The 2011 Cloud Networking Report

Part 4: The Management of Cloud Computing

By Dr. Jim Metzler
Ashton Metzler & Associates
Distinguished Research Fellow and Co-Founder
Webtorials Analyst Division

ipanema Technologies

Produced by:

Webtorials

Table of Contents

The Management of Cloud Computing

Executive Summary

The **2011 Cloud Networking Report** will be published both in its entirety and in a serial fashion. This is the fourth of the serial publications. One goal of this publication is to describe the management challenges created by cloud computing and to identify the importance that IT organizations place on these challenges. Another goal of this publication is to describe how the adoption of cloud computing increases the difficulty of traditional management processes such as root cause analysis. The third goal of this publication is to describe some of the possible approaches that IT organizations can take to better manage in a cloud computing environment and to identify the value that IT organizations see in those approaches.

In order to quantify how IT organizations are responding to the challenges of managing a cloud computing environment, this publication includes the results of surveys that were given to the subscribers of Webtorials.com in 2010 and 2011. Throughout this publication, those two groups of respondents will be respectively referred to as The 2010 Webtorials Respondents and The 2011 Webtorials Respondents.

Importance of Managing Cloud Computing

One of the questions that was administered to The 2011 Webtorials Respondents was "Please indicate how important it is to your organization to get better at each of the following tasks over the next year." The question included twenty wide-ranging management tasks, many of which were included in a similar question that was administered to The 2010 Webtorials Respondents. The possible answers were to the question were:

- Extremely important
- Very important
- Moderately important
- Slightly important
- Not at all important

In order to avoid restating that question each time it is referenced in this section of the report, it will be referred to as The Question. Three of the twenty tasks that were included in The Question were managing private, managing hybrid and managing public cloud computing solutions. The responses of The 2011 Webtorials Respondents are summarized in Table 1.

Table 1: Importance of Managing Cloud Solutions

	Private Cloud	Hybrid Cloud	Public Cloud
Extremely	16.5%	9.2%	5.3%
Very	35.7%	31.1%	23.9%
Moderately	21.7%	25.2%	23.9%
Slightly	11.3%	15.1%	23.9%
Not at All	14.8%	19.3%	23.0%

One observation that can be drawn from the data in Table 1 is that

The majority of IT organizations believe that getting better at managing private cloud computing solutions is either very or extremely important.

Another observation that can be drawn from the data in Table 1 is that managing a private cloud is more important than managing a hybrid cloud which is itself more important than managing a public cloud. One of the reasons for this phenomenon is that enterprise IT organizations are making more use of private cloud solutions than they are of either public or hybrid cloud solutions. Another reason for this phenomenon is that as complicated as it is to manage a private cloud, it is notably more doable than is managing either a hybrid or public cloud and IT organizations are placing more emphasis on activities that have a higher chance of success.

The Evolving Management Environment

The Increased Focus on Services

Just as IT organizations are getting somewhat comfortable with managing the performance of applications, they are being tasked with managing the performance of services. IT professionals use the term *service* in a variety of ways. For example, the ITIL definition of service[1] states that a service:

- Is based on the use of Information Technology.
- Supports one or more of the customer's business processes.
- Is comprised of a combination of people, processes and technology.
- Should be defined in a Service Level Agreement (SLA).

In part because the ongoing adoption of virtualization and cloud computing has created the concept of everything as a service (XaaS), the term service as used in this section of the report will sometimes refer to services that IT organizations acquired from a public cloud computing provider; e.g., compute, storage, applications.

In order to quantify the interest that IT organizations have in managing this type of service, three of the twenty tasks that were included in The Question were:

- Effectively monitoring and managing compute services acquired from a third party such as Rackspace.

- Effectively monitoring and managing storage services acquired from a third party such as Rackspace.

- Effectively monitoring and managing applications acquired from a software-as-a-service provider such as Salesforce.com.

The responses of The 2011 Webtorials Respondents are summarized in Table 2.

Table 2: Importance of Effectively Monitoring and Managing Cloud Solutions

	Compute Services	Storage Services	SaaS Based Applications
Extremely	8.1%	2.9%	9.4%
Very	20.7%	20.0%	30.8%
Moderately	25.2%	28.6%	23.9%
Slightly	22.5%	20.0%	18.8%
Not at All	23.4%	28.6%	17.1%

As shown in Table 2, 28.6% of The Webtorials Respondents responded with “not at all important” when asked about the importance of getting better at monitoring and

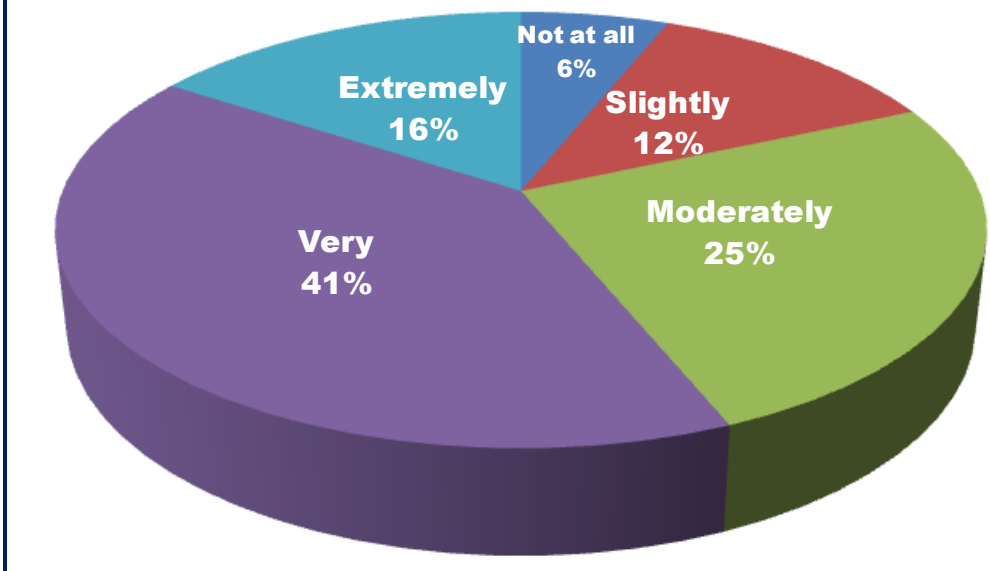
[1] ITIL definition of service

managing storage services that they acquire from a public cloud computing vendor; a.k.a., an Infrastructure as a Service (IaaS) vendor.

