Cloud Computing Deep Dive

The Journey to the Private Cloud

© Copyright InfoWorld Media Group. All rights reserved.
The journey to the private cloud

Virtualization is not enough. For IT to achieve qualitatively greater economies of scale, it must borrow technologies and techniques from the public cloud and apply them to the data center.

WHY DO YOU need a private cloud? What could possibly induce you to go to the time and expense of transforming a big chunk of your data center into an environment resembling that of a cloud service provider? Well, for one thing, you’re probably partway there already. According to most surveys, roughly half of the x86 servers operated by enterprises have been virtualized. You could be forgiven, in fact, for imagining that “private cloud” is just a fancy catchphrase for “on-premise virtualization.”

But virtualization alone, despite the agility and scalability it can bring to the table, can’t solve all of your problems. In fact, if left unmanaged, virtualization’s strong points can actually cause such problems as rampant virtual machine creep, an inability to charge for virtual infrastructure usage, dangerously over-permissive network security, and a whole host of other headaches.

Though virtualization almost always forms the foundation to a private cloud, stopping there misses the larger picture of what the private cloud has to offer. If you support business units that contain some level of IT expertise or regularly deal with third-party vendors that are charged with the installation of new business applications, incorporating some aspects of the private cloud into your infrastructure will save you a tremendous amount of time and, amazingly, make your users happier.

ONE SIZE DOES NOT FIT ALL

What actually separates a solid virtualization infrastructure from a “true” private cloud? If you’re looking for an easy answer to that question, you won’t find it here. Just as clouds found in nature come in an endless variety shaped by the conditions that surround them, so too do private clouds.

Key cloud characteristics range from smart design and management of server virtualization (using tools many enterprises may already own) to fully integrated environments complete with feature-rich, self-service customer portals, fully autonomous server and storage provisioning, and automated chargeback.

It should come as no surprise that the fully integrated, high-end implementations apply almost exclusively to very large enterprises—and not just due to cost. Much of the benefit of cloud computing lies in enabling more infrastructure to be managed by fewer people, so that one admin might be responsible for thousands of servers, many more than most smaller enterprises maintain. Moreover, the notion of self-service, where stakeholders provision their own resources, demands a level of expertise that line-of-business personnel in smaller enterprises typically lack.

Yet a common thread of shared resources, more efficient management, and greater business agility unifies all private cloud implementations and can be applied to any size environment. In practice, how you view and understand these benefits depends on your role within the organization.

TWO PERSPECTIVES ON THE PRIVATE CLOUD

As always seems to be the case in IT, at least two different perspectives surround the adoption of any new data center technology: the view from the CIO’s desk, which largely concerns itself with business goals; and the view from within the data center, which is driven by technology and the struggle to manage an ever-expanding workload. The good news is that the private cloud, when implemented for the right reasons, can meet both sets of needs.

FROM THE CORNER OFFICE

A CIO might see the private cloud as a means to deliver better service levels, improve responsiveness, and allocate resources among business units more effectively. Among other things, the newfound agility and efficiency of the
A private cloud can decrease the likelihood internal business units will “get tired of waiting for IT” and adopt public cloud services willy-nilly – weakening the IT organization, creating new silos and redundancies, and opening potential security vulnerabilities.

A 2010 survey of IT decision makers conducted by Forrester Research concluded that only 13 percent of enterprises surveyed were using cloud-based IaaS (infrastructure as a service) offerings, but Forrester believes the true number to be nearly double that. “It often comes as a big shock to the infrastructure and operations people [within IT] to find they grossly underestimated the cloud services in use at their organizations,” says Galen Schreck, Forrester vice president and principal analyst, “They realize they have no idea what the application owners [in business units] and developers are up to.”

That’s a dramatic statement and CIOs are definitely taking notice. Day by day, they risk losing control of their organization’s data — data they are ultimately held responsible for managing and protecting. In the old days, “rogue” projects typically took the form of departmental servers hiding underneath someone’s desk; today, data migrates to third-party public cloud providers without planning or oversight, risking data loss or regulatory violation. To the CIO, the private cloud seems like the silver bullet to stop this.

In environments where server virtualization has already gained serious traction, many of the key goals of the private cloud have already been realized. Even a basic virtualization infrastructure offers many of the agility and scalability benefits, although server virtualization alone cannot deliver all that a private cloud can offer.

Unless additional software and policies are brought to bear, such benefits as self-service, pay-as-you-go chargeback, and secure multitenancy can’t be fully realized. Obviously, not all organizations need that. If they do, a full-scale private cloud may bring significant benefits.

Also note that moving to a private cloud needn’t be an all-or-nothing affair. It can be rolled out incrementally. Even though a number of hardware and software vendors imply that you need a whole rack of new hardware and an industrial-size barrel of software spaghetti to build a real private cloud, you can easily graft individual cloud features onto an existing virtualization infrastructure with very little effort.

