IDC TECHNOLOGY SPOTLIGHT

Dynamic Resource Provisioning for Virtualization and Private Cloud Computing

August 2010

Adapted from *Worldwide Datacenter Network 2009–2013 Forecast* by Lucinda Borovick, IDC #220397

Sponsored by Brocade Communications

Enterprises need new networking solutions to enable intelligent, automated provisioning of a dynamic infrastructure for on-demand IT resource allocation. Dynamic provisioning technology, embedded in Layer 4–7 switches, can add a level of intelligence so that enterprises better use network, virtualization, and application resources to meet the business application demands of an ever-changing workforce and customer base. This Technology Spotlight examines why the market is ready for dynamic provisioning of applications and other services, enabling enterprises to support a private cloud delivery model, and the role Layer 4–7 switches can play in enabling that model. The paper also looks at Brocade's ADX family of Layer 4–7 switches and its accompanying Application Resource Broker software, as well as the strategic importance of this technology for today's datacenter.

Introduction

Enterprise datacenters are in the midst of an unprecedented transformation as organizations look to evaluate and rearchitect their IT investments to lower costs and increase business agility. As such, larger enterprises will drive continued, albeit slower, growth in networking equipment by making ongoing investments in new consolidated datacenters and associated infrastructure.

As application-aware networks are called upon to address a litany of challenges, such as improving network reliability in the face of a growing mix of applications and increased mobility, organizations will need to improve infrastructure. Networks will have to meet the growing demand for increased speed, performance, and security. Platform upgrades to support newer, denser computing configurations — especially those incorporating virtualization and "on-demand" applications in consolidated datacenters — will also place greater pressure on network infrastructure.

The investment in a dynamic datacenter network for on-demand resource allocation will fulfill the following three related but separate goals:

- **Reduce costs by increasing the utilization rates and efficiency of equipment and facilities.** The new datacenter designs and energy-efficient configurations being implemented by IT require an intelligent network. The demand for increased utilization rates is the starting use case for server and storage virtualization, creating much denser configurations that require the highest-performing network.
- **Create dynamic network policies.** Consolidated and virtualized servers are at a higher risk of oversubscribed application loads and thus in greater need of dynamic resource allocation. During times of peak load, there is a risk that multiple applications will be competing for shared IT components (servers, storage, network). On-demand resource allocation enables IT to realize the cost savings benefits of virtualization and consolidation while maintaining service-level commitments.

- **Increase the velocity at which IT can respond to business needs.** Technology and service innovations are coming together to meet business requirements for rapid response times. Virtual server mobility, coupled with innovations from virtual server and datacenter infrastructure suppliers, provides the foundational tools to rapidly provision new workloads based on business requirements, enabling business agility. Datacenter network architectures must adapt to these new requirements with virtual technology and advanced network provisioning capabilities.

These requirements drive the need to architect a dynamic datacenter that can adapt to change. The term that is most commonly used to describe this new paradigm is private cloud, which is an evolution from a single focus on virtualization to include automation and provisioning. IT understands all too clearly that the pace of innovation in Web 2.0 and service-oriented architecture (SOA) environments means that the network can no longer be statically designed for specific device types, applications, or, for that matter, security threats. IT is challenged to find the most optimal timing to invest in new network infrastructure. Investing too early causes unnecessary expenses; investing too late creates downtime.

As a result, the datacenter network has the following requirements:

- **Support for private cloud computing.** IDC defines "public cloud" computing as accessing IT on demand over the Internet with pay-as-you-go models where IT infrastructure is shared across multiple customers. The IT infrastructure is typically highly standardized and the management highly automated. A "private cloud" is defined as having an IT strategy similar to that of public clouds, but applied within a single enterprise or an extended enterprise as an internal shared set of infrastructure.

- **Scalability.** The use of virtualization has become so pervasive that the number of virtual servers deployed has already surpassed the number of physical servers deployed in a year. While the most common initial driver of server virtualization is to optimize compute resources, the network is still required to scale to this explosion of virtual servers. Meanwhile, the transaction load on these virtual servers is no longer predictable, causing IT to find solutions that can both scale up and scale down based on time of day or time of year.