The 28.6% was the largest percentage to respond with "not at all important" for any of the twenty management tasks that were presented to The Webtorials Respondents. Given that, it is possible to conclude that monitoring and managing the services obtained from an IaaS vendor is not an important task. However, that conclusion is contradicted by the fact that almost a quarter of The Webtorials Respondents indicated that getting better at monitoring and managing storage services acquired from an IaaS vendor was either very or extremely important. A more reasonable conclusion is based on the observation that many companies don't make any use of storage and compute services from an IaaS vendor and the ones that do often make only minor use of such services. Based on that observation, the data in Table 2 suggests that if a company makes significant use of the services provided by an IaaS vendor, then monitoring and managing those services is indeed an important task.

The term service as used in this section of the report will sometimes refer to business services that involve multiple inter-related applications. One of the management tasks that was included in The Question was "Manage a business service, such as CRM, that is supported by multiple, inter-related applications." The answers of The 2011 Webtorials Respondents are summarized in Figure 1.

Figure 1: Importance of Getting Better at Managing Services



One observation that can be drawn from Figure 1 is that:

The majority of IT organizations believe that getting better at managing inter-related applications that comprise a business service is either very or extremely important.

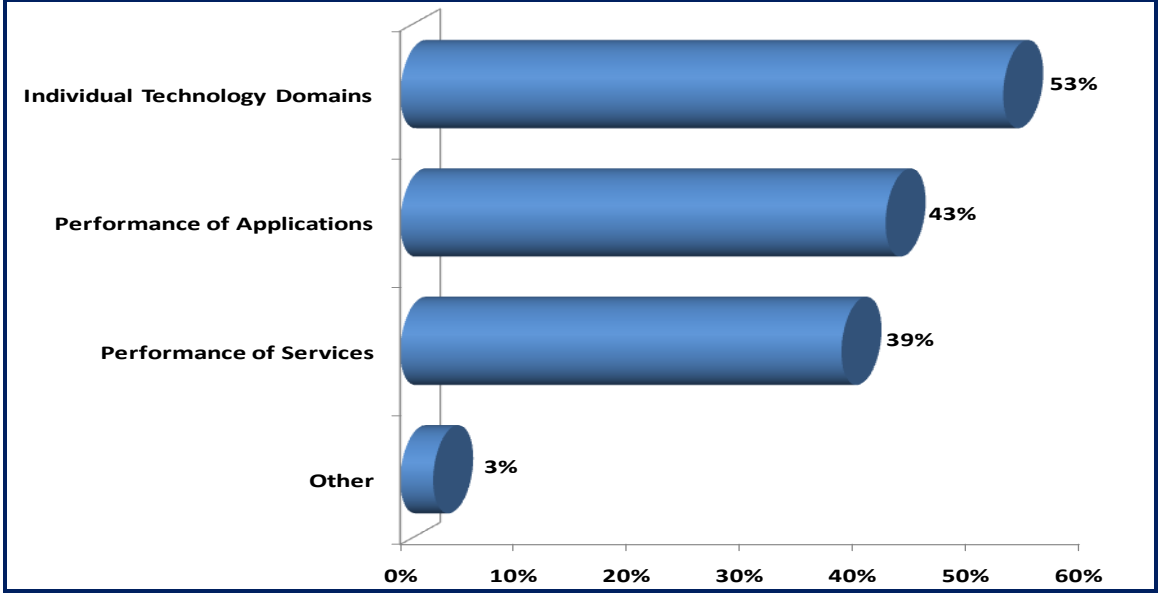
Unfortunately, the adoption of cloud computing will further complicate the task of managing the inter-related applications that comprise a service. That follows because in a cloud computing environment, the applications that comprise the service will increasingly be supported by an infrastructure that is virtual. The challenges that are associated with managing server virtualization are discussed below. In addition, as is also discussed below, managing application performance in a cloud computing environment is extremely complex.

The 2010 Webtorials Respondents were asked to indicate the approach their organization takes to management. They were given the following choices and allowed to choose all that applied to their environment.

- We have a focus primarily on individual technology domains such as LAN, WAN and servers
- We have a focus on managing the performance of applications as seen by the end user
- We have a focus on managing the performance of services as seen by the end user, where service refers to multiple, inter-related applications
- Other

Their responses are summarized in Figure 2.

Figure 2: Focus of Management



The data in Figure 2 indicates that the most frequent approach that IT organizations take to management is to focus on individual technology domains. However:

A significant percentage of IT organizations focus their management activities on the performance of applications and/or services.

The Growing Importance of Application Performance Management

In order to quantify how successful IT organizations are with their growing focus on managing the performance of applications and services, The 2011 Webtorials Respondents were given a set of statements and were asked to indicate which of the statements described their organization's approach to application performance management (APM). They were allowed to indicate all that applied.

Only about fifteen percent of The 2011 Survey Respondents indicated that their organization currently does a good job of APM. In addition, The 2011 Survey Respondents indicated by a significant margin that the approach that their organization takes to APM is that each technical discipline does its own thing vs. their using an approach that is top down and pretty tightly coordinated.

There is growing discussion in the industry about the best technical approach to implement APM. One approach is to be able to infer application performance based on management data, such as NetFlow, that is routinely collected by the network elements. An alternative approach is to use specialized agents to gather more sophisticated management data. Approximately twelve percent of The 2011 Survey Respondents indicated that their approach to APM makes heavy use of specialized agents to monitor the status of the various components of the application delivery chain.

