3: Inside the box, servers may fail over but services remain available.**

## Flexibility

And yet virtualization's availability features are no longer its most compelling. Even more compelling to the fast-moving enterprise is flexibility in how IT fulfills its business needs. You gain this flexibility by expanding virtualization's core technologies into a complete platform for business computing.

Getting there requires a hardware investment. "Just a few virtual hosts" quickly becomes a fully-realized private cloud *as the scale of its hardware increases*. More equipment means more resources that can be brought to bear in support of business needs. Thus, a private cloud can be perceived by its owner as a collection of computing resources that can be called upon at any moment for reconfiguration or the incorporation of new services.

This flexibility represents a boon to businesses who too often suffer under the financial, technical, and manpower limitations that have been traditionally associated with IT. The following sidebar tells the all too familiar tale of how the unprepared IT organization quickly becomes the source of scalability limitations.

**Traditional IT: A Drag on Business Agility**

This story repeats itself far too often across many businesses: A new opportunity presents itself. Perhaps it is related to a new partnership or a new line of business that could bring significant profit to the company. The business itself can move with the necessary agility to take advantage of the window of opportunity.

But there's a problem: *Its IT cannot.*

Spinning up a new service requires new servers and new infrastructure components. It might require massive networking or storage reconfiguration. It likely requires the purchase of hardware, all of which takes time to arrive and ultimately get configured, tested, and ready for service delivery. As a result, the necessary IT services that mobilize that business venture simply aren't brought online fast enough.

IT becomes a drag on business agility when its technologies cannot be reconfigured at the speed of business. The traditional physical server approach to service delivery means that every new service automatically requires each of the sidebar story's time-consuming and painful steps. A private cloud infrastructure, however, (see Figure 1.4) represents an in-place collection of computing resources—computing, networking, storage, and so on—that can be trivially reconfigured at will and extended when new services are required. Although virtual machines are the mechanism in which IT services are provided, the private cloud infrastructure is the platform that enables those virtual machines to be created and managed at the speed of business.

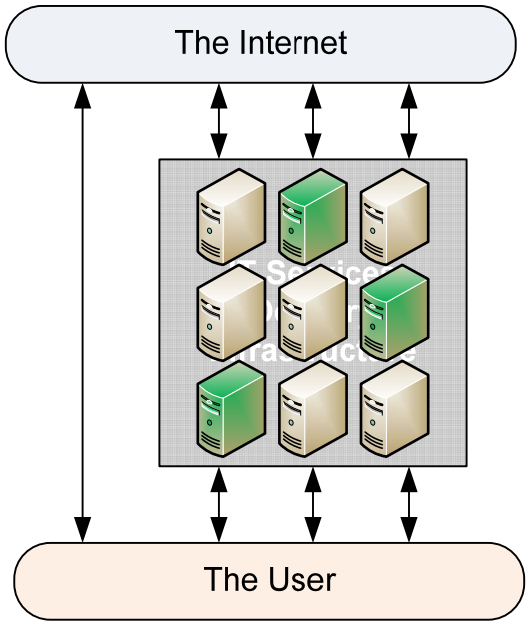


**Figure 1.4: A private cloud is a trivially-reconfigurable collection of computing resources.**

## Scalability

The laws of physics obviously still apply. Even with the capacity for trivial reconfiguration and rapid deployment, there comes the time when available hardware resources simply run out. Once your IT services have consumed the quantity of resources dedicated to your private cloud, you are at some point going to need an expansion.

It is in this situation where a private cloud provides particular value. Recall that a private cloud infrastructure exposes a set of resources into which IT services will be installed. This means that adding more resources requires grafting on more hardware. Figure 1.5 shows a graphical representation of this expansion. Scalability in a private cloud can happen with minimal extra engineering effort because its platform provides the mechanism to trivially expand the cloud through the insertion of additional hardware.

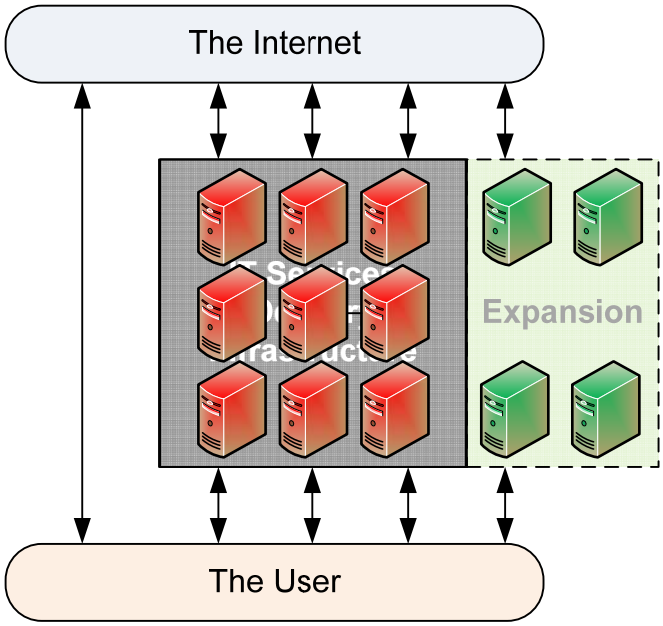


**Figure 1.5: A private cloud can be trivially expanded as necessary.**

Individual IT services themselves are not directly reliant on the private cloud infrastructure, only the services it provides. Thus, once added, that hardware becomes *an additional resource that the environment can use*.

