Application Performance Management: New Challenges Demand a New Approach

By ***Jim Metzler, Ashton Metzler & Associates***
Distinguished Research Fellow and Co-Founder,
Webtorials Editorial/Analyst Division

Introduction

Historically IT management has focused on individual technology domains; e.g., LAN, WAN, servers, firewalls, operating systems, etc. This approach made the implicit assumption that if each technology domain was performing well, that the application was performing well. Unfortunately, too often that assumption isn't correct as it is possible for each technology domain to be working well and yet the application performance as seen by the end user is unacceptable. One way that this could happen is that due to the overall complexity of the environment, the cumulative end-to-end latency between a user and the back-end database becomes so high that backend database queries time out.

Over the last five years, in an attempt to improve service quality as seen by customers and business partners and to also reduce cycle times and service delivery costs, many IT organizations have begun to reduce their emphasis on managing just individual technology domains and to implement Application Performance Management. While the vast majority of IT organizations have already implemented Application Performance Management, many IT organizations have not yet fully realized the promised benefits of Application Performance Management. One indication of that unrealized promise is that as highlighted in the 2012 Application and Service Delivery Handbook[1], in the vast majority of instances when the performance of an application is degrading, it is noticed first by the end users and not by the IT organization. The handbook also underscored the importance of IT organizations becoming more proficient at Application Performance Management by presenting recent market research that showed that in roughly two thirds of the instances in which a business critical application is performing badly, the company loses revenue.

This white paper has three goals. One goal is to describe the traditional approach to Application Performance Management and to identify some of the key weaknesses of that approach. The second goal is to describe some of the factors that are causing IT organizations to make some fundamental changes in terms of how they implement Application Performance Management. The third goal is to identify the key attributes of the next generation of Application Performance Management. To help to achieve those goals, this white paper will incorporate the results of a survey that was distributed in April 2012 to the subscribers of Webtorials. Throughout this white paper, the 190 IT professionals that completed that survey will be referred to as ***The Survey Respondents***.

[1] http://www.webtorials.com/news/2012/07/2012-application-service-delivery-handbook-1.html

Traditional Application Performance Management

As application architectures evolved from client/server to n-tier web-based applications, application functionality on the server was usually divided up into two or three segments. These segments are the web front-end (a.k.a., presentation tier or tier1), the business logic processes (a.k.a., the application tier or tier 2) and database operations (data tier or tier 3).

In an n-tier web based application, the user interacts with the presentation tier and the presentation tier in turn communicates to the logic tier, which in turn communicates to the data tier. Each tier uses servers that are optimized to the characteristics of their tier. A presentation tier server, for example, is optimized for network I/O and web traffic; e.g. multiple network cards, large network buffers, etc. A logic tier server is optimized for logic computations; e.g. high speed CPUs, large memory size, etc. A data tier server is optimized for database operations; e.g. multiple disk I/O controllers, large disk cache, large memory size, etc.

Traditional Application Performance Management was typically performed separately from network performance management. For example, when application degradation occurs, the triage process typically assigns the incident to either the network or server areas for resolution. Each area then examines their basic internal measurements of network and server performance and a pronouncement is made that the source of the issue is either the network or the application server or both or neither. Since these tasks are typically done by different parts of the IT organization using different toolsets and management frameworks, it is quite possible that conflicting answers are given for the source of application performance issues.

Traditional Application Performance Management solutions have limitations. Those limitations include the fact that that traditional Application Performance Management solutions:

- Do not integrate network performance data between tiers to monitor and analyze application performance problems.

- Cannot attribute CPU, disk I/O, network I/O nor memory utilization to specific classes of transactions. Only aggregate server performance information is available.

The Current Environment

As previously mentioned, a survey was recently distributed to the subscribers of Webtorials. As part of that survey, The Survey Respondents were given twenty management tasks and were asked to indicate how important it was to their organization to get better at each task over the next year. They were given five options:

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

Table 1 lists the three management tasks that were the most important to The Survey Respondents and the percentage of The Survey Respondents that indicated that getting better at those tasks was either very or extremely important.

