Get the Whole Picture

Why Most Organizations Miss User Response Monitoring—and What to Do About It

Nimsoft Whitepaper

Get the Whole Picture

Table of Contents

Nimsoft Whitepaper

Get the Whole Picture

Executive Summary

End user response. To borrow a phrase, it's where the rubber meets the road. You can be armed with vast amounts of performance metrics, but if you don't know what users are actually experiencing, you don't have the real performance picture. While this measure is critical, it is one many organizations fail to consistently capture. Why? This guide looks at the challenges of user response monitoring, and it shows how you can overcome these challenges and start to get a real handle on your infrastructure performance and how it impacts your users' experience.

Introduction

A co-worker just called to say her email is down. You look at the server and network stats and everything seems to working fine. What's the real story? Is it user error, or is some other system within the infrastructure an issue? How do you know?

With all the intricacies and interdependencies within your organization's infrastructure, there's a lot to keep track of. Ironically, though, the one measurement that ultimately really matters is the one that many organizations are still not monitoring: end user response. How long does it actually take a user to submit a request to an application and have that request completed and returned?

Employees can only be productive if the applications and services they rely upon are responsive and always available when they need them. Customers and would-be customers demand fast responses, or they'll go elsewhere. It is end user response measurement that ultimately provides the most precise measure of user experiences, and thus it ultimately has the biggest implications for the business.

Why haven't more organizations been monitoring this crucial aspect? How can organizations begin to leverage end user response monitoring to better ensure optimal service levels and user productivity? This guide looks at the challenges that have kept organizations from successfully implementing end user response monitoring. It then offers guidance for overcoming these challenges. Finally, it looks at the key deployment characteristics that are required to make monitoring solutions effective in your organization.

Roadblocks to Monitoring End User Response

Historically, IT organizations have had success in monitoring specific applications, servers, or databases. Yet, as vital as it is to monitor such metrics as the processing capacity of the database server, the memory utilization of the application server, and CPU usage of a virtual machine, all these measurements become secondary if an application isn't responding to a user's request. These and a host of other metrics may hold the key to determining why an application isn't responsive, but without first having visibility into what's happening at the user level, prioritizing and mitigating issues is much more difficult.

Given that, why don't more organizations routinely and comprehensively monitor this aspect? Following are a few reasons:

- Cost. While legacy management solutions offer end user response monitoring capabilities, they are complex and time consuming to implement. For many organizations, adding another facet to the monitoring mix using these types of products represents too big of a cost and resource burden, in spite of the obvious benefits of doing so.
- Complexity. Many point solutions are available to do user response monitoring, but these products represent one more monitoring solution that needs to be purchased, deployed, and maintained. Ultimately, this approach results in higher complexity, which significantly diminishes the potential benefits of user response monitoring.