- **Performance.** Always-on networking is not just a requirement but a prerequisite. It's paramount that the network be able to meet peak traffic demands to ensure quality of experience (QoE), including faster response times, and also help meet service-level agreements (SLAs) while intelligently rerouting traffic in the face of system failure or security threats.

- **Flexibility.** Given the dynamic requirements of an evolving network, with protocols as well as voice and video applications evolving, the network must have a solid foundation at Layer 2 while possessing the flexibility to adapt to changes in applications and security threats at Layers 4–7.

- **Visibility.** Visibility into the current network environment, including auditing, logging, and bottleneck detection, as well as the ability to analyze for future growth, is a strategy for success. Layer 4–7 switches are a point of visibility into the application.

Layer 4–7 switching is an integral part of this new wave of network investment. Layer 4–7 switching is also referred to as application delivery controllers or load balancers. Despite these uncertain economic times, IDC believes datacenter Layer 4–7 switching will remain a strategic resource. IDC expects the market, which reached an estimated $1 billion in 2008, to ultimately see modest growth to $1.2 billion by 2014, after a downturn in 2009. Critical to this growth is continued innovation in networking that goes beyond increased performance and security to include products that enable the pay-as-you-grow philosophy of IT, scalability support for new Web 2.0 applications, and platform longevity that will meet TCO demands.

From a user perspective, growth will be driven by support for cloud services, convergence of storage on the Ethernet network, support for virtual network services, and the overall need to support mission-critical workloads on x86 platforms. Additionally, the expansion of network-based businesses is generating an exponential increase in the criticality of the network. The number of Internet devices is exploding and will reach 3.5 trillion in 2013. As organizations look to incorporate these users and devices, capacity planning problems and the risk of oversubscription become acute. More users accessing the network anytime, anywhere leads to more sudden increases in traffic loads. These drivers are helping to fuel the growth in the Layer 4–7 market.

Layer 4–7 Switching: Toward More Dynamic Infrastructure

Layer 4–7 switching is designed to address the unique requirements of the dramatic increase in traffic and complexity caused by distributed virtualized and hosted enterprise applications. This content- or application-aware switching intelligently load balances traffic across virtual servers in a datacenter or multiple datacenters based on availability and load. As such, Layer 4–7 switches go beyond the connectivity and routing services provided by Layer 2–3 switches by routing messages to individual machines after inspecting the contents of the messages and forwarding them based on preset policies.

Layer 4–7 switches monitor device health and can provide automatic failover by enabling enterprises to set protocols for routing messages to other devices if needed. The policies, not found in traditional LAN switches and routers, can provide intelligent traffic management capabilities, including local and global server load balancing, access control, quality of service, and bandwidth management.

The intelligence being built into Layer 4–7 switches also implicitly measures user experience with an application through response time measurements, load factors, synthetic transactions, and other factors. Thus the switch can be a source of intelligence to drive an on-demand infrastructure, enabling organizations to build datacenters with dynamic provisioning for changing workloads.

As enterprises and communications services providers move to a more on-demand computing model, including private cloud computing, Layer 4–7 switches provide the following benefits:

- Utilize resources in multiple locations
- Increase the utilization rate of existing systems
- Integrate with datacenter orchestration systems
- Demystify traffic patterns between virtual servers
- Provide necessary intelligence to enable dynamic infrastructures
- Anticipate unpredictable traffic patterns resulting from mobile Internet/anytime connectivity

Servers are used more efficiently and Web site performance can be improved, with the added bonus of the ability to perform content-based server health checks. And as organizations move to more on-demand application delivery models, they have the traffic management and security capabilities to offer differentiated services when and where they are needed.

As organizations look to build private clouds, datacenter Layer 4–7 switching will play a fundamental role. Real-time network intelligence is a fundamental building block technology to provide network intelligence with datacenter orchestration systems. The combination of measuring user experience with an application and system and network load factors is a powerful tool to drive an on-demand infrastructure.

