



## **New Frontiers in Web Performance Monitoring**

By Corey Ferengul, Senior Program Director, META Group

Fortune 1000 companies worldwide are realizing that their bottom line can be heavily influenced by the performance of their Web-based applications. With so many of these applications now interfacing directly with customers, companies must ensure that they deliver the performance that customers demand and provide a positive end-user experience. To achieve this goal, IT groups and lines of business (LOB) must begin working more closely together. More important, IT groups, which typically provide the LOB with highly technical information about service levels, will need to start providing information that quantifies the end-user experience.

In their quest to quantify the end-user experience, LOB and IT groups now have the benefit of a new metric—end-user response time. This response time measures the performance of an application from the end-user perspective (e.g., the amount of time required for the end user to receive a response or to execute a business transaction). End-user response time is valuable to both LOB and IT groups. For the LOB, it identifies clearly the application's level of performance and availability as experienced by the end user. For IT, end-user response time serves as a barometer of performance and an early indicator of problems within the application.

### **Performance Monitoring—The Key to Measuring End-User Experience**

META Group believes that organizations that do not monitor and manage response times leave their e-business applications at risk for serious performance problems. First, by not monitoring, organizations do not have a clear or complete picture of their application's performance and cannot differentiate between good or poor performance. Second, they cannot isolate potential and existing bottlenecks, since this data can be collected only through monitoring. As a result, failing to monitor application response time can result in system outages, slowdowns and lost revenue.

Multiple methods of collecting and monitoring response times are now available. The most common types of monitoring are active (or synthetic) and passive (or observational). By using these two types of monitoring, IT can broaden its coverage to gain an accurate and complete picture of the end-user experience. Organizations need to be educated on both types, as each will play a role in the overall application management plan and service level reporting.

This paper will provide an overview of active and passive monitoring, including the relative advantages and disadvantages of each.

### **Active Monitoring**

Active monitoring is currently the most widely used method of response time monitoring. It creates an emulated (or simulated) client and executes transactions against the application. These transactions are executed on a regular basis (which could be every 5 minutes to once a day) and provide a clear baseline of expected performance for key transactions. These transactions are measured in detail (object load times, page load times, wait time, from which server) and also provide availability data (as a transaction will fail if the site is not available). Most organizations identify several key transactions (we recommend 5 to 10 minimum – as this varies from situation to situation) and create simulated clients for each. Thresholds are then set on the result (usually derived from the baseline), and real-time alerts are sent if a transaction execution time exceeds that threshold. This provides an excellent early warning signal to IT, enabling them to attack any potential problem. The resulting metrics are also used to provide the LOB service level reports (SLRs). With a single collection method, both groups then can benefit and work from a common ground.

Response Time Monitoring Advantages		
Active	Passive	
X		24x7 monitoring
X		Baselining
X		Proactive
X		Constant controlled collection
X	X	Root cause analysis
X	X	Multiple viewpoints
	X	Based on actual user data
	X	Detailed end user data
	X	PC time captured

**The advantages of active monitoring include:**

- **24x7 monitoring**  
When active agents are deployed they are set to execute the simulated transaction on a pre-defined interval. This results in round-the-clock monitoring, necessary in e-business applications that are consumed 24x7.
- **Constant controlled collection**  
It is necessary to have a constant data flow to ensure accurate evaluation of the application's performance. Also, at peak usage times (or seasons) it may be required to collect data more frequently.
- **Baselining ability**  
With the constant controlled results, IT groups can baseline performance and establish thresholds. This is needed to identify when the application is

experiencing performance trouble and to set alarm triggers accordingly.

- **Proactive in nature**  
Because the data is based on simulated clients and executed constantly, it is likely to identify a slowdown before it becomes critical or is perceived by external users. This enables IT to proactively take action and correct the problem before it impacts end users.

- **Multiple points of view collected**  
End users can be located anywhere worldwide, therefore it is important to get a view that most closely represents their experience.