Table 1: Primary Management Challenges	
Management Task	**Percentage**
Rapidly identify the root cause of degraded application performance	67.9%
Identify the components of the IT infrastructure that support the company's critical business applications	62.9%
Obtain performance indicator metrics and granular data that can be used to detect and eliminate impending problems	51.6%

Some of the observations that can be drawn from the data in **Table 1** include:

- Rapidly identifying the root cause of degraded application performance has been a critical issue for IT organizations for many years. It remains a critical issue because even though IT organizations are continually getting better at it, its importance continues to grow as an organization's key business processes increasingly rely on applications and the supporting infrastructure.

- The importance of identifying the components of the IT infrastructure that support the company's critical business applications reflects the fact that most IT organizations are focusing their Application Performance Management efforts primarily on their company's business critical applications.

- The importance of obtaining performance indicator metrics and granular data that can be used to detect and eliminate impending problems demonstrates that IT organizations are continually attempting to implement a more proactive approach to Application Performance Management.

In addition, the fact that such a high percentage of The Survey Respondents stated that it was either very or extremely important for their organization to get better at the tasks listed in **Table 1** is further evidence that the majority of IT organizations have not yet fully realized the promised benefits of Application Performance Management.

Emerging Challenges

The data in **Table 1** highlights some of the Application Performance Management related challenges that IT organizations face in the current environment. The current IT environment, however, is very dynamic and a number of emerging trends are going to both further exacerbate the existing Application Performance Management related challenges as well as create some new ones. Some of the key emerging trends that IT organizations need to consider when modifying their approach to Application Performance Management are:

- Server Virtualization
- Cloud Computing
- Mobility

Server Virtualization

In order to quantify the interest that IT organizations have in server virtualization, 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, which are shown in **Table 2**, highlight the strong and growing interest that IT organizations have in virtualizing their data center servers.

Table 2: Deployment of Virtualized Servers

	None	1% to 25%	26% to 50%	51% to 75%	76% to 100%
Have already been virtualized	18%	30%	25%	16%	11%
Expect to be virtualized within a year	11%	28%	24%	25%	12%

The way to read **Table 2** is that currently 27% of IT organizations (i.e., 16% + 11%) have already virtualized the majority of their data center servers and that percentage is expected to increase to 37% within a year. The fact that the percentage of IT organizations that have already virtualized the majority of their data center servers will increase significantly over the next year while the percentage of IT organizations that have not virtualized any servers will drop significantly over the next year underscores the ongoing strength and breadth of the movement to adopt server virtualization. The strength and breadth of this movement means that the impact of the management challenges described below will continue to grow and that they will continue to demand an effective response from IT organizations.

Server virtualization creates a number of management challenges. One such challenge is that a virtual switch (vSwitch) is used to switch the traffic that goes between virtual machines (VMs) on a physical server. In the vast majority of instances, vSwitches have minimal management functionality and so IT organizations lose all visibility into the inter-VM traffic. As a result, IT organizations lose the ability to manage the performance of multi-tiered applications that reside entirely on a virtualized server.

Another challenge is that until recently, all aspects of IT management, including Application Performance Management, was based on the assumption that the IT organization performed tasks such as monitoring and troubleshooting on a server-by-server basis. Now, given the widespread adoption of server virtualization, the traditional approach to Application Performance Management must change to where it is focused on a virtual machine VM-by-VM basis. Failure to do so means that IT organizations cannot perform key Application Performance Management tasks such as baselining the performance of an application that runs in a VM.

Another assumption that underpinned the traditional approach to Application Performance Management was that the data center environment was static. For example, it was commonly assumed that an application resided on a given server, or set of servers, for very long periods of time. However, part of the value proposition that is associated with server virtualization is that it is possible to migrate VMs between physical servers, both within the same data center and between disparate data centers. This ability to migrate VMs between physical servers is just one example of why IT organizations need to adopt an approach to Application Performance Management that is based on the assumption that the components of a service, and the location of those components, can and will change frequently. Not adopting that assumption means that IT organizations lose the ability to manage the performance of an application that has been moved between physical servers.

