Automated, Standardized, and Private Cloud Management

August 2010

By Mary Johnston Turner, Research Director, System Management Software

Sponsored by Novell

Overview

IDC forecasts that 69% of enterprise datacenter workloads will be virtualized by 2013. Already 75% of organizations that use virtualization have made virtual servers their default environment for new application and workload deployments.

Virtualization has enabled IT organizations to hold the line on capital spending by allowing them to share physical server resources across multiple workloads and applications. As the utilization of physical servers increases, IT organizations save money on hardware, power and cooling, and facilities costs. IDC estimates the average virtualized server currently supports roughly six virtualized workloads, and that number will only go higher over time.

Virtualization and cloud computing initiatives are creating new types of management challenges across enterprise datacenters. Unfortunately, what we consistently see is that the rate and pace of operational change and complexity that results from virtualization makes it very difficult for IT teams to significantly improve operational effectiveness and IT staff efficiency. Equally disappointing, extensive use of virtualization does not always translate into better business outcomes, as spinning up a virtual machine (VM) is only part of what it takes to deliver a usable business service.

As we discuss in this iView, cloud computing represents the next step in delivering enterprise IT and business services more efficiently and dynamically than has previously been possible. Unlike virtualization, which is focused on the optimization of physical resources and individual workloads, cloud computing targets the delivery of end-to-end services used by end users. These services are generally made up of multiple workloads which frequently reside on different, heterogeneous computing platforms.

Public cloud services, including infrastructure-as-a-service, compute-on-demand options, and software-as-a-service solutions provide customers with computing and applications resources as needed, enabled via shared infrastructure and delivered over Web interfaces. Customers pay based on the services used.

Private clouds use the same concepts except the resources and services are dedicated to the needs of a single organization. Private clouds share infrastructure resources across multiple business services, using self-service portals and service catalogs, as well as policy-driven automation and configuration management tools to provision workloads on-demand.

Private clouds enable pooling and dynamic assignment of datacenter resources as needed in order to optimize business service performance and costs. Within private cloud environments, workloads move transparently across resources as needed, and end users focus more on the cost of the
resources they consume and the service levels they experience, rather than on where the associated infrastructure resources are located and how they are paid for.

This shift away from focusing on the operation of components, and toward the delivery of services using shared infrastructure, fundamentally changes many aspects of how IT teams must design, plan, deploy, and operate datacenter environments. IT teams need holistic insight into service levels, and need to be able to provision and orchestrate services across heterogeneous platforms. The ability to rapidly migrate services and to automate self-service provisioning of end-to-end services requires organizations to standardize service offerings and aggressively automate as many operational activities as possible.

For most established datacenter environments, the move to cloud, and the accompanying standardization and automation of management processes and tools, will be an evolutionary effort that takes place over several years. In speaking with IT decision makers who are already implementing private-cloud architectures, we consistently hear that the effort to define and standardize services, workflows, and SLAs is worth the effort, as those investments allow the organization to take full advantage of state-of-the-art cloud automation and management technologies.

This iView describes the types of operational changes IT teams need to make in order to benefit from the promise of private-cloud architectures, and identifies a key set of capabilities and attributes to look for when considering management tools to help implement, automate, and operate these environments.

From Virtualization to Cloud

At its core, cloud computing is about creating shared-resource pools to dynamically and cost-effectively provision and support business service requirements. Public cloud services have received lots of attention, and by IDC’s own estimates, public cloud services represent one of the fastest growing of all IT spending categories, with an estimated CAGR of 26% over the next five years.

However, many enterprises are also beginning to implement private cloud architectures. In fact, a recent IDC survey showed that that 73% of enterprise datacenters are evaluating, planning, or have already implemented private cloud strategies. Most of these initiatives focus on deploying workloads on pools of shared resources in order to drive up utilization, improve service levels, and create more flexible business environments while maintaining control over security and compliance requirements.

Eventually, many organizations expect to make use of hybrid cloud approaches where they opt to use public services for some functions where the public nature and global reach of the business make sense, but still rely on internal private cloud architectures for critical business activities. Concerns about security, compliance, and maintaining competitive business advantage have significant impact on whether organizations feel comfortable using public cloud services today.

In most organizations, private cloud architectures build on the dynamic workload migration and server consolidation capabilities provided by virtualization. In today’s environments, individual VMs can often be spun up quickly, but end users may wait weeks to get their services while different IT teams and disconnected processes work on deploying the application, the security, the network connections, and the storage resources needed to deliver a fully functioning service. With a private cloud, the goal is to standardize and automate most configuration, deployment, and workload migration activity and eventually get to the point where business users can self-provision, or at least IT can quickly order and automate deployment of standard services.

As a result, implementing cloud requires IT operations to go to a whole new level of business life-cycle integration and awareness. Rather than each silo focusing on the availability, performance, or cost of its own specific domain, private cloud computing focuses the entire organization on the rapid,
flexible delivery of business services. This new approach to datacenter planning, provisioning, and operation requires new approaches to IT workflows and demands widespread use of highly automated infrastructure management tools.

