The Mandate for Network Infrastructure Automation

By Jim Metzler, Cofounder, Webtorials Editorial/Analyst Division

Executive Summary

In the last couple of years, server virtualization has become very popular in part because it reduces cost. Another reason for the popularity of server virtualization is that it enables IT organizations to provision virtual machines and to move virtual machines among physical servers, both within a given data center and between disparate data centers, in a matter of seconds or minutes. As described in this white paper, many IT organizations already make regular use of the capability to move virtual machines between servers and the interest that IT organizations have in dynamically provisioning virtual machines is increasing greatly.

The ability to dynamically provision and to dynamically move virtual machines between servers is an important step towards the ultimate goal of elastic or on-demand computing. When this goal is met, IT organizations will be able to respond to the demand for infrastructure resources with instant-on, real-time delivery of virtualized network services. However, a complex computer system is more than a virtual machine and although compute cycles have become virtual and dynamic, the rest of the overall computer system is still largely physical and static. As a result, while it is possible for an IT organization to provision or move a virtual machine in a matter of minutes, it can take weeks or months for the organization to implement the necessary changes to the rest of infrastructure.

This white paper identifies some of the challenges relative to implementing a dynamic infrastructure. This includes VLAN proliferation, the loss of visibility into the traffic within virtualized servers, the need to transfer configuration state information along with a virtual machine as well as the network and storage support that is needed. This white paper also identifies the progress that is being made relative to responding to these challenges. This includes the evolution of dynamic infrastructure management, virtualized performance management, virtualized security management, distributed virtual switching as well as orchestration and provisioning. This white paper concludes by highlighting the fact that just implementing new technologies will not enable IT organizations to realize the goal of elastic computing. IT organizations also need to implement streamlined processes that reflect the reality of a dynamic, virtualized IT infrastructure.

Introduction

Server virtualization is well on its way to becoming universally adopted within enterprise data centers. Initially, the primary factor that drove IT organizations to implement server virtualization was the significant cost savings that can be realized through hosting multiple virtual machines (VMs) on a single physical server. These savings include reducing the cost of servers, as well as the cost of the supporting real estate, power and cooling.

While cost savings is still a significant driver, today two other factors are also significant drivers of the movement to virtualize data center servers. These factors are the ability to provision VMs and the ability to move VMs among physical servers, both within a given data center and between disparate data centers, in a matter of seconds or minutes. The combination of the ease and speed with which VMs can be provisioned and moved has led many IT organizations to create initiatives to further leverage virtualization throughout their IT infrastructure. The goal of these initiatives is to have an infrastructure that has the ability to provide each application and network service with the required resources even as the demand for each service fluctuates dynamically. The ultimate in *elastic computing* (a.k.a., on-demand computing) is realized when the demand for infrastructure resources can be met with instant-on, real-time delivery of virtualized network services.

There is no doubt that an on-demand, dynamic, virtual IT infrastructure is very appealing. However, the traditional approach to network design and the associated labor-intensive management tools that are typically used to control and manage the IT infrastructure are not able to keep pace with the frequent, dynamic changes that are required. For example, the traditional approach to data center network design is based on the concept of interconnecting and managing static physical devices. This approach has two fundamental limitations when used to support virtualized servers. One limitation is that the workload for the operational support staff can spiral out of control due to the constant stream of configuration changes that are needed to support the dynamic provisioning and movement of VMs. The second limitation is that even if IT organizations had enough support staff to implement the necessary configuration changes, the time to support these changes is typically measured in days and weeks. In order to truly have a dynamic IT infrastructure, these changes must be made in the same amount of time that it takes to provision or move a VM; i.e., seconds or minutes.

One goal of this white paper is to quantify the interest that IT organizations have in implementing network infrastructure automation. Another goal of this white paper is to show that in order to implement network infrastructure automation that the network needs to respond to a number of challenges that are created by server virtualization. In particular, the network needs to evolve its own approach to abstraction and virtualization. The approach that IT organizations take to abstract and virtualize the network must allow the network to participate in an end-to-end virtualized infrastructure that features the tightly-integrated and automated management of the services that are delivered by a wide variety of system, storage, and network resources. The third goal of this white paper is to demonstrate that the promise of elastic computing will not be realized just by the implementation of new technologies. In order to realize the promise of elastic computing, IT organizations will also have to modify their management processes.