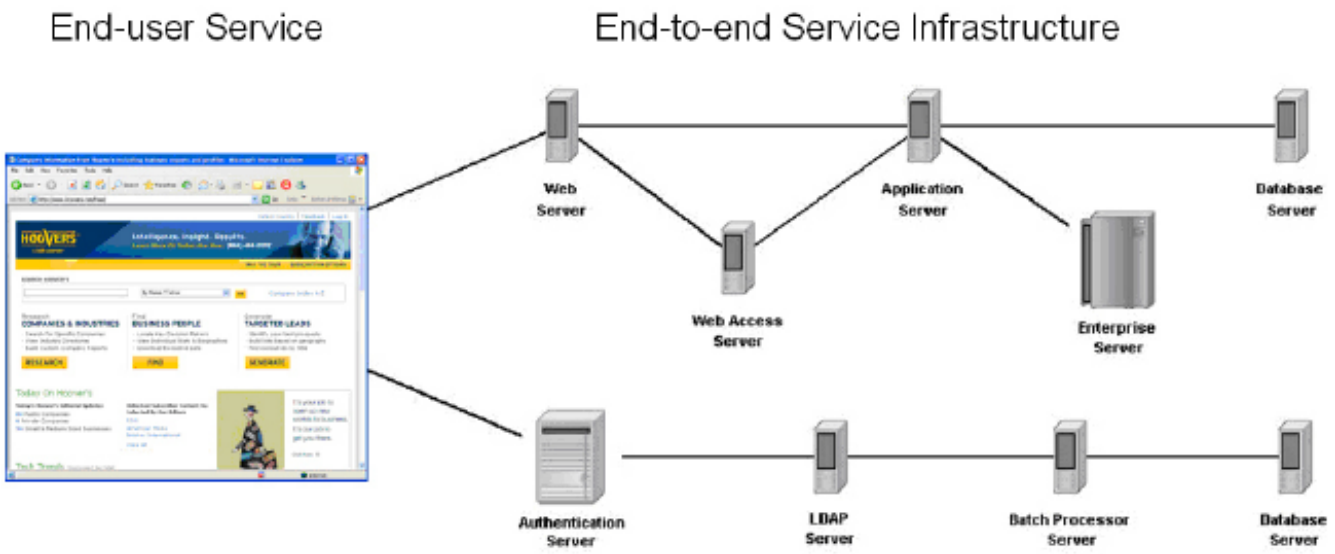
- Limited visibility. Having end user response metrics is only part of the picture. Using one solution to monitor end user response and other solutions to monitor various other systems within the infrastructure provides an inherent limitation: it only provides limited visibility. For example, if user response is slow, but an administrator can't easily view metrics for the rest of the infrastructure, troubleshooting remains difficult.

To address these challenges, your organization needs to be equipped with a single solution that delivers a broad set of monitoring capabilities. In the following section, you can find out what those capabilities are, and why they're important.

Capabilities Required for User Response Monitoring

To effectively gauge user response time and the complete user experience, you must be able to monitor several critical infrastructure elements—including applications, servers, and services—because all these elements play a role in the level of responsiveness and availability the user experiences. You need to be able to track application performance as the transaction moves from application to server, from server to service, and so on. This measurement needs to span your entire infrastructure, until the request is ultimately returned to the user. In this section, we detail the key capabilities required to effectively measure performance from the user's perspective.

Figure below: Effectively gauging user response times requires a clear understanding of the key applications and business services and their supporting end-to-end service infrastructure.



Application Response Time Monitoring

One of the keys to user response monitoring is gaining visibility into the application server to ensure it is processing transactions in a consistent and timely fashion. Monitoring application servers for user responsiveness may reveal whether the application server, with its current capacity, is fit to serve the workload of the user base, both in the near term and as that user base grows. Following are a couple specific examples of what's required for effective application response time monitoring:

- Application service port connection. It is important for monitoring solutions to poll an application service port to check its availability, i.e. whether it is active or inactive, and the time it takes to respond to a poll request. You should be able to test any

application service port. Look for pre-packaged monitoring capabilities for such common protocols as FTP, SMTP, POP, HTTP, and more.

- Multi-phase application response. In addition, monitoring solutions should be able to record, play back, and monitor the response times for each phase of a multi-phase, multi-transaction application. For example, an application may consist of several distinct phases, including application launch, application log in, application database query, and application log out. In this case, each application phase represents a unique customer experience, one that may rely upon a unique underlying technology infrastructure. For optimal application performance monitoring, each phase should be tracked.

Web-based Services Response Time Monitoring

Web-based services are increasingly representing, vital, business-critical services such as e-commerce, online banking, online support, and much more. If a Web server is serving a large user community or encountering heavy usage spikes, it may be taxed to the point of delaying service requests, which can degrade the user experience—and directly affect a firm's bottom line. That's why you need a monitoring solution that can track the performance of Web-based applications and business services. The solution should include capabilities for monitoring these aspects:

- Web server connection. Effective monitoring solutions need to test the amount of time it takes to connect to a designated Web server on a network.
- Web service port connection. You need to be able to test the availability and connectivity response time for critical Web services, such as HTTP, FTP, and others.
- Web page load. It is also important to monitor a collection of the URLs of specific Web pages, and measure the time it takes to load a page.
- Multi-phase Web application response. Just as with application response monitoring described above, it is important to monitor each distinct phase of multi-phase Web applications. This complete, partitioned approach allows administrators to measure the total application response time—and isolate particular phases of the Web application that are exhibiting sub par performance.