One conclusion that can be drawn from the data discussed in the preceding two paragraphs is that

> ***APM is a work in progress. By that is meant that in spite of its importance, the vast majority of IT organizations don't do a good job of it.***

The statement that APM is both important and a work in progress is supported by the fact that a third of The 2011 Survey Respondents indicated that it was important to their organization to get better at APM over the next year.

Communications Based Applications

Communications based applications are an important class of application in part because these applications tend to be highly visible and their performance can degrade quickly if they experience impairments such as undo delay, jitter or packet loss. These applications are also important because as explained in the section of this report entitled ***The Wide Area Network***, over the next year almost 80% of IT organizations will increase their use of video, and in many cases the increased use of video will be substantial.

Another reason why communications based applications are an important class of applications in general and important relative to cloud computing in particular is that as discussed in the section of this report entitled ***The Emergence of Cloud Computing and Cloud Networking***, services such as VoIP and unified communications are now available from a cloud computing service provider (CCSP). As that section of the report also discussed, there is significant interest on the part of IT organizations to acquire both VoIP and unified communications from a CCSP.

To quantify the challenges associated with supporting a range of communications traffic, The 2011 Webtorials Respondents were asked to indicate how important it was over the next year for their IT organization to get better at managing the use of VoIP, traditional video traffic and telepresence. Their answers are summarized in Table 3.

Table 3: Importance of Managing the Use of Communications Based Traffic			
	VoIP	**Traditional Video Traffic**	**Telepresence**
Extremely Important	13.4%	6.8%	4.8%
Very Important	33.9%	20.3%	25.6%
Moderately Important	29.9%	29.7%	25.6%
Slightly Important	14.2%	28.0%	24.8%
Not at all Important	8.7%	15.3%	19.2%

The data in Table 3 shows that almost 50% of The Survey respondents indicated that getting better at managing the use of VoIP traffic is either very or extremely important to their IT organization. This is a significant percentage, particularly given that VoIP is not a new application. The challenge of managing VoIP will increase in those situations in which VoIP is acquired from a CCSP. In those instances, the IT organization will have to be able to gather and correlate management data from the CCSP, the network or networks that carry the VoIP traffic and the users' devices.

Internal SLAs

As recently as two or three years ago, few IT organizations offered an SLA to the company's business and functional managers; a.k.a., an internal SLA. However, that situation has changed and now it is common for IT organizations to offer internal SLAs. To understand the prevalence and effectiveness of internal SLAs, The 2010 Webtorials Respondents were asked to indicate their agreement or disagreement with three statements. The three statements and the percentage of The Webtorials Respondents that agreed with the statement are shown in Table 4.

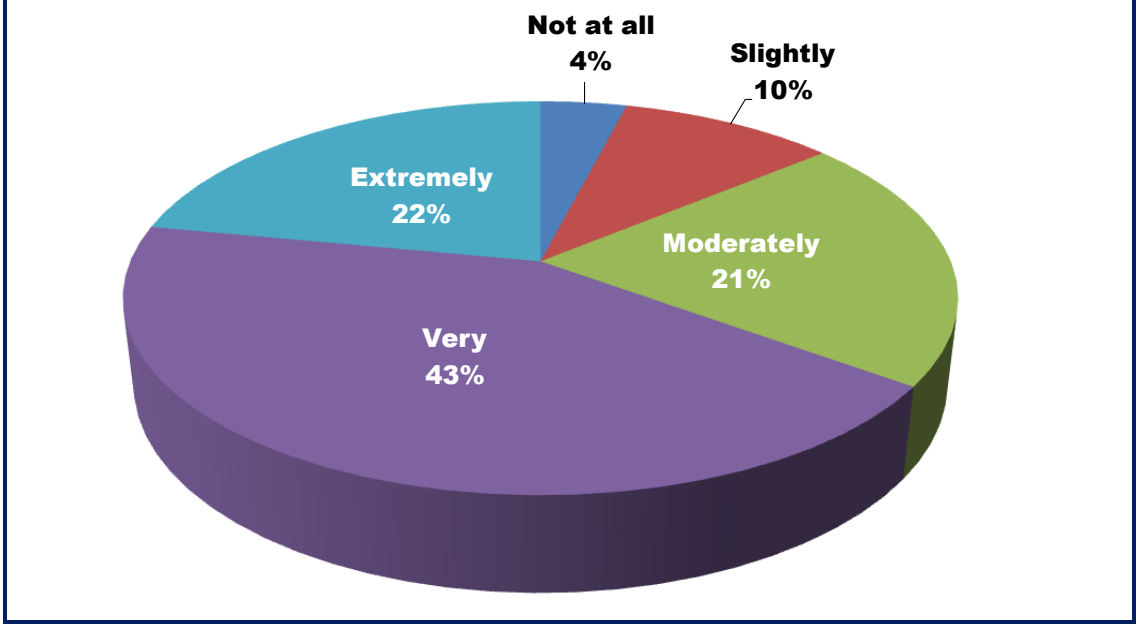
Table 4: Status of Internal SLAs	
Statement	**Percentage**
We provide an SLA internally for every application that we support	30.0%
We provide an SLA internally for at least some applications	69.9%
We do a good job of managing our internal SLAs	55.8%

The data in Table 4 highlights the growing interest that IT organizations have in providing internal SLAs for at least some applications.

The vast majority of IT organizations provide an internal SLA for at least some applications.

One of the answers to The Question was managing internal SLAs for one or more business-critical applications. The responses of The 2011 Webtorials Respondents are summarized in Figure 3.

Figure 3: The Importance of Getting Better at Managing Internal SLAs



The data in Figure 3 leads to two related conclusions. One conclusion is that

Two thirds of IT organizations believe that it is either very or extremely important to get better at effectively managing internal SLAs.