A Virtualization Reality Check

The purpose of this section is to quantify the interest that IT organizations have in implementing network infrastructure automation. To achieve that objective, this section will utilize the results of a survey that was administered in early 2010 to the attendees of the Interop conference. Throughout this white paper, the IT professionals who completed that survey will be referred to as The Survey Respondents.

The Survey Respondents were asked to indicate the percentage of their company's data center servers that have either already been virtualized or that they expected would be virtualized within the next year. Their responses are shown in Table 1.

	None	1% to 25%	26% to 50%	51% to 75%	76% to 100%
Have already been virtualized	**30%**	**34%**	**17%**	**11%**	**9%**
Expect to be virtualized within a year	**22%**	**25%**	**25%**	**16%**	**12%**

Table 1: Deployment of Virtualized Servers

The data in Table 1 shows the deep and ongoing interest that IT organizations have relative to deploying virtualized servers. Two observations that can be drawn from Table 1 are that within the next year:

- The number of IT organizations that have not implemented server virtualization will be cut by over twenty five percent.
- The number of IT organizations that have virtualized the majority of their servers will grow by forty percent.

As previously noted, two of the factors that are currently driving the movement to virtualize data center servers are the ability to provision VMs and the ability to move VMs among physical servers. As a result of being able to rapidly provision VMs, IT organizations can respond to the business requirement for additional computing resources in a matter of seconds or minutes instead of days, weeks or months. The mobility of VMs means that system administration tasks such as backup and restore, system upgrades, as well as hardware and software maintenance can be performed without impacting the availability of applications or services. Mobility can also be leveraged to ensure high application availability and workload balancing across a cluster of virtualized servers.

In order to quantify how important these two factors are to IT organizations, The Survey Respondents were asked two questions. One of the questions asked them to indicate how important the dynamic provisioning of virtual machines (VMs) is currently to their organization and how important it would be in a year. Their responses are shown in Table 2.

	Importance Currently	Importance in a Year
Very Significant Importance	**13%**	**23%**
Significant Importance	**16%**	**23%**
Moderate Importance	**20%**	**24%**
Slight Importance	**25%**	**15%**
No Importance	**26%**	**14%**

Table 2: Importance of Dynamically Provisioning VMs

As shown in Table 2, today only 29% of The Survey Respondents believe that the dynamic provisioning of VMs is either significantly or very significantly important. However, that situation will change dramatically in the near term. In particular, as is also shown in Table 2, 46% of The Survey Respondents believe that the dynamic provisioning of VMs will be either significantly or very significantly important to their organization a year from now.

Just over a third of The Survey Respondents have implemented VMotion or a similar technology. Those who had were asked a second question. They were asked to indicate how often they use that capability either manually or automatically. Their responses are shown in Table 3.

	Use Manually	**Use Automatically**
One or more times a day	**9%**	**24%**
One or more times a week	**24%**	**19%**
One or more times a month	**37%**	**17%**
One or more times a quarter	**11%**	**20%**
We rarely if ever move a VM	**20%**	**21%**

Table 3: Movement of VMs

The data in Table 3 indicates that the vast majority of the IT organizations that have implemented VMotion or a similar technology use it at least one or more times a quarter, either manually or automatically. The data also indicates that well over a third of these organizations use it at least once a week.

Almost 60% of The Survey Respondents indicated that their IT organization was interested in implementing more automated management functionality. They were also asked to indicate the primary driver of that interest. Their answers are shown in Table 4.

Driver	**Percentage of Respondents**
General interest in becoming more agile	**30%**
The desire to reduce cost	**26%**
The belief that automation improves quality	**18%**
Supporting the data center server virtualization that you have already implemented	**13%**
Supporting the data center server virtualization that you will have implemented by the end of the year	**12%**

Table 4: Drivers of Automated Management

One observation that can be drawn from the data in Table 4 is that there are a number of significant factors that are driving IT organizations to automate management tasks. In addition, combining the bottom two rows of Table 4 shows that supporting either existing or planned data center server virtualization is one of the dominant factors driving management automation.

