Five Reasons Why Network-Based Application Performance Management (APM) Is Necessary for Cloud-Enabled IT Environments

Abstract

Application owners and IT Operations managers need to reassess how they manage application performance when migrating their applications to virtual or cloud-based environments. While traditional products that gather host-based performance data face significant challenges in these environments, a new approach called network-based application performance management (APM) offers a much more manageable and flexible solution. With network-based APM, IT teams can keep track of a constantly changing environment, automatically track performance baselines and trends, diagnose and fix esoteric performance issues related to virtualization, and ensure compliance with service-level agreements (SLAs).

Table of Contents

Introduction and Background 2
Network-Based Application Performance Management 3
Five Reasons for Network-Based APM 4
   #1: Keeping track of assets in the cloud 5
   #2: Managing physical-to-virtual (P2V) migrations 5
   #3: Provisioning, optimizing, and capacity planning for applications in cloud environments 5
   #4: Achieving unified performance management across the datacenter and the cloud 6
   #5: Performance auditing and accountability in the cloud 6
Summary 6
About ExtraHop Networks 6
Introduction and Background

Virtualization and cloud computing have morphed quickly from business buzzwords into IT reality. While most IT teams have made the leap based on promises of efficiency and cost savings, the transition has proven much more difficult than anticipated. In these environments, the management and optimization of application performance is especially challenging because virtual machines offer less operational visibility than traditional deployments on physical infrastructure. This white paper provides an overview of how IT managers can regain lost visibility and keep up with the rapid pace of change by implementing network-based application performance management (APM) solutions that adapt to dynamic environments and do not require complex agent-based deployments.

Virtualization and cloud computing promise optimal resource utilization and agile scaling via dynamic workload allocation. However, these promises are easy to keep only when workloads are identical, small, and simple, with requirements that remain entirely flat over time—in other words, only in sterile laboratory conditions. Out in the unpredictable, unruly real world, IT teams find significant risks associated with moving new application workloads onto virtual machines initially, followed by ongoing risks entailed in moving these workloads around, in duplicating them, and in linking application and server virtual instances to storage and other critical resources.

Most IT professionals familiar with virtualization and cloud computing can enumerate at least some of the potential issues with these environments. The obvious challenges are local: Dropping a workload requiring a 3GHz CPU, 3GB of free RAM, or a very high-speed, low-latency I/O onto an underpowered, under-connected machine is clearly inadvisable (though it happens more often than people might admit). Less obvious but sometimes more critical long-term problems may occur, for example, in situations where an I/O-demanding workload has been moved and must now connect to its supporting database along paths that are adequate except at peak-load times. This scenario often results in delayed-action failure at arguably the worst possible moment for business productivity.

As application requirements become more complex, IT teams find an exponential increase in the likelihood of migrating to a virtualized environment that seems to work fine, but in fact hides numerous new threats. For example, consider the implications when multipath access to high-availability storage is configured (requiring available paths if failures occur), or when identically configured virtual servers running on separate hardware are configured for active/active mutual fallback. Compounding the management challenges in these situations is the fact that, although virtualization hypervisors and their increasingly single-pane-of-glass supervisory console applications are efficient for managing and deploying virtual private server (VPS) containers, they provide only limited visibility into what applications inside those containers are doing (see Figure 1).

Figure 1. Application performance details can be obscured within virtual environments.
In theory, these details are all supposed to be understood well before anything goes onto a VPS and under a hypervisor. Many authoritative blog posts and presentations with titles such as “Getting Ready for Enterprise Virtualization” advise that IT teams create an inventory of application instances and do a meticulous analysis of memory, CPU, I/O, resiliency, security, and other requirements and constraints in order to identify likely workloads for conversion and flag outliers as problematic. Unsurprisingly, however, this kind of extensive analysis is difficult, and most IT managers simply do not have the time or resources for due diligence. As a result, the IT organization cannot set stakeholder expectations about what reasonably can and cannot be virtualized. Insufficient analysis also can lead to situations where applications get virtualized when they should not. In a worst-case scenario, this can result in unacceptably poor performance, along with compromised resiliency and security. In a best-case scenario, accommodating the unique requirements of this application will push the virtualization or cloud architecture off its optima, to the point where the IT organization is not getting all the benefits it is paying for.