As described above, if IT organizations don't change their approach to Application Performance Management to respond to the management challenges that are associated with server virtualization then they are faced with an unacceptable choice. That choice is to either stop implementing server virtualization and give up all of the attendant benefits or accept that they will have little if any ability to manage a virtualized server environment, even though that environment will increasingly be used to support customer facing, revenue generating applications. Making the later choice significantly reduces the value of the IT organization.

Cloud Computing

There are three forms of cloud computing[2]: public, private and hybrid. While there is not a definition of cloud computing that is universally agreed to, it is generally agreed to that the goal of cloud computing is a dramatic improvement in the cost effective, elastic provisioning of IT services. It is also generally agreed to that some of the key characteristics of a cloud computing solution are virtualization, standardization, self-service provisioning and automation.

The most common categories of public cloud computing are Software-as-a-Service (SaaS) and Infrastructure-as-a-Service (IaaS). The initial set of SaaS solutions consisted primarily of enterprise applications such as customer relationship management and supply chain management. While enterprise applications still represents the majority of the SaaS market, today it is also possible to acquire a wide range of basic IT functionality from a SaaS provider. This includes:

- Voice over IP
- Unified Communications
- Security

The initial set of IaaS solutions that were brought to market by IaaS Providers were the basic compute and storage services that are necessary to run applications. However, the IaaS market is highly dynamic and IaaS providers are deploying myriad new services including:

- Disaster Recovery
- Virtual Private Data Centers
- High Performance Computing

The phrase *private cloud computing* refers to IT organizations implementing the previously mentioned characteristics (e.g., virtualization, standardization) within their own environments. The phrase *hybrid cloud computing* refers to solutions that are a combination of public cloud and private cloud solutions. For example, a company that has an n-tier application may decide that in order to protect their business critical data that they will run the application and database tiers in their private cloud, but that in order to have the Web tier be close to their users, that they will host the web tier at an IaaS provider's facilities. Alternatively, a company may decide that it can save money by building its infrastructure to support the average demand for service and then burst to a public cloud provider in those periods of peak demand.

Many of the management challenges that are associated with private cloud computing are the previously discussed challenges that result from server virtualization. Managing server virtualization is also a challenge for both public and hybrid cloud computing. However, the adoption of those forms cloud computing creates a new set of management challenges. Some of these new challenges stem

[2] A discussion of the varying forms of cloud computing can be found in the 2011 Cloud Networking Report, http://www.webtorials.com/content/2011/11/2011-cloud-networking-report.html

from the fact that IT organizations are typically held responsible for the performance of these public and hybrid cloud solutions even though in most cases they don't have the same access to the enabling IT infrastructure that they would have if the application was entirely intra-company. Another part of the new set of management challenge stems from the sheer complexity of the public and cloud environments. What this complexity means is that in order to manage end-to-end in either a public cloud or a hybrid cloud environment, management data must be gathered from the enterprise, one or more Network Service Providers (NSPs) and one or more cloud computing service providers.

The downside of IT organizations not evolving their approach to Application Performance Management to respond to the challenges that are associated with cloud computing is similar to the downside of IT organizations not responding to the challenges of server virtualization. In this case, IT organizations are faced with the choice to either stop implementing cloud computing and give up all of the attendant benefits or accept that they will have little if any ability to manage a cloud computing environment, even though that environment will increasingly be used to support customer facing, revenue generating applications. Making the later choice significantly reduces the value of the IT organization

Mobility

Over the last few years there has been a dramatic growth in the number of mobile employees. Up until a couple of years ago, most IT organizations attempted to control the types of mobile devices that could access the corporate network. For example, it was common two years ago for IT organizations to either not allow any user owned devices to access the network or to standardize on one device, usually a Blackberry, which would be allowed network access. The last few years, however, have seen a dramatic shift driven by the desire on the part of employees to bring an increasing array of smartphones and tablet computers to work and to use those devices to access applications. Now most organizations have adopted the Bring Your Own Device (BYOD) movement and as a result, these organizations now allow a wide variety of user owned devices to access their network.