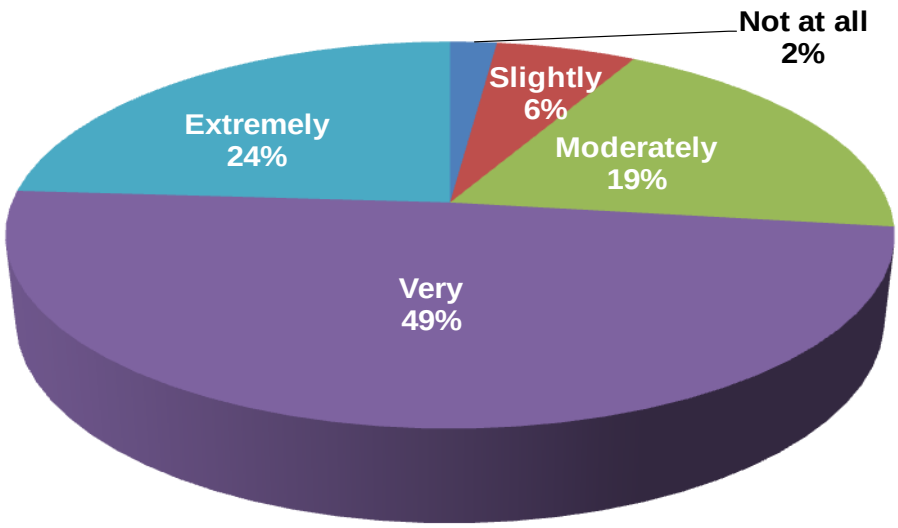
A somewhat more subtle conclusion is that managing internal SLAs is difficult or else the majority of IT organizations would already be doing a good job of managing these SLAs and hence would not be striving to get better at the task. Unfortunately, the movement to utilize public cloud computing services greatly increases the difficulty associated with managing an internal SLA. That follows in part because of the difficulty of gathering all of the management data on an end-to-end basis that is necessary to effectively monitor an SLA. It also follows because as pointed out in the section of this report entitled ***The Emergence of Cloud Computing and Cloud Networking***, it is common for CCSPs to deliver their services over the Internet and no vendor will provide an end-to-end performance guarantee for services and applications that are delivered over the Internet.

The lack of meaningful SLAs for public cloud services is a deterrent to the Global 2000 adopting these services for delay-sensitive, business-critical applications.

Root Cause Analysis

The 2011 Webtorials Respondents were asked how important it was over the next year for their organization to get better at rapidly identifying the causes of application degradation. Their responses are shown in Figure 4.

Figure 4: The Importance of Getting Better at Root Cause Analysis



Comparing the answers that The 2011 Webtorials Respondents gave to the this management task to the other nineteen management tasks shows that:

Getting better at doing root cause analysis is the most important management task facing the vast majority of IT organizations.

It is not surprising that rapidly identifying the root cause of degraded application performance is so important to IT organizations in part because on an ever increasing basis a company's key business processes rely on a handful of applications. That means that if those applications are not running well, neither are those key business processes.

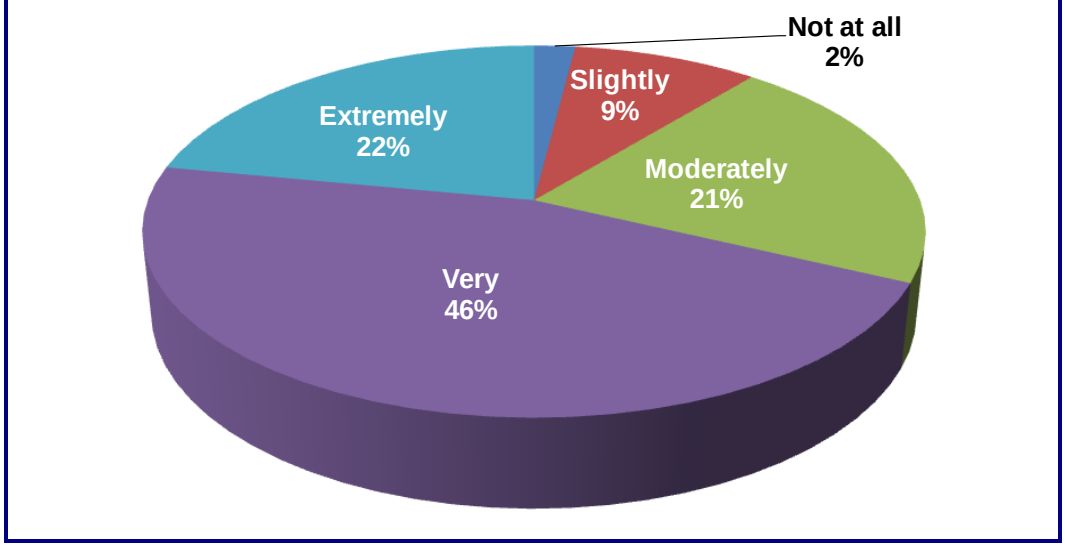
A prerequisite to being able to perform effective root cause analysis is the automatic discovery of all the elements in the IT infrastructure that support each service or application. If IT organizations can effectively identify which components of the infrastructure support a particular application or service, monitoring can much more easily identify when services are about to degrade due to problems in the infrastructure. As part of this approach, predictive techniques such as heuristic-based trending of software issues and infrastructure key performance indicators can be employed to identify and alert management of problems before they impact end users. In addition, outages and other incidents that generate alerts can be prioritized based on their

potential business impact. Prioritization can be based on a number of factors including the affected business process and its value to the enterprise, the identity and number of users affected and the severity of the issue.

Once the components of the infrastructure that support a given application or service has been identified, triage and root cause analysis can be applied at both the application and the infrastructure levels. When applied directly to applications, triage and root cause analysis can identify application issues such as the depletion of threads and pooled resources, memory leaks or internal failures within a Java server or .NET server. At the infrastructure level, root cause analysis can determine the subsystem within the component that is causing the problem.