You might scoff at the notion that any environment can be "trivially expanded," considering some of IT's previous promises along this line. However, there are commonly-accepted tactics for best accomplishing this expansion with a private cloud. One such tactic involves the use of modular hardware, as is explained in the following sidebar.

**You Get Scalability Out of Modularity**

Expanding your private cloud can be a trivial activity when new resources are required. Purchase and install a couple of new servers and their resources are yours. See that you're nearing full capacity on the resources you have, and start the purchase-and-install process long before you run out.

The challenge with private cloud expansion is not necessarily in inserting those pieces of hardware into the environment. It is instead determining the net new quantity of resources that new hardware exposes. Or, essentially, *quantifying how many resources you've added*. As you'll discover throughout this guide, one smart approach towards easily quantifying that "how many" question is in using modular hardware as the expansion.

Think of modular hardware like a set of building blocks that contributes a known quantity of resources to a private cloud. If you need three quantities of processing, you simply purchase three quantities. If you need two, buy two. Modular hardware from top-tier vendors should arrive with an assertion regarding how many resources that hardware will contribute to the private cloud.

Leveraging modular hardware for this activity eliminates the guess-and-check engineering that is required when constructing a private cloud expansion out of individual pieces and parts. I'll talk more on this topic later in this chapter as well as in Chapter 3.

## Hardware Resource Optimization

I do quite a bit of custom training on virtualization technologies, both to technologists and director and C-level stakeholders. In those engagements, there's a story I like to tell about how the traditional physical server approach is a lot like a poorly-performing employee. That story goes something like this:

> The average Windows server, one that isn't running a high-performance workload, tends to operate on average at about 5 to 7% processor utilization. Whatever service that server is running usually doesn't require much attention by the hardware. But 5 to 7% utilization is a remarkably low number. Although this might be an acceptable average for a computer system, what would you do if you had an employee in your organization that was consistently running at 5 to 7% utilization?

The crowd's answer to this question—particularly with managers—is often a chuckle along with the recognition that such an employee wouldn't be working very long at their company. But if this is the case with employees, why shouldn't it be the case with computer systems as well?

With a private cloud, it doesn't need to be. Computer systems in a private cloud enjoy the ability to host exactly the right workload to employ the resources they have available. In a well-tuned private cloud, your hardware will be correctly load balanced so that each piece of hardware will be used to its most-appropriate extent (see Figure 1.6). In this example, you can see how each piece of hardware has been assigned the correct number of virtual machines so that its utilization remains at a safe 85%. In this situation, no resources need be wasted through inactivity.

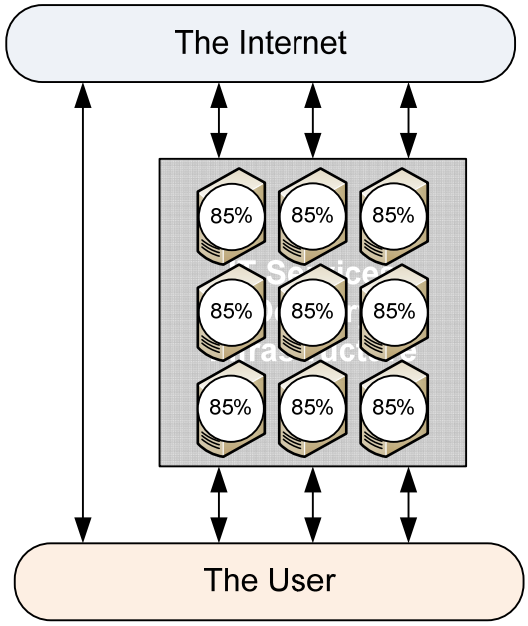


**Figure 1.6: A private cloud ensures that hardware resources are optimally used with the least-possible quantity of waste.**

Obviously, some level of non-use is critical in such an environment. A well-designed private cloud will include enough unused resource capacity so that any potential business request can be fulfilled. Right-sizing that level of spare capacity with anticipated business requests is an activity that requires technology in cooperation with process. Your private cloud's technology should provide you the metrics for identifying available capacity in terms that your engineers can understand, while your business processes will define what resources your expected services might need.