Challenges in Implementing a Dynamic IT Infrastructure

The additional complexity of a virtualized environment coupled with the desire to leverage the potential it offers for *elastic computing* poses a number of high-level challenges for an IT organization's operational staff. These challenges generally fall into two categories:

- Exercising control over a virtualized service.
- Implementing that control in a notably more efficient manner than is possible with the traditional, labor-intensive, siloed tools sets that are not well integrated or automated.

Some of the specific challenges include:

VM Sprawl and VLAN Proliferation

The ease with which new VMs can be deployed has often led to VM proliferation, or VM sprawl. This introduces new management challenges relative to tracking VMs and their consumption of resources throughout their life cycle. In addition, the normal best practices for virtual server configuration call for creating separate VLANs for the different types of traffic to and from the VMs within the data center. While not all of these VLANs need to be routable, they all must be managed. The combined proliferation of VMs and VLANs places a significant strain on the manual processes traditionally used to manage the ranges, capacity, partitioning, allocation and de-allocation of IP addresses. As the environment becomes increasingly dynamic, efficient and effective address management becomes even more critical to the success of the IT organization. These challenges are further exacerbated by the concurrent on-going proliferation of IP-enabled mobile devices, such as VoIP phones, PDAs, and RFID entities at the edge of the enterprise network.

Visibility of Traffic Within Virtualized Servers

Prior to server virtualization, IT organizations were able to leverage their data center LAN switches in order to monitor the traffic that flowed between servers. With traditional hardware switches, however, it is not generally possible to monitor traffic or to apply network security policy to the traffic that is switched between VMs by the traditional hypervisor virtual switch (vSwitch). This creates challenges for IT organizations relative to the security filtering of VM-to-VM traffic and relative to the performance monitoring that is necessary in order to identify problems such as limited CPU cycles or I/O bottlenecks. While these performance problems can occur in a traditional physical server, they are more likely to occur in a virtualized server because the consolidation of multiple VMs onto a single physical server significantly increases both the average and the peak levels of traffic.

Network Policy Support for VM Migration

As previously discussed, many of the benefits of on-demand computing depend on the ability to migrate VMs among physical servers located in the same data center or in geographically separated data centers. The task of moving a VM is a relatively simple function of the virtual server management system. There can, however, be significant challenges in assuring that the VM's network configuration state (including QoS settings, ACLs, and firewall settings) is also transferred to the new location. In the vast majority of instances today, making these modifications to complete the VM transfer involves the time-consuming manual configuration of multiple devices.

Compliance requirements can further complicate this task. For example, assume that the VM to be transferred is supporting an application that is subject to PCI compliance. Further assume that because the application is subject to PCI compliance that the IT organization has implemented logging and auditing functionality. In addition to the VM's network configuration state, this logging and auditing capability also has to be transferred to the new physical server.

Layer 2 Network Support for VM Migration

When VMs are migrated, the network has to accommodate the constraints imposed by the VM migration utility; e.g., VMmotion. Typically the source and destination servers have to be on the same VM migration VLAN, the same VM management VLAN, and the same data VLAN. This allows the VM to retain its IP address which helps to preserve user connectivity after the migration. When migrating VMs between dispersed data centers, these constraints require that the data center LAN be extended across the physical locations or data centers without compromising the availability, resilience, and security of the VM in its new location. VM migration also requires the LAN extension service have considerable bandwidth and low latency. VMware, for example, requires at least 622 Mbps of bandwidth and less than 5 ms of round trip latency between source and destination servers over the extended LAN.

Layer 3 Network Support for VM Migration

Traffic routing to and from the virtual machine may need to be modified so that traffic flows in an optimized way to the virtual machine's new location. If the user's traffic to the virtual machine originates in the same Layer 2 domain, then a Layer 2 extension will suffice. However, if the user's traffic to the virtual machine is traversing a Layer 3 network or the Internet, then granular routes need to be advertised by the destination data center for the migrated virtual machine(s). If these changes are not provisioned, sub-optimal routing may result in additional delay, which may not be acceptable for some applications or to satisfy the requirements of some VM migration utilities.

To help ensure that the traffic from the virtual machine is optimally routed, the IP addresses of the default gateways of the data subnets in both the primary and secondary data centers need to be identical. In those instances in which remote clients are accessing VM-resident applications that are front ended by a load balancing device, the DNS lookup of the application server has to return the virtual IP address of the load balancer where the VM currently resides. These two factors require that the configurations of the DHCP server, DNS server, load balancer and virtual server management systems be kept in synchronization.