The 2011 Webtorials Respondents were asked how important it was over the next year for their organization to get better at identifying the components of the IT infrastructure that support the company's critical business applications. Their responses are shown in Figure 5.

Figure 5: Importance of Identifying Components of Critical Applications



A clear observation that can be drawn from Figure 5 is that

Getting better at identifying the components of the IT infrastructure that support the company's critical business applications and services is one of the most important management tasks facing IT organizations.

Server Virtualization

As discussed in the section of this report entitled ***The Emergence of Cloud Computing and Cloud Networking***, there isn't a universally accepted definition of what is meant by cloud computing. That section of the report included a number of characteristics of a cloud computing solution, but also pointed out that there is not a litmus test to determine if a particular service is indeed a cloud computing service based on how many of the characteristics it supports. That said, the vast majority of private, public and hybrid cloud computing solutions are based at least in part on server virtualization. Hence, the management challenges that are associated with server virtualization can reasonably be regarded as management challenges for cloud computing.

As pointed out in Virtualization: Benefits, Challenges and Solutions, server virtualization creates a number of management challenges. For example, the need to manually reconfigure the network to support VM migration that was discussed in the section of the report entitled ***The Emerging Data Center LAN*** can be regarded as either a LAN challenge or a management challenge. Additional management challenges that are associated with server virtualization include:

Breakdown of Network Design and Management Tools
The workload for the operational staff can spiral out of control due to the constant stream of configuration changes that must be made to the static date center network devices in order to support the dynamic provisioning and movement of VMs.

Limited VM-to-VM Traffic Visibility
The first generation of vSwitches doesn't have the same traffic monitoring features as does physical access switches. This limits the IT organization's ability to do security filtering, performance monitoring and troubleshooting within virtualized server domains.

Poor Management Scalability
Many IT organizations have experienced VM proliferation sometimes called VM sprawl. 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. The combined proliferation of VMs and VLANs places a significant strain on the manual processes that are traditionally used to manage servers and the supporting infrastructure.

Contentious Management of the vSwitch
Each virtualized server includes at least one software-based vSwitch. This adds yet another layer to the existing data center LAN architecture. It also creates organizational stress and leads to inconsistent policy implementation.

Inconsistent Network Policy Enforcement
Traditional vSwitches lack some of the advanced features that are required to provide a high degree of traffic control and isolation. Even when vSwitches support some of these features, they may not be fully compatible with similar features that are offered by physical access switches. This situation leads to the implementation of inconsistent end-to-end network policies.

Multiple Hypervisors

It is becoming common to find IT organizations using multiple hypervisors, each of which comes with their own management system and their own management interface. In addition, the management functionality provided by each hypervisor varies as does the degree to which each hypervisor management system is integrated with other management systems.

Management on a per-VM Basis

IT organizations typically perform management tasks such as discovery, capacity planning and troubleshooting on a per server basis. While that is still required, IT organizations must also perform those tasks on a per-VM basis.

In order to quantify the interest that IT organizations have in responding to the management challenges that are created by server virtualization, three of the twenty tasks that were included in The Question were:

- Manage the traffic that goes between virtual machines (VMs) on a single physical server.
- Support the movement of VMs between servers in different data centers.
- Perform traditional management tasks such as troubleshooting and performance management on a per VM basis.

The responses of The 2011 Webtorials Respondents are summarized in Table 5.

Table 5: Importance of Managing Server Virtualization

	Traffic Between VMs	Move VMs Between Servers	Manage on a per VM Basis
Extremely	7.3%	15.4%	12.9%
Very	29.0%	32.5%	37.9%
Moderately	29.8%	20.5%	29.8%
Slightly	17.7%	18.8%	16.1%
Not at All	16.1%	12.8%	3.2%

One conclusion that can be drawn from the data in Table 5 is that managing the traffic that goes between VMs on a single physical server is not a very important task for the majority of IT organizations. Another conclusion is that

Half of the IT organizations consider it to be either very or extremely important over the next year for them to get better performing management tasks such as troubleshooting on a per-VM basis.

Management Challenges Associated with Cloud Computing

Even in the traditional IT environment[2] when the performance of an application is degrading the degradation is typically noticed first by the end user and not by the IT organization. In addition, when IT is made aware of the fact that application performance has degraded, the process to identify the source of the degradation can be lengthy.

Unfortunately:

The adoption of cloud computing makes troubleshooting application performance an order of magnitude more difficult than it is in a traditional environment.

One of the challenges associated with managing in any environment is that it is difficult to know the end-to-end path that packets take across a network. This management complexity comes in part from the distributed nature of IP. In particular, routers exchange reachability information with each other via a routing protocol such as OSPF (Open Shortest Path First). Based on this information, each router makes its own decision about how to forward a packet. There is, however, no single repository of routing information in the network. This lack of knowledge complicates tasks such as troubleshooting. The difficulty of knowing the path from origin to destination is greatly increased in a cloud computing environment because applications and services can be dynamically moved between servers both within and between data centers.

One of the fundamental issues relative to managing in a cloud computing environment is that the network topology becomes even more complex and hence understanding the end-to-end path becomes even more difficult.

In order to illustrate some of the other challenges of managing a cloud computing environment, assume that a hypothetical company called SmartCompany has started down the path of implementing private cloud computing by virtualizing their data center servers. Further assume that one of SmartCompany's most important applications is called BusApp and that the users of the application complain of sporadic poor performance and that BusApp is implemented in a manner such that the web server, the application server and the database server are each running on VMs on separate physical servers which have been virtualized using different hypervisors.