**The disadvantages include:**

- **Not based on true end-user data, rather a representative baseline**  
As the execution is simulated, it limits the interpretation of the results. It therefore can miss possible aspects of what the end user is experiencing (e.g., slow PC).
- **Limited insight into the back-end infrastructure performance**  
Active monitors record the time required by a transaction, but they do not specify which back-end infrastructure delivered the entire transaction (e.g., database). (We are beginning to see tools on the market adding these metrics.)
- **Difficult to execute "entire" transaction**  
Simulated agents cannot necessarily complete a transaction (e.g., actually buy an item) without interfering with the normal operations of an application. Hence, the data can possibly miss an element of performance. Some organizations attempt to address this issue by using a ghost transaction, one that is a known ghost user.
- **Creation and maintenance of scripts**  
The scripts that are used to simulate the transactions must be built and updated as the application changes.

**Deployment of Active Monitoring Tools**

Deployment of the active/simulated agents needs to be carefully planned to ensure the most successful usage of this method. Because an e-business application is used both outside and inside an organization, the monitoring strategy needs to cover both of these perspectives. Therefore, the META Group recommends that organizations collect active data both internally and externally (inside and outside of the firewall). External monitoring is more complicated, however, since the end users may be located anywhere worldwide. For the external collection, we believe that the data should be collected from a number of external points that best represent the potential end users' locations.

## Passive Monitoring

Many organizations that have implemented active monitoring now are looking to expand their insight into application performance and enhance their management data. Passive monitoring helps them achieve this goal by collecting data from actual end-user activity. This data represents the actual experience and thereby provides views into all consumed aspects of an application. All data collected on end-user activity is stored in a common database, centralizing management. Passive monitoring often produces more performance data since there is no limit to where an end user may go within an application. Organizations utilizing passive monitoring must also plan to consolidate or harvest the large volumes of data. The resulting data is often combined with usage analysis data to provide the LOB a robust view of end-user activity or habits within the application.

Multiple methods exist for passive monitoring, including the use of client-based agents and network-level collection. Client-based agents often use a browser plug-in, which is difficult to implement and maintain. Other client-based agent methods embed a Java script into a Web page, which then acts as the collection agent. A less complicated passive monitoring method is network-level collection. This method collects data at the network layer, via packet analysis. Network-level collection is complementary to many common network management techniques and is therefore more likely to be adopted by an organization. Network-level collection is also very useful in determining the architectural tier that is responsible for slowdowns in the environment.

### The advantages of passive monitoring include:

- Based on actual end-user data  
Identifying what the end users are truly seeing not only validates the results of active monitoring, but also is useful for troubleshooting. Additionally, this data is useful when combined with application usage data, providing an even clearer picture of what the end users are doing.
- High degree of detail on end user  
Not only does passive monitoring clearly show the level of performance end users are experiencing, it identifies precisely when/where they are experiencing it. This passive data is necessary because it is impossible to have an active agent for every page or transaction.
- Capture PC time  
Some performance problems result from the end user's own machine. Understanding that aspect will lead to an understanding of whether the performance slowdown is widespread or limited to specific end users.