Storage Support for VM Migration

The data storage locations, including the boot device used by the virtual machine, must be accessible by both the source and destination physical servers at all times. If the physical servers are at two distinct locations and the data is replicated at the second site, then the two data sets must be identical. One approach is to extend the storage area network (SAN) to the two sites and maintain a single data source. Another option is to migrate the data space associated with a virtual machine to the secondary storage location. In either case, it is necessary to coordinate the VM and storage migrations.

Progress in Addressing the Challenges

At the present time there is no overarching solution for the comprehensive management of a computing environment composed of virtualized servers, storage, and networks. Vendors, however, are beginning to address the challenges previously described by enhancing the functionality of their products with virtualization features as well as automation and support for some level of integration - primarily with the virtual server management system. Some developments of this type include:

Network Infrastructure Automation

A dynamic virtualized environment can benefit greatly from a highly scalable and integrated DNS/DHCP/IPAM solution, which is also well integrated with the virtual server management system. Where DNS/DHCP/IPAM share a common database, the integration obviates the need to coordinate records in different locations and allows these core services to accommodate any different addressing and naming requirements of physical and virtual servers. Potential advantages of this include the automated generation of IP addresses for newly created VMs, the automated allocation of subnets for new VLANs, and the population of an IP address data base with detailed information about the current location and security profiles of VMs. The integration of infrastructure utilities with the virtual server management system can also facilitate automated changes to the DHCP and DNS databases. This results in the optimal routing of client traffic to and from virtualized servers, as was described previously.

Virtualized Performance Management

Virtual switches currently being introduced into the market can export traffic flow data to external collectors in order to provide visibility into the network flows between and among the VMs in the same physical machine. Performance management products are currently beginning to leverage this capability by collecting and analysing VM-to-VM traffic data. Another approach to monitoring VM traffic is to deploy a virtual appliance within the virtualized server. While changes in the virtual topology can be gleaned from flow analysis, a third approach to managing a virtualised server is to access the data in the virtual server management system. Gathering data from this source can also provide access to additional performance information for specific VMs, such as CPU utilization and memory utilization.

Virtualized Security Management

Virtual firewall appliances can be leveraged to allow firewalls to be dynamically migrated in conjunction with VM migration where this is necessary to extend a trust zone to a new physical location. This is one approach that can enable firewall security policies to be enforced even as virtual machines dynamically migrate between hardware devices.

Distributed Virtual Switching (DVS)

As noted earlier, with server virtualization each physical server comes with a virtual switching capability that allows connectivity among VMs on the same physical platform. The traditional virtual switch includes a data plane implemented in software as well as an integral control plane with fairly limited functionality. With DVS the control and data planes of the embedded virtual switch are decoupled. This allows the data planes of multiple virtual switches to be controlled by an external centralized management system that implements the control plane functionality.

Some of the advantages of this approach include:

- If the DVS functionality is compatible with other data center switches, there can be a consistent set of networking features and a common provisioning process that extends from the core of the network all the way to the VM's virtual NIC (vNIC).
- With DVS, the responsibility for configuring and managing the virtual switches can be transferred from the system administration staff to the network administration staff, ensuring more consistency in end-to-end network performance and security.
- Virtual servers can leverage the same network configuration, security policies, diagnostic tools, and operational models as their physical server counterparts that are attached to dedicated physical network ports.

- With the appropriate functionality, the DVS data planes can provide visibility for VM-to-VM traffic allowing the capture and analysis of VM traffic flows.

- Via linkages with the virtual server management system, the DVS can ensure that QoS and switch-resident security policies are transferred to the destination virtual switch's data plane as part of a VM migration among physical severs. With the combination of DVS and virtual firewalls and possibly other security appliances, the same security policies can be enforced wherever the VM migrates.

Orchestration and Provisioning

Service orchestration is an operational technique that helps IT organizations automate many of the manual tasks that are involved in provisioning and controlling the capacity of dynamic virtualized services. By automatically coordinating provisioning and resource reuse across servers, storage, and networks, service orchestration can help IT streamline operational workloads and overcome technology and organizational silos and boundaries. Orchestration engines use business policies to define a virtual service and to translate that service into the required physical and virtual resources needed for deployment. The orchestration engine then disseminates the needed configuration commands to the appropriate devices across the network in order to initiate the requested service. The orchestration engine can automatically initiate the creation of the required virtual machines while simultaneously deploying the network access and security models across all required infrastructure devices, including routers, switches, security devices, and core infrastructure services. The entire process can allow setup and deployment of network routes, VPNs, VLANs, ACLs, security certificates, firewall rules and DNS entries without any time consuming manual entries via device-specific management systems or CLIs.