All of these issues are challenging for enterprises keeping virtualization and private cloud services in-house and are even more troubling for businesses that want to experiment with public cloud services of the IaaS (Infrastructure as a Service) or PaaS (Platform as a Service) varieties. Currently, there are huge deltas between common-sense business expectations for application performance and availability and the sort of generalized, dynamic, service-level guarantees and SLAs offered by cloud vendors. There are logical reasons for these disparities; most center on the incontrovertible fact that the enterprise customer—not the service provider—actually controls the application, with the exception of SaaS (Software as a Service) clouds.

Even where clouds are internal only, SLAs to the larger organization can be challenging to develop and adhere to closely. As noted above, it is the nature of clouds to be dynamic, change frequently, and constantly seek more and more efficient utilization of existing resources and better resource options. The system as a whole is a constantly moving target, demanding not only efficient mechanisms for determining momentary application performance but also methods for extending this scrutiny over time to ensure continued adequacy in the face of changing platform details and varying end-user demands.

Experts such as Bernd Harzog, of The Virtualization Practice, correctly point out the need for an outside-in view to gather deterministic performance metrics. For public clouds, Harzog suggests measuring the response time between the layer of software owned by the cloud provider and the layer of software owned by the cloud customer. For private clouds, he advocates measuring the time the virtualized infrastructure takes to complete the application request. In both cases, the goal is to measure actual application performance from the viewpoint of the user.

**Network-Based Application Performance Management**

Ascertaining real application performance for the cloud requires a solution engineered to help IT teams measure, manage, predict, visualize in detail, and relate application performance metrics to real end-user experience. With network-based application performance management (APM), a relatively new category of APM, IT teams gain much-needed visibility into how the modern high-density, virtualized, and cloud-extended datacenter is serving business needs. Network-based APM also helps to manage workloads with a much-improved ability to predict results, a significant benefit in dynamic environments.
Many vendors offer APM tools, and one of the ways in which solutions differ is in how they acquire the information they need to function. The most common approach, software instrumentation, installs agents onboard the devices they monitor. While this approach is adequate for certain types of monitoring and data acquisition, it adds computational and network overhead, not to mention management hassles. More importantly for virtual and cloud environments, software instrumentation contends for CPU and other machine resources with monitored workloads. This arrangement distorts the timing of the metrics gathered, due to clock skew. In addition, host-based instrumentation will miss subtle but serious communications faults such as virtual packet loss (see below).

Probe-based or inline monitoring on network legs is another classic strategy. Depending on the configuration and the need for permanent, 24/7 monitoring, equipment cost may be an issue for these tools, and inline monitors may add network latency. Data retrieved by such legacy methods is usually limited from an application performance standpoint; consisting primarily of L4 TCP packet header information combined with NetFlow or SNMP data from network elements. While useful for some types of analysis, this lack of application detail makes it impossible to trace certain kinds of interactions or interpret complex data flows in real-world enterprise datacenter scenarios.

Network-based APM is the best approach for data-acquisition for virtualized or cloud environments as it captures all application transactions flowing over the network at wire speed via a network tap, VACL capture, SPAN, or similar method of attaching to a central switch. Data then can be examined in totality, viewing interactions from L2 to L7 of the OSI stack. Capturing traffic out of line in this way is 100 percent passive, giving an accurate view of production data flows in virtual environments. A challenge to scale, with this arrangement, is processing application-level transactions at wire speed—requiring a proprietary networking microkernel—and recording metrics, samples, and other reportable data to disk without falling behind or overwhelming the datastore. For this purpose, leading network-based APM products feature specialized direct block-transfer disk management systems whose performance scales radically better than conventional relational databases.

Network-based APM systems are mostly self-configuring through auto-discovery of applications from their data flows and identification of IP addresses, VLANs, and other criteria relevant to grouping.

Additional information can be supplied by the user to label and group transactions optimally for analysis. Network-based APM systems can produce a wide range of useful metrics out of the box, and optional protocol modules and associated logic help to enable tracing and deep analysis of specific L7 application data flows. Modules may be offered for enterprise directory transactions, database communications, network-attached storage protocols, web application protocols, and other areas of interest. Web-based user interfaces give IT personnel a window into current application performance status and historical records, and wizard-like tools assist in investigating undesirable application or network behavior.