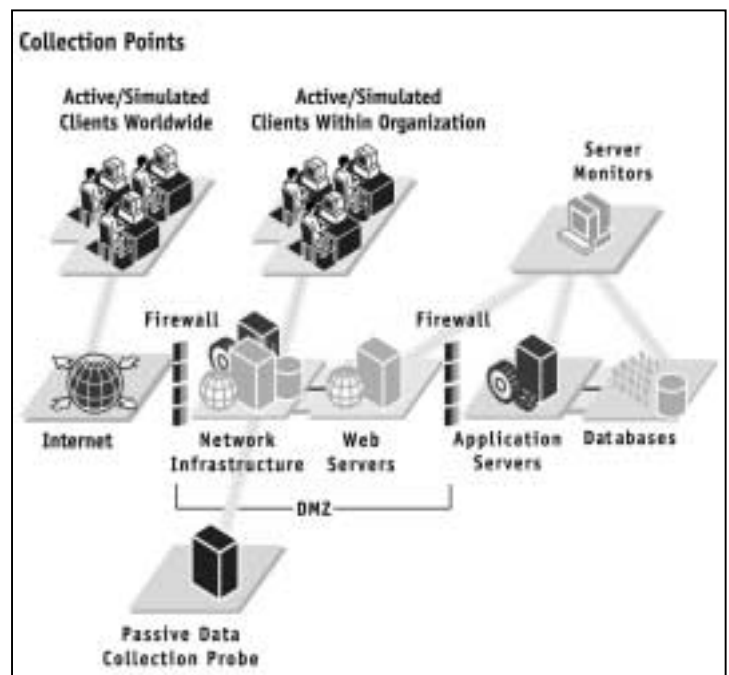
- Aids in root-cause analysis

When a performance slowdown is detected, passive monitoring data provides details on where end users are experiencing bottlenecks. With a more complete representation of the overall health of the infrastructure (when compared to active), passive monitoring data will enable problem isolation.

### The disadvantages include:

- Data is opportunistically collected

Different end users connect at different times and locations. End users also do not access the same elements of an application. The result is that the data collection is unpredictable; there is no guarantee that a given part of the application will be accessed or exercised. This leads to potential "blind spots" within the application.



- Reactive by nature

When a threshold is triggered, it is a reaction to an actual end-user problem.

- Potentially excessive data

Data is generated from every end user call to every part of the application. The resulting data is massive, and it is quite difficult to analyze the results without a clear plan of correlation, aggregation and data reduction.

### **Deployment of Passive Monitoring Tools**

When deploying passive monitoring tools, it is necessary to understand what view of the enterprise is needed—the client’s machine, application or network. Each of these methods is beneficial and offers something unique. Monitoring from the client perspective, for example, offers insight into the end user’s machine. Monitoring from the application perspective (JavaScript) provides a complete view of the transactions path. And monitoring from the network perspective offers the clearest view into the application infrastructure. To ensure success with passive collection, organizations need to create a data aggregation and correlation strategy. Aggregation with usage data is also beneficial and should be shared with the LOB.

### **Deployment Strategies for Complete Application Performance Monitoring**

Organizations will need to deploy both passive and active monitoring methods in order to obtain the most complete and accurate views of the end-user experience. Only then can they effectively manage application performance. Which method is deployed first will depend both on the type of application being monitored and the monitoring goals.

META Group sees a majority of business-to-consumer applications initially deploying active monitoring in order to gain the controlled baseline view of performance. External deployment of active monitoring (e.g., outside the corporate firewall, minimally at the service provider’s location) should be deployed early in the life of business-to-consumer and/or business-to-business applications. Collection should be from many geographic points (as appropriate), with a limited number of transactions. In business-to-business applications, not as many external monitoring points are required, but having the correct ones, representative of the user audience, is critical.

Organizations also should attack monitoring internally (e.g. inside the firewall) by deploying passive monitoring tools. This method provides coverage of additional transactions with shorter intervals of monitoring. It also allows internal control over the monitoring tool, enabling end users to interact with the technology directly when troubleshooting. We believe that application-centric passive monitoring as well as network-centric passive monitoring should be deployed behind the firewall. At this point, troubleshooting and root-cause analysis will become more important (as the application has matured some and usage has predictably increased). A strategy needs to be in place to ensure that all the data collected is merged with each other (as well as other monitoring data) to allow the most complete view of the end-to-end operation. META Group believes organizations need to implement all types of response time monitoring in order to ensure their application’s success.

### **Summary**

Organizations need to deploy response time management tools for their e-business applications. Failing to do so will lead to outages, poor end-user experiences and lost revenue, since there will be no understanding of how the application is performing from the end-user perspective. With so many different methods of monitoring available to them, organizations need to understand the advantages and disadvantages of each. Moreover, both active and passive monitoring will eventually need to be deployed within an enterprise. Finally, organizations must plan to integrate many tools, as their synergy will be powerful for complete performance management.