In order to manage BusApp in the type of virtualized environment described above, an IT organization needs detailed information on each of the three VMs that support the application and the communications amongst them. For the sake of example, assume that the IT organization has deployed the tools and processes to gather this information and has been able to determine that the reason that BusApp sporadically exhibits poor performance is that the application server occasionally exhibits poor performance. However, just determining that it is the application server that is causing the application to perform badly is not enough. The IT organization also needs to understand why the application server is experiencing sporadic performance problems. The answer to that question might be that other VMs on the same physical server as the application server

[2] This refers to an IT environment prior to the current wave of virtualization and cloud computing.

are sporadically consuming resources needed by the application server and that as a result, the application server occasionally performs poorly.

Part of the challenge associated with troubleshooting this scenario is that as previously noted, in most cases once an IT organization has virtualized its servers it looses insight into the inter-VM traffic that occurs within a physical server. Another part of the challenge is that as was also previously noted, each of the hypervisors comes with their own management system.

Staying with this example, now assume that SmartCompany has decided to evaluate the viability of deploying BusApp using either a public or hybrid cloud computing solution. For the sake of this example, consider two alternative approaches that SmartCompany might implement. Those approaches are:

1. Public Cloud Computing

   SmartCompany acquires BusApp functionality from a SaaS provider. The employees of SmartCompany that work in branch and regional offices use an MPLS service from a network service provider (NSP) to access the application, while home office workers and mobile workers use the Internet.

2. Hybrid Cloud Computing

   SmartCompany hosts the application and data base servers in one of their data centers and the web servers are provided by a cloud computing service provider. All of the users access the web servers over the Internet and the connectivity between the web server layer and the application server layer is provided by an MPLS service.

In order to monitor and manage either deployment, consistent and extensive management data needs to be gathered from the cloud computing service provider(s), the MPLS provider(s) and the provider(s) of Internet access. In the case of the first option (public cloud computing) similar management data also needs to be gathered on the components of the on-site infrastructure that are used by SmartCompany's employees and supported by the IT organization. In the case of the second option (hybrid cloud computing) similar management data also needs to also be gathered on both the on-site infrastructure as well as the web and application servers that are supported by the IT organization. In either case, effective tools are also necessary in order to process all of this data so that IT organizations can identify when the performance of the application is degrading before end users are impacted and can also identify the root cause of that degradation.

Another fundamental issue relative to managing either a public or hybrid cloud computing service is that the service has at least three separate management domains: the enterprise, the WAN service provider(s) and the various cloud computing service providers.

Cloud Management Solutions

The Growing Use of Cloud Networking Services

As pointed out in the section of this report entitled ***The Emergence of Cloud Computing and Cloud Networking***, a new class of solutions has begun to be offered by CCSPs. These are solutions that have historically been provided by the IT infrastructure group itself and include management, security, network and application optimization, VoIP, Unified Communications (UC) and virtualized desktops. This new class of solutions is referred to as Cloud Networking Services (CNS). That section of this report also presented the results of a survey in which The 2011 Webtorials Respondents were asked to indicate how likely it was over the next year that their company would acquire specific CNSs. The survey respondents were given nine types of services. Table 6 below highlights the interest that The 2011 Webtorials Respondents have in acquiring three specific CNSs.

Table 6: Interest in Cloud Networking Services

	Will Not Happen	Might Happen	50/50 Chance	Will Likely Happen	Will Happen
Security	39.0%	16.9%	16.9%	14.0%	13.2%
Network Management	38.8%	26.6%	7.2%	17.3%	10.1%
Application Performance Management	35.8%	28.4%	15.7%	12.7%	7.5%

One observation that can be drawn from the data in Table 6 is that:

Over the next year, more than a quarter of IT organizations will either likely acquire or will acquire security and/or management functionality from a CCSP.

Security as a Cloud Networking Service

Security is a very broad topic. That said, one of the largest, if not the largest sources of security vulnerabilities is Web based applications. Part of the growing security challenge associated with Web based applications is the continually increasing business use of social media sites such as Facebook and of major Webmail services such as Yahoo. A company could implement a simple acceptable use policy that either allows or denies access to these sites. However, such a policy ignores the fact that these sites typically provide a variety of functions, some of which fall into the acceptable use policies of a growing number of organizations. To deal with the evolving use of multi-faceted social media sites, a security based CNS needs to be able to allow access to a social media site such as Facebook, but block specific activities within the site, such as gaming or posting. Analogously, the CNS needs to have the granular controls to be able to allow users to send and receive mail using Yahoo, but block email attachments.

Another one of the security challenges associated with the use of Web based applications that is rapidly increasing in importance is the growth of malware. To protect against malware, a security based CNS should be able to identify sites that are either suspecious or are known to distribute malware. In order to be effective, a CNS that provides Web content filtering or malware protection needs a source of intellectual capital that identifies known and suspected vulnerabilities. To be effective, this source needs to be both dynamic and as extensive as possible.

One component of the value proposition of a CNS that provides web filtering and/or malware protection is the standard value proposition of any cloud based service. That value proposition is that a cloud based service has the potential to lower the cost of providing the service, reduce the time it takes to implement the service and give the company that is using the service access to functionality that they couldn't otherwise acquire. Another component of the value proposition of a CNS that provides web filtering and/or malware protection is that

Unlike a traditional security solution that relies on the implementation of a hardware based proxy, a security based CNS can also protect mobile workers.

The security based CNS does this by leveraging functionality that it provides at its cloud data centers as well as functionality in a software agent that is deployed on each mobile device.

In many cases, the best use of a CNS is as part of a hybrid solution. For example, in some cases, the IT organization already has functionality such web filtering or malware protection deployed in CPE at some of their sites. In this case, the IT organization may choose to implement a CNS just to protect the sites that don't have security functionality already implemented and/or to protect the organization's mobile workers. Alternatively, an organization may choose to implement security functionality in CPE at all of their sites and to also utilize a CNS as part of a defense in depth strategy.