Even implementing agility features, such as autoprovisioning of physical servers, can be added to an existing environment without a rip and replace — often using the same software components included in the industrial-size barrel you’d get with an all-in-one solution.

No matter how you get there, constructing a private cloud can be as much of a boon to the IT practitioner as it is to the rest of the business. Anything that results in a decreased workload for data center techs (think automation and self-service provisioning) allows them to focus on things that matter much more than grunt work.

**ASPECTS OF THE PRIVATE CLOUD**

Before delving into what a private cloud looks like from a rack-and-sheet-metal perspective, it’s important to understand what problems a private cloud is designed to solve and how that sets it apart from a traditional, even fully virtualized, on-premise infrastructure.

**AGILITY**

Business units like to complain to CIOs and IT practitioners that it always takes too long to provision new services, and they often decry the up-front cost associated with them. A business unit seeking to deploy a new application may spend months or even years choosing a software vendor and lining up development resources
and consultants. But once contracts are signed and plans are put in motion, business stakeholders expect IT to react quickly and fulfill infrastructure needs.

For a traditional IT department, unanticipated requirements can be extremely difficult to manage. Business stakeholders often underestimate the server, storage, and data-protection resources that the new application will require, and they may not account for the time it takes to order, receive, configure, and implement. Alternatively, the contract for the software may have included hardware intended to be dedicated to the new application. In the latter case, not only will IT be saddled with managing that hardware, there’s also an excellent chance the software vendor will have massively over spec’d it — resulting in even less operational efficiency.

At best, this process is an expensive waste of time. At worst, it can have a lasting negative impact on the working relationship between the business unit and IT. It’s easy to say that the solution lies in better communication between IT and the business units sourcing applications, but very few IT organizations manage to fully cross that chasm.

The private cloud essentially allows everyone to have their cake and eat it, too. Project sponsors can access various types of server and storage resources that IT has made available through a self-service portal. They can review the specifications and costs of each and share them with the software vendor, which can make recommendations. When it’s time for the application to go live, the business unit “orders” the services, which are automatically provisioned and immediately available for use, all without IT needing to do anything or even necessarily be involved.

Configuring the portal, policy, and automation magic that makes it all work requires time and effort. But the efficiency benefit can be big, especially when system provisioning is a common task. From a political standpoint, the benefits are much more obvious: IT no longer has to dedicate resources to each business unit individually. Instead, they can pool the entire corporate infrastructure — servers and storage — and manage a single pool of spare capacity.

It’s easy to see how this can decrease overall costs. Just as business units can deploy a new application with little lead time, they can also increase the amount of resources granted to one that they have already deployed — even to satisfy a short-term increase in load — and then contract them afterward.

MULTITENANCY

One of the few good things about traditionally deployed dedicated infrastructure is that it’s fairly easy to maintain divisions between the infrastructure serving various applications and business units. These divisions may simply consist of installing applications on different servers, inherently providing security and performance segregation. But they may also extend all the way down through the network and storage infrastructure. Such physical separation allows IT to implement a high degree of security easily, but it also results in an incredible amount of waste.

Although resources are pooled on the same server, network, and storage hardware in a private cloud, IT must still maintain appropriate performance and security segregation between the various workloads for the resulting product to be acceptable to business units. This segregation is accomplished through automatic configuration of the virtualization, network, and storage hardware as the services are provisioned.

During the provisioning process, the automation engine will build out a virtual machine with processor and memory allocations, limits, and reservations that match the specifications the business unit chose for the system. In
addition, it will automatically configure a secure network for the system, generally using a software-based firewall for edge security. It will also, based on policy, configure the storage for that virtual machine. Although the level of direct storage integration varies from product to product, the service level for storage can be based on either known service levels for various pools of storage that users can choose between, or, ideally, on actual service-level configuration within the back-end storage itself.

GOVERNANCE
In traditional IT environments, IT governance – really just an explicit set of policies – is often seen as an obstruction in the path of business units seeking quick deployment. The lumbering nature of governance often derives from the fact that IT must apply and reapply the same policies over and over as each business unit brings in a new application or upgrades an old one.

How much performance is required? How will the application be backed up? How will the data be stored? What are the data retention policies if a system is decommissioned? What kind of redundancy will exist? These are only a few questions that IT really must ask to do its job, but the business unit almost always sees them as obstructionist and, above all, expensive.

The private cloud does not eliminate these governance requirements, but it does simplify them by allowing IT to effectively answer them once for the entire shared infrastructure and build those infrastructural costs into the usage fees that business units pay.