Considering Brocade

Brocade Communications, based in San Jose, California, has released a new software upgrade, the Brocade Application Resource Broker (ARB), which will be bundled with Brocade's ServerIron ADX Series of Layer 4–7 switches. The ServerIron ADX platform is designed for high-performance application performance monitoring, acceleration, and optimization in both real and virtual server environments. The Brocade ARB software module plugs in to VMware's vCenter product and essentially enables a private cloud for enterprises through a combination of application resource monitoring and resource commissioning or decommissioning as loads increase or decrease, respectively.

The ARB software analyzes application health and performance across enterprise networks and virtual machine (VM) infrastructure. The software tracks multiple metrics, including application response time, inbound connection levels, and VM CPU utilization. The system then logs, alerts, and provisions additional application resources accordingly.

ARB provides visibility into the network and on-demand provisioning of application resources. In the example of a financial reporting application, companies typically initiate complex financial reports at the end of a quarter, significantly increasing the application load and slowing response time from application instances running in VMs. With Application Resource Broker, the ServerIron ADX platform automatically redirects new requests to other existing VMs that can provide faster response times. Should the load increase further still, the ARB software alerts the IT manager and commissions additional VMs from a shared pool. After the end of a quarter, application loads decrease dramatically and the specially created VMs are decommissioned and released for use by other applications.

The ServerIron ADX Series with Application Resource Broker provides the following key capabilities for enterprises seeking private cloud performance:

- **Agile infrastructure** — Application performance monitoring and on-demand virtual resource provisioning
- **Application intelligence** — Validated designs for application acceleration, as well as customizable application handling and an online community for collaboration
- **Application infrastructure visibility** — A single management console that spans physical and virtual infrastructures

The combination of the ServerIron ADX Series and Application Resource Broker software enables highly secure and scalable service infrastructures to help applications run more efficiently and with higher availability — streamlining operations, increasing business agility, and significantly reducing costs.

Challenges

Brocade faces some market challenges, however. First, as with any vendor in challenging economic times, Brocade must continue to demonstrate that it can deliver products that help enterprises gain greater utilization out of the compute and storage assets they have already deployed.

Second, as the nature of enterprise networks and networking in general continues to change, Brocade needs to align its future road map with the tidal wave of applications and devices that continue to demand network services. Key to the company's future success will be providing a platform that can scale with the transaction explosion and simultaneously increase specific application intelligence in a virtualized datacenter while still being mindful of constrained IT budgets.

Conclusion and Essential Guidance

Despite a muted economic recovery, enterprises are still evaluating changes to their datacenter architectures — as well they should. It's imperative that the datacenter continue to respond to critical needs as the nature of business changes. As the model for enterprise applications and their delivery moves toward more dynamic architectures such as on-demand resource allocation, new architectures will be required.

Typically, changes will include the physical space as well as the virtual, including servers, storage, applications, and, importantly, the network. Ultimately the datacenter must reflect today's business model, a dynamic environment that can quickly respond to changes while still using resources efficiently. This will require smart networking solutions that include Layer 4–7 switching to match newer computing models.

For larger enterprises looking for networking solutions that provide greater resiliency and visibility for performance management, IDC recommends finding a supplier that truly understands the application environment and can therefore help justify the cost of new equipment or upgrades. The supplier should be able to provide customer case histories and be willing to spend time to ensure that the right solution is utilized.

IDC advises doing research to understand the need, development, installation, use, and management of dynamic networking solutions. Because new business models demand flexibility, there is a risk of failing to focus solutions sufficiently and accurately. On-demand applications, virtual computing, and cloud computing all need underlying rules and processes that clearly specify business needs and appropriate levels of service. Without this understanding, dynamic provisioning technology will not deliver its true value.

Layer 4–7 switches, augmented with the right software to analyze applications and application traffic, will increasingly become a key component of dynamic networks. To the extent that Brocade can meet the market challenges described, the company's ADX platform and Application Resource Broker solution should be on the short list of evaluations by enterprise datacenters.

Global Headquarters: 5 Speen Street Framingham, MA 01701 USA P.508.872.8200 F.508.935.4015 www.idc.com