Orchestration engines are generally limited in the range of devices with which they can interface due to differences in device and/or vendor management interfaces. Therefore, orchestration solutions mirror to some extent the constraints of virtual data center solutions that result from vendor partnerships among manufacturers of virtual server software, networks, and networked storage. The initial focus of such partnerships has been on promulgating validated network designs and architectures rather than on fully integrated or automated management. The next logical step for such partnerships is to include orchestration capabilities.

Orchestration solutions would benefit greatly from the emergence of an open standard for the exchange of information among the full range of devices that may be used to construct a dynamic virtual data center. In the Cloud Computing arena there are a number of standards under development, including the Open Cloud Computing Interface (OCCI) from the Open Grid Forum. These standards activities may also prove to provide value within the enterprise virtual data center, since the stated scope of the specification is to encompass "all high level functionality required for the life-cycle management of virtual machines (or workloads) running on virtualization technologies (or containers) supporting service elasticity".

IF-MAP is another emerging standard proposed by the Trusted Computing Group and implemented by a number of companies in the security and network industries. It is a publish/subscribe protocol that allows hosts to lookup meta-data and subscribe to service or host-specific event notifications. IF-MAP can enable auto-discovery and self-assembly (or re-assembly) of the network architecture. As such, IF-MAP has the potential to support automation and dynamic orchestration of not only security systems but also other elements of the virtual data center. For example, IF-MAP could facilitate automation of the processes associated with virtual machine provisioning and deployment by publishing all of the necessary policy and state information to an IF-MAP database that is accessible by all other elements of the extended data center.

A Day in the Life

The concept of dynamically provisioning a VM in seconds or minutes is very appealing. That appeal stems in large part from the fact that currently large IT organizations often take weeks or months to install a new server. Some of the extended time that it takes to install a server is due to manual nature of tasks such as configuration and address management. However, part of the extended time is due to the byzantine processes that are often associated with implementing a new server inside of a mid to large sized company. This section will describe the server implementation process for a hypothetical company – BigBiz. As was discussed in the preceding sections of this white paper, IT organizations won't be able to fully leverage server virtualization until they have effectively abstracted and virtualized their network infrastructure. However, as will be demonstrated in this section of the white paper, IT organizations won't be able to fully leverage the abstraction and virtualization of their network infrastructure if they don't modify their current management processes.

In this hypothetical example, an application development team inside of BigBiz makes a request for five new servers. As part of making the request, the application development team has to take the time to document why it needs five additional servers and why it needs servers of the type requested. It also needs to document the necessary ACLs, whether or not load balancing is required and a variety of other conditions, such as how often the new servers need to be backed up. This request goes to the BigBiz hardware review committee. Here it is reviewed from a variety of perspectives. This includes whether or not the added capacity is justified and whether or not there is budget for the new servers. This review process also looks at whether or not the requested systems are part of the standard product set that is supported by the IT organization, whether or not the request includes appropriate security functionality and whether or not there is space for the new servers in the data center requested by the application development team. In some instances there are additional time-consuming reviews that take place. For example, if the servers are intended to support an application that is jointed developed with one or more of BigBiz's partners, then the committee will also review the integration testing that the request indicates will be done with the relevant partner(s). Unfortunately, since the review process is as much about politics and bureaucracy as it is about technology, the BigBiz hardware review committee has a tendency to send the request back to the application development team if any piece of the documentation is missing or deemed to be incomplete.

Once the BigBiz hardware review committee has approved the request, the request is sent to the BibBiz logistics committee. This committee takes the request and creates a set of tasks that need to be performed. This includes tasks such as identifying the VLAN that each server will reside in and determining if there is enough power for the new systems and if there is not, initiating a project to provision more power. The committee determines if there are enough power outlets and network switches to support the new servers and if there is not, the committee orders the necessary equipment. The BibBiz logistics committee also creates order forms for the necessary peripherals; i.e., disks and cabinets. Hopefully a blanket purchase order (PO) exists with the supplier(s). If not, one has to be created. In any case, the BibBiz logistics committee gets involved with the supplier(s) to negotiate ship dates.