Remember that provisioning within a private cloud is driven entirely by the business units, largely without direct IT involvement. Therefore, IT must be very careful when it constructs the policies that define the different compute and storage products business units can choose from – and in defining the SLAs attached to them. So although IT can improve its own customer relations by requiring less information from business units, it also has a far larger internal policy burden to bear.

PRIVATE CLOUD INGREDIENTS
Now that we have a set of requirements, we can look at how different pieces of on-premise IT infrastructure can be marshaled to build a private cloud. All the basic components of the data center must be provisioned appropriately.

STORAGE
Storage is the bedrock of any IT infrastructure. At first glance, it seems that deploying storage for a private cloud would be relatively simple. Instead of requiring a fleet of different storage resources, each dedicated to a different business unit or application, a single integrated storage pool can shoulder the load of the entire cloud infrastructure.

True enough, but that fact alone demands storage solutions specifically tuned to accommodate a large number of disparate workloads. The storage must be able to scale extremely easily, must be capacity-efficient (typically by making use of eager zeroed thin provisioning), must manage performance and tiering autonomously, and, ideally, should be easy to integrate with cloud management software. To be sure, this is a tall order. Only a few storage products available in the market today satisfy all or most of these requirements.

Regardless of what kind of storage is used, monitoring and managing storage capacity and performance levels are extremely important in private cloud environments. Since IT may not have any warning that large influxes of new workloads are going to spin up, it needs to be able to turn on a dime to add capacity. Failure to adequately manage storage capacity and performance in a private cloud environment can have far-reaching impact on a wide range of users. That storage, after all, is a huge pooled resource.

NETWORK
Most implementers of high-density private cloud environments will find that the cost and effort required to manage separate storage and application networks are prohibitive – and simply unnecessary. As a result, most private clouds are built on top of high-bandwidth, converged network infrastructures. This class of network infrastructure offers far better scalability and flexibility, in addition to significantly better resource utilization.

In practice, this usually involves the implementation of redundant 10Gbps Ethernet switches to which each host or storage array has at least a pair of redundant connections. This kind of configuration may initially be overkill for many environments for which multiple 1Gbps Ethernet connections might do the trick, but simplicity and ease of scalability are critical in cloud environments where growth patterns are rarely known in advance.

Precisely what kind of switching hardware will work best depends largely upon what kind of storage hardware
you’re using – if your back-end storage is going to be based on Fibre Channel or Fibre Channel over Ethernet, switches that support the full spread of DCB or CEE (Data Center Bridging or Convergence Enhanced Ethernet) features (such as the Cisco Nexus line) is a requirement. Even some iSCSI-based storage arrays are now able to benefit from some of the standards included within the various DCB and CEE standards, so getting switching hardware that can support them right off the bat is generally prudent.

SERVERS
Virtualization is the underpinning of most private clouds. This has a number of implications when selecting a server. Many virtual machines on one physical host results in many threads, so look for CPUs with a large number of cores (such as Intel’s Sandy Bridge and Westmere-EX architectures) and rank the importance of multiplicity far above clock speed. By the same token, large memory and high-bandwidth I/O increase the VM capacity of each physical host machine.

Blade servers make ideal virtualization hosts. It’s much easier to add a blade to a blade server than to add a server to a rack, and not just because you don’t need to mess with as many interconnects – the software that comes with blade servers typically makes scaling out much easier.

Note that virtualization is not an absolute requirement for every server in a private cloud. A few cloud automation software stacks even support the deployment and management of physical servers – thus satisfying the requirements of applications that aren’t suitable for deployment on virtualization platforms.

MANAGEMENT SOFTWARE
Up to this point, the hardware that’s been described would be adequate for operating any high-density virtualized data center. Software is what truly transforms that hardware from a capable, traditionally managed IT infrastructure into a true private cloud. Though choices made in hardware selection are important, in the long run, they are not nearly as critical as determining which software will integrate them into a single self-managing affair.

Cloud management software handles such tasks as: presenting a self-service portal with which business units can order services, automating the deployment and modification of services, ensuring that policy-driven SLAs are delivered, providing a means for accurate chargeback and reporting, and ensuring secure separation of workloads. It’s easy to see that any software tasked with all these jobs needs to integrate seamlessly with every level of the infrastructure. It’s also easy to imagine how a poorly cobbled-together software framework could devolve into disaster.

As with any large software project, a great deal of planning, research, and thorough testing is required. Defining exactly what your organization hopes to get out of the implementation of a private cloud is the most important thing to do well. That will almost always involve getting feedback from business units when self-service cloud automation software is in the pilot phase. Implementing a fancy new portal won’t do any good if business units refuse to use it. Likewise, investing the time and energy in building a system that can autoprovison physical server hardware may not deliver much ROI if the data center rarely spins up new servers.