By combining all these capabilities with a private cloud's potential for expansion, *your IT infrastructure can enjoy the least possible quantity of wasted hardware resources*. That's good for the bottom line.

## Resiliency

If applications and data are the lifeblood of the modern business, than it goes without saying that their loss represents a critical impact. Businesses that experience a massive data center loss tend to not recover. It is for this reason businesses today, from the very large to the very small, are looking to disaster recovery solutions for protecting their applications and data.

Throughout IT's short history, real resiliency solutions have been attainable only by those with the deepest of pocketbooks (and, arguably, the most to lose should a loss occur). Until recently, creating a fully-duplicate computing infrastructure in an alternate location has been a fantastically-expensive solution both in terms of financial cost and human capital.

That is, until the introduction of virtualization. Virtualization, as you know, abstracts a virtual machine from its underlying hardware. This process enables that virtual machine to run atop any available hardware. Today's software and hardware solutions enable the extension of that hardware (see Figure 1.7) to alternate locations, without the technical and personnel limitations suffered by previous generations.

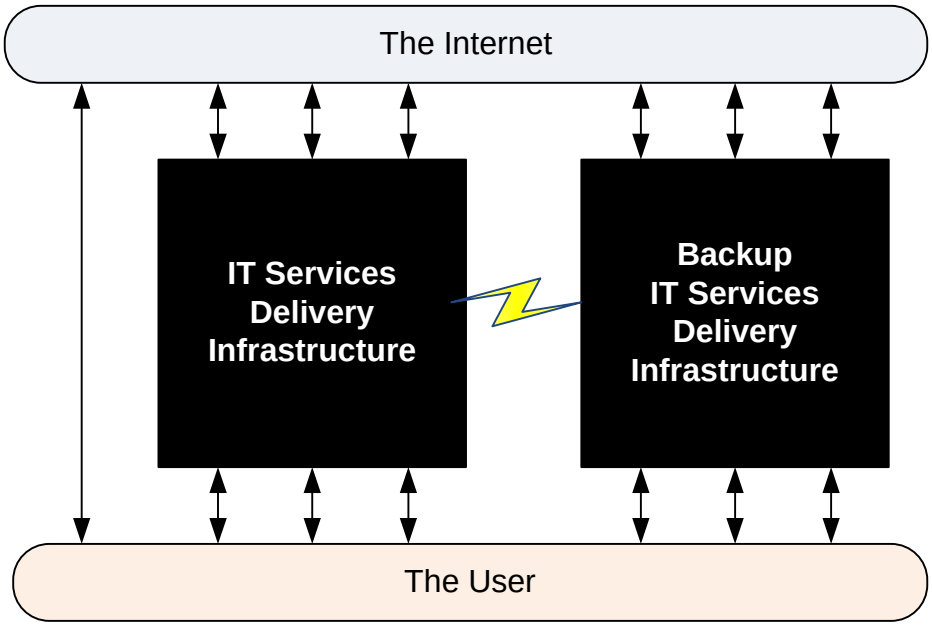


**Figure 1.7: Extending a private cloud to a backup site.**

This process occurs using the same framework that enables the incremental expansion of your private cloud. Using modular hardware and effective management toolsets, creating a resilient IT services delivery infrastructure can be accomplished with minimal effort (at least in comparison with the options available prior to virtualization and private clouds).

## Globalization

Finally is the promise of extending your private cloud to other user locations around the world. This expansion needn't necessarily comprise what the industry considers "cloud computing," where your organization provides services for others. Rather, it creates a global computing infrastructure for use by your multinational users.