Other situations in which a security centric CNS can serve to either be the only source of security functionality, or to compliment CPE based implementations include cloud-based firewall and cloud-based IPS services. Such a service should support equipment from the leading vendors. Given the previously mentioned importance of hybrid solutions, the service should allow for flexibility in terms of whether the security functionality is provided in the cloud or from CPE as well as for flexibility in terms of who manages the functionality – a CCSP or the enterprise IT organization.

Management as a Cloud Networking Service

As is the case with security, management is a very broad topic and hence it is possible to find a CNS that provides almost any possible form of management capability. For example, the preceding subsection discussed how a security based CNS could support mobile employees. In a similar fashion, a management based CNS can add value by helping IT organization to manage the burgeoning deployment of mobile devices.

One class of management based CNS is focused on managing specific types of devices, such as branch office routers, WiFi access points, mobile devices or security devices. In

some cases, the CNS supports customer-owned CPE from a wide range of vendors. In other cases, the CNS could be bundled with CCSP-owned devices located at the customer's premise. A variation on the latter approach involves a CNS vendor that provides devices, such as branch office routers, that have been specifically designed to be centrally managed from the cloud via a web portal. In this case, the vendor can move the device's control plane into the cloud in a manner analogous to the separation of control plane and data plane provided by OpenFlow, as discussed in the section of this report entitled ***The Emerging Data Center LAN***.

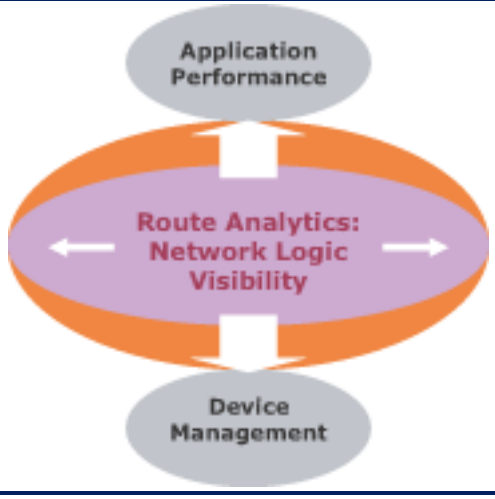
A second class of management based CNS is focused on managing other CNS services provided by a CCSP. These services typically are aimed at addressing the weaknesses in management capability generally associated with early CCSP provided services. For example, the initial wave of CCSP services came with little if any commitment on the part of the service provider relative to an SLA. One example of this class of management based service is a CNS that provides an enhanced level of management for a VoIP service that an IT organization acquires from a CCSP.

Route Analytics

As was previously mentioned, due to the distributed nature of IP it is sometimes difficult to know the end-to-end path that packets take across a network. While that is a challenge in any IT environment, it is a particularly difficult challenge in a cloud computing environment due to the dynamic nature of creating and moving virtual machines.

As shown in Figure 6, route analytics provides IT organizations and service providers with insight into the routing layer.

Figure 6: The Positioning of Route Analytics



The value proposition of route analytics is that

Route analytics provides visibility, analysis, and diagnosis of the issues that occur at the routing layer in complex, meshed networks.

A route analytics appliance draws its primary data directly from the network in real time by participating in the IP routing protocol exchanges. This allows the route analytics device to compute a real-time Layer 3 topology of the end-to-end network, detect routing events in real time and correlate routing events or topology changes with other information, including application performance metrics. As a result, route analytics can help both IT organizations and service providers determine the impact on performance of both planned and actual changes in the Layer 3 network.

Dynamic Infrastructure Management

A traditional environment can benefit from implementing dynamic infrastructure management. However, due to the challenges that are associated with cloud computing:

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 approach 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 database 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.

Virtualized Performance and Fault Management

In a traditional IT environment it is common to implement adaptive performance thresholding solutions that can identify systemic deviations from normal patterns of behaviour as well as time over threshold violations and can also automatically update thresholds based on changes to historic levels of utilization. That same capability is needed in a virtualized environment so that IT organizations can monitor the performance of individual VMs.

Virtual switches currently being introduced into the market can export traffic flow data to external collectors in order to provide some 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 intra-VM traffic data. Another approach to monitoring and troubleshooting intra-VM traffic is to deploy a virtual performance management appliance or probe within the virtualized server. This approach has the advantage of potentially extending the fault and performance

management solution from the physical network into the virtual network by capturing VM traffic at the packet level, as well as at the flow level.

While changes in the virtual topology can be gleaned from flow analysis, a third approach to managing a virtualized server is to access the data in the server's management system. Gathering data from this source can also provide IT organizations with access to additional performance information for specific VMs, such as CPU utilization and memory utilization.

Management Solutions Packaged with Converged Infrastructure

An increasingly popular approach to building cloud data centers is based on pre-integrated and certified infrastructure packages from either a broadly-based IT equipment vendor, a group of partners or a joint venture formed by a group of complementary vendors. These packages typically are offered as turn-key solutions and include compute, server virtualization, storage, network, and management capabilities. Other data center functions such as WOCs, ADCs, APM and security functionality may also be included.

One of the primary reasons why IT organizations implement a converged IT infrastructure is to reduce the overall complexity of a pervasively virtualized infrastructure. The reduction in complexity makes it feasible for IT organizations to fully capitalize on the virtualized infrastructure's inherent potential to serve as an agile, demand-driven platform that can deliver dynamic IT services with unprecedented levels of control, security and compliance, reliability, and efficiency. In order to realize the full potential of the converged IT infrastructure, the management system must provide a unified, cross-domain approach to automated element management, provisioning, change management and operations management. Some of the most critical aspects of managing a cloud data center include:

- **Integrated and Automated Infrastructure and Service Management:** Integrated management reduces the number of management interfaces that are involved in implementing administrative workflows. Automation allows services to be dynamically provisioned, modified or scaled without requiring time-consuming manual configuration across the various technology domains of the data center; e.g., compute, network, storage and security. The management suite should also include application and service level management capabilities that will support end-to-end SLAs. From an operational management perspective, the management system should provide additional capabilities, such as cross-domain root cause analysis and service impact analysis, in order to support the highest levels of service reliability.