Once you’ve identified your goals, try to find the simplest way to accomplish them. You may find that if your true goal is to provide a scalable virtualization infrastructure that incorporates chargeback functionality (but not self-service), you may simply be able to add a relatively inexpensive third-party software module to your existing virtualization infrastructure to fulfill the goal (VMware’s vCenter Chargeback is a great example).

If your goals really center around automation and you have the in-house software development skills, don’t be afraid to develop your own management tools. Nearly every virtualization hypervisor, storage platform, and server hardware platform sold today comes with some flavor of API that makes it relatively easy to script common administrative tasks.

Even if outright scripting isn’t your forte, there are still options that may not require buying anything. In the VMware world, all vCenter server editions ship with an often-overlooked tool called vCenter Orchestrator, a feature-rich automation engine that allows you to build complex tasks with multiple inputs and outputs. In fact, it’s the same engine used by VMware’s vCloud Director software to manage virtualization-related automation. Then there’s Puppet, a wildly popular configuration management framework designed to automate almost any repeatable task in the data center. Puppet can create fresh installs and monitor existing nodes; push out system images, as well as update and reconfigure them; and restart your services – all unattended.
ALL-IN-ONE MANAGEMENT SOFTWARE STACKS
If you’ve examined the requirements you intend to establish for your private cloud infrastructure and determined that no small amount of scripting or a la carte third-party tools will do the job, then you may be in the market for a full cloud automation suite. The marketplace for these all-encompassing management stacks is evolving rapidly, but a partial list would include Abiquo, Cloud.com (owned by Citrix), Eucalyptus, Red Hat CloudForms, OpenStack, and of course, VMware’s offering, which includes vSphere, vShield, vCloud Director, and vCenter Operations.

Among these, the OpenStack private cloud solution is a standout, in part because it follows a Linux-like open source model. Today, under an Apache license, the OpenStack “kernel” has three components: Compute (for managing large networks of virtual machines), Object Storage (for massive storage clusters), and Image Service (for managing virtual disk images). Around that kernel – as with Linux – vendors add value. The leading commercialized version of OpenStack is Project Olympus from Citrix; startup vendors Internap, Nebula, and Piston Cloud Computing also use the OpenStack core.

Between its debut in October 2010 and today, OpenStack has already undergone four revisions. The fifth, code-named Essex and scheduled for release in spring 2012, will include two new components: Identity, for authentication and authorization, and Dashboard, a UI for managing OpenStack services.

But OpenStack is hardly the only game in town. Its best-known competitor is Eucalyptus, a private cloud implementation of Amazon Web Services that enables you to move workloads back and forth between Amazon EC2 and Eucalyptus (which also comes in an open source version).

If you’re willing to pay the licensing fees, you can even build an all-VMware private cloud utilizing their vCloud Director stack. Virtualization is the underpinning of the private cloud – and VMware still offers the most advanced virtualization management tools. In October 2011, VMware announced three new suites to “simplify and automate IT management,” including vCenter Operations Management Suite (an update of vCenter Operations for monitoring infrastructure and managing configuration), vFabric Application Management Suite (mainly devops tools), and IT Business Management Suite (reports on operating expenses, services levels, and so on).

Which suite you choose depends entirely on what features you need to get out of your private cloud offering – including whether you intend to try to integrate with public cloud offerings and what skill sets you already have cultivated within your staff. For example, although an Amazon or Google approach may yield excellent efficiency and allow the use of very low-end server and storage hardware, managing such an architecture may be entirely foreign to an IT shop used to the full-service nature of a more traditional VMware-based virtualization architecture.

Whatever you do, never adapt your requirements to the availability of corresponding features in a management suite. Chances are, “settling” for what you can get from an incomplete cloud management suite will only compound the same challenges you’re seeking to solve.

MAKE THE PRIVATE CLOUD WORK FOR YOU
Building a full-blown private cloud is not for the meek. Reaping the full range of benefits offered by the private cloud demands hard, careful work, both in specifying and implementing the technology and, perhaps to a larger extent, in reforming the way that IT is run from a business policy perspective. Even if your environment isn’t large or complex enough to require all the self-service and automation functionality of a full private cloud, you can still come out ahead by designing in some cloudlike agility and scalability.

As time goes on, the distinction between large and small enterprises will fade. Just as the earliest adopters of large-scale server virtualization a decade ago were major enterprises, the earliest adopters of private cloud technology today will be those same large corporations. Today, small enterprises often enjoy more advanced virtualization technology than that of big business.

The same trickle-down scenario will play out in private cloud solutions. As those solutions mature, they will work their way into the fabric of virtualization and storage solutions and eventually become the de facto way of doing IT. If you don’t see the value of building a private cloud right now, don’t worry – whether or not you go to it, it will inevitably come to you. ☝️