A key concern relative to supporting mobile workers is how the applications that these workers access has changed. At one time, mobile workers tended to primarily access either recreational applications or applications that are not delay sensitive; e.g., email. However, in the current environment mobile workers also need to access a wide range of business critical applications, many of which are delay sensitive. This shift in the types of applications that are accessed by mobile workers was highlighted by SAP's announcement[3] that it will leverage its Sybase acquisition to offer access to its business applications to mobile workers.

One of the key management implications of the extensive and growing adoption of mobile devices is that IT organizations need to enable the company's business unit managers to use their mobile devices to access metrics about the overall health of the applications that they rely on to drive their business units. Another key management implication of the adoption of mobile devices is that any Web site that is customer facing, whether that is for sales or customer service, must run flawlessly when accessed using a wide array of smartphones and tablets. In order to ensure this flawless behavior, IT organizations need the ability to monitor how the Web site is performing when accessed using a wide range of popular browsers; e.g., Safari 4, Android 2.3 and iOS4.

Whether or not IT organizations change their approach to managing mobility, the use of a growing array of mobile devices to access business critical applications will continue to expand. If IT organizations

[3] Wall Street Journal, May 17, 2012, page B7

don’t evolve their approach to managing mobility, they will significantly reduce the value that they provide to the company’s business unit managers.

Summary and Call to Action

While the vast majority of IT organizations have already implemented Application Performance Management, only a small minority of IT organizations have realized the promised benefits. The inability of IT organizations to fully realize the promised benefits of Application Performance Management has a dramatic business impact. This follows because in the majority of instances in which the performance of a business critical application is performing badly, the company loses revenue.

While IT organizations try to modify their approach to Application Performance Management to respond to the traditional management challenges, a number of emerging trends are going to both further exacerbate the existing Application Performance Management related challenges as well as create some new ones. These trends include:

- Server Virtualization
- Cloud Computing
- Mobility

Relative to their implementation of Application Performance Management, IT organizations are at a crossroads and can choose one of two options. One option is to adopt an approach to Application Performance Management that effectively responds to the existing and emerging challenges. The other option is to continue with the current approach to Application Performance Management. Maintaining the current approach to Application Performance Management means that degraded application performance will continue to cause companies to lose revenue. It also means that IT organizations will only be able to provide best effort support to key initiatives such as server virtualization, cloud computing and mobility. As a minimum, supporting these initiatives on a best effort basis will result in more instances of degraded application performance, which in turn will result in more instances in which the company loses revenue. In some cases, supporting these initiatives on a best effort basis will result in dramatic user dissatisfaction and the possible abandonment of those initiatives. In either case, supporting these initiatives on a best effort basis will result in a significant reduction in the perceived value of the IT organization.

Because IT organizations are under continual pressure to demonstrate value to the company's business units, the new approach to Application Performance Management that IT organizations need to adopt must have a focus not on technology domains, but on business transactions and must provide the ability to perform deep-dive problem diagnostics. In order for this new approach to Application Performance Management to enable IT organizations to achieve their goal of being able to rapidly identify the root cause of degraded application performance, the new approach must include the ability for end-to-end monitoring of every end-user transaction from the end user through to the mainframe or to the cloud. Since a SaaS provider will process a growing number of business transactions, having the ability to do synthetic transaction monitoring of SaaS providers is a key component of the Application Performance Management functionality that IT organization need to have.

In order for this new approach to Application Performance Management to achieve the other two goals identified in **Table 1**, the new approach must also enable IT organizations to gather management data that links the performance of a transaction as seen by the end user with all of the applications that are involved in supporting the transaction as well as all of the supporting infrastructure components. This management data must be gathered across both physical and virtual infrastructure components and from both within the enterprise as well as from network and cloud service providers.

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.