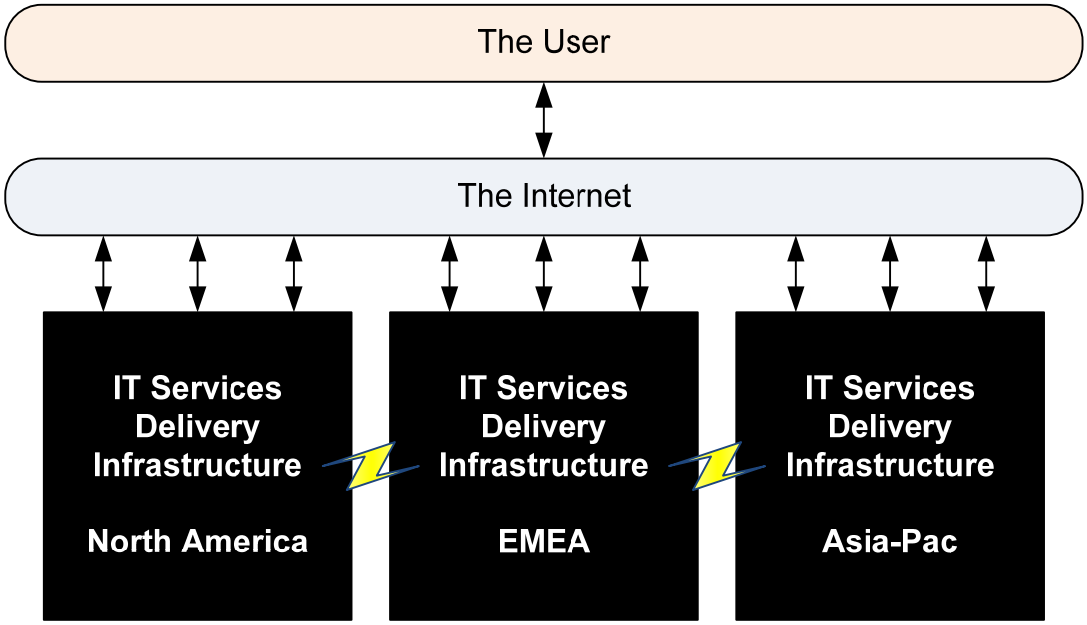


**Figure 1.8: Leveraging the Internet as the communications path between users and globally-distributed sites.**

Globally-distributed private clouds come with their share of complexities. Network connectivity between globally-distributed sites remains at a level far below LAN speed, and is expected to remain so for the foreseeable future. Resources that are needed by users in one geographic area have historically been housed in geographically local data centers.

However, recent advancements in remote protocols have brought about great improvements to user experiences across previous geographical boundaries. These enhancements, coupled with a well-designed and well-connected private cloud infrastructure, can create a local-like user experience for many types of applications and their data. The right combination of hardware and software can enable user connections with an acceptable user experience and necessary security.

## What Is the Private Cloud User Experience?

From the perspective of the user, the private cloud experience is arguably no different than what they've seen in the past. If you've constructed your private cloud effectively, the familiar experience that users know (as seen in Figure 1.1) should be no different.

Although what's seen and felt in existing services shouldn't be that different, the private cloud advantage instead comes in the availability of those services in combination with IT's enhanced abilities to provide them. Recall that primary mission of IT as outlined early in this chapter:

> IT is responsible for the assured and secured delivery of applications and data with an acceptable user experience.

A private cloud enables IT to manage the services already under their purview with an enhanced expectation of availability. Previously critical situations such as server failure, service failure, and component failure can be compensated for inside the private cloud infrastructure. Even greatly-challenging situations such as entire-site failures can be compensated without substantial incremental cost to the business. Virtualization and its abstraction of resources simply makes many of these problems go away.

Although not heavily discussed in this chapter, security is also enhanced through the construction of a private cloud. Today's security faults are sourced more often by inappropriate accesses, incorrect update and patch levels, and most often by incorrect configurations. Each of these faults is made more secure through the convergence of an entire infrastructure under a unified private cloud. There, access can be better controlled, update and patch levels can be more comprehensively monitored and fixed, and incorrect configurations needn't even occur through the use of server templates and unified management toolsets.

Lastly is the all-important user experience requirement. In the end, users really just want access to applications and data. The fact that IT gives it to them via a physical desktop, a thin client, a hosted virtual desktop, or any of the other delivery mechanisms available today is less important than ensuring that it is delivered with an acceptable user experience.

Although a private cloud in and of itself won't automatically guarantee an acceptable user experience across all delivery mechanisms (an IT organization still needs to engineer each delivery mechanism correctly), it does assist in resolving performance issues as well as protecting against spikes in resource demand. For example, when a large number of users accesses the same resource simultaneously, a private cloud has the flexibility to swing more resources to fulfill needs than does any individual physical server.

## IT's Mission Is Inside the Box

When you look at that black box that represents an IT services delivery infrastructure, you learn a few things about what people want. Ultimately, the business cares about the arrows going in and out of the black box. The business is willing to pay whatever it takes (within reason) to ensure that their services are delivered with availability, security, and a good user experience. It's IT's mission to care about what's inside that black box.

It has been this chapter's goal to highlight the fact that virtualization and private clouds are today's most-optimized mechanism for accomplishing that mission. With a private cloud, an enterprise gains flexibility in service delivery, scalability in resource assignment, and enhanced availability in keeping the applications running.

The next step in your learning starts with Chapter 2's analysis of exactly how to construct that black box. There have been a number of lofty statements in this chapter about how a private cloud can dramatically change how business needs are fulfilled and IT services are managed. Chapter 2 provides the evidence to those claims. In it, you'll learn what components you'll need to build your own black box, and see the benefits you've learned about here.

## Download Additional Books from Realtime Nexus!

Realtime Nexus—The Digital Library provides world-class expert resources that IT professionals depend on to learn about the newest technologies. If you found this book to be informative, we encourage you to download more of our industry-leading technology books and video guides at Realtime Nexus. Please visit http://nexus.realtimepublishers.com.