- **Secure Multi-tenancy:** A robust multi-layer security architecture is required to ensure confidentiality and integrity of the services and the subscriber's data, particularly in a multi-tenant environment.

- **Support for Enterprise Co-Management:** The service management system should provide a web portal supporting the self-service provisioning of new services or the scaling of existing services. The portal should also include

dashboards that provide real-time visibility of application and service performance as well as the consumption of on-demand services. The service management system should also facilitate turning off resources such as VMs that are acquired from a CCSP when they are not needed so that the company using the resources does not incur unnecessary expenses.

- **Compatibility with Enterprise Cloud Implementations:** The efficiency of hybrid clouds is optimized where there is a high degree of consistency across the private and public portions of the solution in terms of the cloud management systems, the hypervisors and the hypervisors' management systems. This consistency facilitates the movement of VMs between enterprise data centers and service provider data centers, and this movement also enables the dynamic reallocation of cloud resources.

Management systems for a converged infrastructure typically support APIs for integration with other management systems that may be currently deployed in order to manage the end-to-end data center. These APIs can provide integration with enterprise management systems, automated service provisioning systems, fault and performance management systems and orchestration engines.

While IT departments or CCSPs can themselves achieve some degree of cross-domain management integration by leveraging available element manager plug-ins and APIs, ad hoc automation and integration across the end-to-end infrastructure is quite time-consuming and involves considerable specialized programming expertise. Therefore, the completeness and effectiveness of pre-integrated management functionality are likely to be two of the key differentiators among converged infrastructure solutions.

Cross-domain integrated management of the converged infrastructure will bring added benefits in those situations in which a single administrator has the authority to initiate and complete cross-domain tasks, such as provisioning and modifying infrastructure services. The use of a single administrator can eliminate the considerable delays that are typical in a traditional management environment in which the originating administrator must request other administrators in the other domains to synchronize the configuration of elements within their domains of responsibility. However, a well-known cliché describes the difficulty of realizing these benefits.

Culture eats strategy for breakfast.

That cliché refers to the fact that in many cases the culture of an IT organization resists any changes that involve changing the roles of the members of the organization. Exacerbating the challenge of the IT organization's resistance to change is the fact that, as was pointed out in the section of this report entitled ***The Emergence of Cloud Computing and Cloud Networking***, the culture of an IT organization typically changes very slowly.

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. Orchestration engines are available as standalone management products or as part of complete suites of management tools that are focused on the data center. In addition, the management systems that are integrated with converged infrastructure solutions typically include some orchestration capabilities.

By automatically coordinating provisioning and resource reuse across servers, storage, and networks, service orchestration can help IT organizations streamline operational workloads and overcome technology and organizational silos and boundaries. The value proposition of an orchestration engine is that

Orchestration engines use business policies to define a virtual service and to translate that service into the required physical and virtual resources that are 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 of the required infrastructure components. This includes routers, switches, security devices and core infrastructure services. The entire process can allow for the 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 available that are pre-configured to interface with certain families of infrastructure devices. Therefore, it is possible to think of the orchestration engine as providing some degree of management integration for non-converged infrastructure. As such, orchestration engines might be a highly desirable approach in those instances in which an existing heterogeneous (i.e., non-converged) data center infrastructure is being transitioned to perform as a cloud data center.

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[3]. These standards activities may also 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[4] and implemented by a number of companies in the security and network industries. It is a

[3] http://www.gridforum.org/
[4] http://www.trustedcomputinggroup.org/

publish/subscribe protocol that allows hosts to lookup meta-data and to 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 the automation and dynamic orchestration of not only security systems, but also other elements of the virtual data center. For example, IF-MAP could facilitate the 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.

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.Webtorials.com

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

nano|engine

Full application control at 10% of the cost

www.ipanematech.com

A unique technology that breaks the price/performance barrier to guarantee business application performance in branch offices

- For the first time it is possible to guarantee application performance with a device compatible with branch office constraints;
- The nano|engines fully integrate with the other components of Ipanema's ANS solution;
- Plug-and-Play devices, nano|engines are managed under SALSA;
- Real-time changes in network performance and each user's behavior are taken into account in real-time.

Algorithms embedded in the nano|engine automatically adapt to real-time changes as they happen on the network:

- Traffic from private data centers mixed with traffic from external public clouds;
- Hybrid networks combining MPLS and Internet;
- Unified Communications branch-to-branch flows;
- Virtual desktops and rich media delivery…

The nano|engine's ability to guarantee application performance at the branch maximizes productivity, prevents brownouts and protects the business.

Ultra compact **nano|engine** appliances are tailored for providing full application control with unmatched performance/price ratio in broadband branch offices.

The **nano|engine** devices target broadband branch offices and provide:

- Application aware, **per connection Control and dynamic QoS** for public and private application flows to guarantee an excellent and stable Quality of Experience to each user;
- **End-to-end visibility** of application performance of each flow with comprehensive KPIs and application quality scores;
- **Dynamic WAN path selection** among up to 3 networks for optimized control of multi-attached branches, local Internet breakouts and hybrid networks.

Self-managed, nano|engines are installed at the edge locations of the WAN, typically between the CPE router and branch office LAN. Fully "Plug and Play," nano|engines require no on-site configuration. They operate under control of the central management software, SALSA. Customers simply need to plug the nano in, and configuration and provisioning are managed by SALSA.

The nano|engine family fits particularly well in B to C sectors like retail, finance and hospitality, where slow response times to access customer data or delays in processing an order lead to customer dissatisfaction and loss of productivity. Nano|engines' ability to guarantee application performance prevents any brownouts and protects the business.

The nano|2 addresses branch offices with up to 20 users and 4 Mbps while the nano|5 targets branch offices with up to 50 users and 20 Mbps.