As the equipment that has been ordered arrives, it is accepted by the BigBiz purchasing organization and stored. When all of the necessary equipment has been received, another BigBiz team integrates the equipment and sends the servers to the designated data center. Once the servers have arrived at the designated data center, the BigBiz operations team prepares them for production. The network operations group, for example, requests a block of IP addresses and permission to assign the servers to one or more VLANs. Assuming that the network operations group receives the requested IP addresses, they assign addresses to each server and make the appropriate entries into DNS. Assuming that the network operations group receives permission to assign the servers to one or more VLANs, the group configures the VLAN(s). As part of this process, the database operations group is responsible for anything related to the database access that has been requested by the application development team. This includes tasks such as configuring fibre channel access. Both the network and database operations groups provide input to the security operations team that is responsible for a variety of tasks, including inputting the appropriate access control lists (ACLs).

After each step in the processes discussed above has completed, the five servers can be placed into production. However, if at any point in the overall process something goes wrong, the process can stall for an indeterminate amount of time.

Conclusion

Based in part on the desire to reduce cost, IT organizations have already made, and will continue to make broad deployments of server virtualization. As is the case with any new technology, server virtualization introduces some new management challenges to which IT organizations need to respond. These challenges include VM sprawl, VLAN proliferation and the loss of visibility into the traffic that flows between or among VMs on a given physical server.

While reducing cost remains a driver of server virtualization, there is a large and growing interest on the part of IT organizations to implement server virtualization in order to become more agile. In particular, IT organizations are implementing server virtualization in order to be able to dynamically provision and move VMs. However, a complex computer system is more than a VM and although compute cycles have become virtual and dynamic, the rest of the overall computer system is still largely physical and static. As a result, IT organizations will gain relatively little added agility from server virtualization if it still takes the organization days or weeks to manually provision the rest of the system.

In order to realize the agility that server virtualization promises, IT organizations need to respond to a number of challenges. This includes dynamically providing:

- Network policy support for VM migration
- Layer 2 network support for VM migration
- Layer 3 network support for VM migration
- Storage support for VM migration

Unfortunately at the present time there is no overarching solution for the comprehensive management of a computing environment composed of virtualized servers, storage, and networks. Some of the areas, however, in which progress has been made include:

- Dynamic infrastructure management
- Virtualized performance management
- Virtualized security management
- Distributed virtual switching
- Orchestration and provisioning

While each of the areas listed above holds great promise to increase the agility of IT organizations, another roadblock exists. That roadblock is the byzantine processes that are often associated with implementing a new server inside of a mid to large sized company. Hence, in order to truly have the advantages of elastic computing, IT organizations must improve their competence in the areas listed above and must also streamline their processes to better reflect the reality of a dynamic, virtualized IT infrastructure.

About the Webtorials® Editorial/Analyst Division

The Webtorials® Editorial/Analyst Division, a joint venture of industry veterans Steven Taylor and Jim Metzler, is devoted to performing in-depth analysis and research in focused areas such as Metro Ethernet and MPLS, as well as in areas that cross the traditional functional boundaries of IT, such as Unified Communications and Application Delivery. The Editorial/Analyst Division's focus is on providing actionable insight through custom research with a forward looking viewpoint. Through reports that examine industry dynamics from both a demand and a supply perspective, the firm educates the marketplace both on emerging trends and the role that IT products, services and processes play in responding to those trends.

For more information and for additional Webtorials® Editorial/Analyst Division products, please contact Jim Metzler at jim@webtorials.com or Steven Taylor at taylor@webtorials.com.

Published by Webtorials Editorial/Analyst Division
www.Webtorias.com

Professional Opinions Disclaimer
All information presented and opinions expressed in this publication represent the current opinions of the author(s) based on professional judgment and best available information at the time of the presentation. Consequently, the information is subject to change, and no liability for advice presented is assumed. Ultimate responsibility for choice of appropriate solutions remains with the reader.

Division Cofounders:
Jim Metzler
jim@webtorials.com
Steven Taylor
taylor@webtorials.com