### Five Reasons for Network-Based APM

The following five reasons make network-based application performance management a must-have technology for enterprise IT, equally valuable for those teams running conventional datacenters with dedicated infrastructure, contemplating a transition to virtualization and the cloud, or already running virtualized and cloud-based systems.
#1: Keeping track of assets in the cloud

Network-based APM, with its auto-discovery and real-time reporting, is a powerful solution for verifying the existence, status, and health of applications in real time and serves as a useful complement to solutions that enable IT teams to visualize and manage connectivity, hardware, hypervisors, and workloads. Combining both technologies frees staff from routine change-management documentation and document maintenance, freeing bandwidth to tackle more important and strategic work, while limiting virtual machine sprawl. Figure 2 shows a list of auto-discovered devices with an activity map report based on the same list.

![Figure 2](image_url)

*Figure 2. Network-based APM can automatically discover both physical and virtual devices and generate reports that simplify IT management.*

#2: Managing physical-to-virtual (P2V) migrations

Because network-based APM can monitor and characterize application status and performance on both conventional and virtual platforms, it can be ideal for performing necessary discovery and due diligence about application characteristics prior to planning and implementing a virtualization strategy, for identifying workload candidates for conversion (and flagging problematic workloads), for ensuring that continuity and adequate performance are maintained as applications are migrated, and for forensic problem solving.

#3: Provisioning, optimizing, and capacity planning for applications in cloud environments

Network-based APM enables you to apply a single set of truly representative performance metrics to all applications, helping you to understand native performance costs associated with virtualization and cloud deployment and to identify problems introduced in the process of configuring applications for migration or emerging from architectural conflicts, incompatibilities, or insufficient planning. The ability to deeply compare the performance of applications running on conventional servers with the performance of the same applications running in a virtualized environment can be extremely helpful in pre-migration confirmation.
testing and later optimization as well as in right-sizing datacenter bandwidth and storage to reflect projected demands of virtualized workloads.

#4: Achieving unified performance management across the datacenter and the cloud

Virtualization can trigger problems of great subtlety—problems almost never encountered in conventional datacenters—that can have a devastating effect on application performance. Virtual packet loss, for example, occurs when the hypervisor oversubscribes the number of physical CPUs with multiple virtual machines vying for the simultaneous execution of deadline-sensitive TCP code. This results in intermittent degradation of application performance that often goes undetected or misdiagnosed by traditional network and application monitoring tools. Easy fixes exist for these devastating conditions, including resynchronization and modifying overly restrictive timing thresholds for congestion-avoidance behaviors. However, before you can implement these modifications, you need to identify and understand the problem. Network-based APM systems can detect these issues by a combination of application awareness and advanced TCP analysis.

#5: Performance auditing and accountability in the cloud

Network-based application performance management offers a way out of the present quagmire surrounding cloud-service SLAs. Network-based APM provides a client-side mechanism for quantifying acceptable application performance before a cloud transition occurs—contributing meaningful data for framing SLAs. It offers an on-premises mechanism for monitoring performance to ensure SLA compliance and for rapidly correcting issues. Additionally, in the client’s hands, network-based APM offers the means to validate compliance or identify failures.

Summary

These five advantages are just some of the many reasons why network-based application performance management is becoming a cornerstone technology for IT departments managing virtualized and cloud-based applications. For more information about the capabilities of network-based APM, visit the ExtraHop website to find out about ExtraHop application performance monitoring solutions.

About ExtraHop Networks

ExtraHop Networks was founded in early 2007 by Jesse Rothstein and Raja Mukerji, engineering veterans from F5 Networks and architects of the BIG-IP v9 product. The company’s pioneering Application Delivery Assurance system is the industry’s first completely passive network appliance that provides application-level visibility with no agents, configuration, or overhead. The privately held company is headquartered in Seattle, Washington, and funded in part by the Madrona Venture Group.

To learn more about ExtraHop and our innovative Application Delivery Assurance system, visit us at www.extrahop.com.

© 2011 ExtraHop Networks, Inc. All rights reserved. ExtraHop and the ExtraHop logo are registered trademarks of ExtraHop Networks, Inc.