**Operational Transformations**

IDC’s work with organizations that are starting to implement private cloud strategies shows that operational transformation is crucial to the success of these projects. Specifically, these organizations find that, in addition to virtualizing the infrastructure, in order to enable dynamic resource sharing, many IT workflows, service definitions, workload and system configurations, security and compliance policies, and SLA metrics need to be highly standardized in order to facilitate extensive use of automation. Decision-making and governance processes need to become tightly focused on business outcomes, and business departments and stakeholders need to be trained to focus on the cost and quality of services rather than on ownership of infrastructure components.

In organizations that have deployed virtualization, but have not made substantive changes to processes, workflows, configurations, and SLAs, IDC finds that the ratio of administrators to managed machines improves only slightly – hovering in the range of about 29 physical servers per administrator versus 33 virtual machines per administrator. By comparison, IT organizations that make the effort to negotiate with business partners to define standard images and configurations, to eradicate one-off implementations, and to expose the true cost of providing customized service levels find they can achieve dramatic staff productivity improvements by automating the provisioning, migration, and ongoing management of those workloads and resources.

Although there are relatively few private clouds operational today, we have recently seen examples of organizations that are operating with ratios of 150 - 200 or more physical and virtual machines per administrator. A few leading-edge organizations that have aggressively standardized, integrated, and automated physical and virtual system and workload management activities are achieving ratios of 350 - 500 or more servers per administrator.

Organizations experiencing these kinds of productivity improvements typically work aggressively to share infrastructure across multiple workloads, implement governance systems to promote standardization of services, configurations, security and related policies, and invest in automated, integrated tools to provision, migrate, and maintain the health of workloads across these shared environments. These organizations are also working to make business groups more aware of their resource consumption and help them to better understand the cost of changing SLAs or assigning dedicated rather than shared resources.

**Cloud Management Priorities**

Most early, private cloud customers are making significant investments in new automated planning, provisioning, and monitoring tools to enable effective cloud operations. When it comes to evaluating solutions, decision-makers should look for the following capabilities:

- The ability to manage, provision, and maintain standardized service models that include security, compliance, and performance attributes. These models need to recognize the dependencies and interconnections across the multiple workloads that compose a service. They are needed to drive automated service provisioning activities
- The ability to migrate running workloads across heterogeneous physical, virtual, and public cloud environments in order to continually optimize resource consumption, reduce VM sprawl, and maintain security and end-to-end service levels
- Automated support for routine patch management and service provisioning activities, which includes the ability to discover configurations and patch levels, evaluate them against the gold
image, and enforce updates and compliance on an ongoing basis using software libraries and templates

- The ability to define and maintain templates for standardized, automated approval and change control workflows to streamline approval, change control, audit and compliance reporting processes
- Self-service provisioning capabilities, as well as integration with service catalogs, service fulfillment systems, and service desks
- The ability to monitor and analyze hardware and workload characteristics, and to match them appropriately in real time in order to maintain optimal service performance and costs
- Usage and consumption metering and reporting to support capacity planning, governance, and chargeback/showback programs

The most effective tools will support role-based interfaces and dashboards that can be used by both IT professionals and end user organizations. By enabling all stakeholders to interact using the same monitoring, provisioning, and workflow platforms, enterprises can reduce the time it takes to provision and support services while increasing IT’s credibility and visibility with business stakeholders.

**Measuring the Benefits**

Most organizations planning to deploy private clouds expect it will be a multi-year effort that begins with a few selected services and applications or departments, and is then extended as the benefits of the concept are proven and IT develops increased confidence in its ability to maintain SLAs while increasing staff productivity and increasing resource utilization.

In order to build the business case for private cloud solutions, IT decision makers need to develop solid data about the costs and benefits of these initiatives. The cost side of the equation includes spending on new tools and platforms, as well as staff time devoted to developing and implementing standards and service definitions.

The benefit side of the discussion needs to include both hard- and soft-dollar impacts, including the following:

- Hardware and power/cooling cost savings resulting from improved utilization due to eliminating VM sprawl, sharing resources across multiple workloads, improving capacity planning, and implementing dynamic workload balancing
- Software license and maintenance savings due to elimination of assigned but unused licenses and elimination of redundant management tools
- Improved security and compliance due to more automated enforcement of policies and standards
- Staff cost savings due to improved productivity and operational efficiencies
- Improved end-user productivity, business performance, and customer satisfaction due to faster provisioning, reductions in downtime, and consistent service levels

As datacenters become more virtualized and enterprises take greater advantage of cloud services, IT organizations will need to invest in standardization and automation initiatives designed to deliver high-quality, cost-effective services as needed to meet business requirements. Policy-based cloud management solutions will be important to getting IT and business stakeholders aligned and moving forward to support the transformation of their organization's infrastructure environments.
IT cannot implement cloud strategies on its own – business stakeholders must buy in and participate at every step. Effective measurement and reporting on costs, services levels, security, and compliance will facilitate this type of interactive decision making and build support for cloud strategies across business and IT leadership teams.