Email Response Time Monitoring

Today, your organization's employees, customers, and partners constantly rely upon email to get business done. In addition, email has also become an integral component of today's Web-based applications and services. For example, email is designed into distributed applications to automate and confirm online purchases, support desk trouble-ticket submissions, and much more. When a specific element within the email infrastructure experiences performance degradation, it may not only slow down or completely halt email delivery, it may result in failures of these vital business applications. Therefore, it is critical to monitor the timeliness and reliability of email communications.

For monitoring solutions to be effective, they need to monitor and verify email send and receive functions. In addition, they need to monitor outgoing mail delivery, incoming mail receipt, and total duration it takes from the time an email is sent to the time it is received. Further, look for coverage of the following common application types:

- SMTP/POP
- IMAP
- Microsoft Exchange
- Lotus Notes

Directory Services Response Monitoring

Numerous application components will depend on the DNS server for domain directions. That's why another vital aspect to monitoring user response is the ability to track the DNS server and other associated directory service applications to ensure lookups are processed in a consistent and timely fashion. An effective solution must monitor a range of popular directory services:

- WINS—tracking directory lookup response times for one or more WINS servers.
- DNS—gauging DNS lookup response times, and also monitoring for missing directory entries for one or more DNS servers.
- DHCP—monitoring the response time of DHCP IP assignments for one or more DHCP servers.
- LDAP—executing LDAP queries and measuring response time and number of records found.

Network Response Time Monitoring

The network is the critical link that holds the application infrastructure together. If routers and switches should fail, or exceed their performance capacity, the user experience will invariably suffer. That's why it's vital to monitor network response times and capacity, so you can ensure network devices are consistently and quickly transmitting data between end points.

To effectively do network response monitoring, you need a solution that...

- Can track connectivity and response times for a broad range of network devices—including routers, switches, servers, applications, and printers.
- Uses the ping command to verify network connectivity between designated hosts.
- Records response times between network devices, services, and applications—to help you quickly pinpoint areas of excessive network latency.
- Records connectivity test results and round trip sample data so you can leverage this information in availability and performance trend reports.
- Monitor VoIP environments and proactively detect and preempt excessive jitter, packet loss, and excessive round trip response times.

Database Response Time Monitoring

In many environments, application servers and Web servers will frequently make calls to a database server to fulfill a user request. If the database processes requests too slowly, the user will experience delays. That's why it's so vital for you to be able to monitor the database server to ensure database queries are being processed consistently and within required response intervals.

To be effective, your solutions should monitor database transaction rates, and database query response times for read and write operations. Finally, make sure the solution you choose covers all major database platforms, including Oracle, Sybase, IBM Informix, IBM DB2, and Microsoft SQL Server.

Required Deployment Characteristics

As vital as all the above capabilities are, a monitoring solution needs to possess a range of deployment characteristics if it is to be practically and effectively deployed in a real-world environment. Following are the critical characteristics to look for in evaluating potential monitoring solutions.

Robust Reporting

Given all the data collection that occurs when monitoring user response levels, it is important for you to get real-time, intuitive views into the most relevant metrics. Look for highly configurable dashboards that allow you to customize end-to-end response time views, both depending upon your user's areas of responsibility and personal preference. Dashboards should provide a wide combination of alert, performance, and SLA compliance status indicators in a single view. Look for Web-based access to these types of reports:

- Quality of service (QoS) reports. QoS reports provide historical reporting of any metric collected through the end user response monitoring, revealing historical response time performance or availability data over a number of periods. You should be able to report based on a range of intervals, including the last 24 hours, last week, last month, prior weeks, prior months, and prior quarters.
- Service level agreement (SLA) reports. SLA reports should offer a graphical and text-based representation of response time performance contrasted against service level commitments. SLA periods may be defined in day, week, or month increments. Look for products that offer the ability to generate alerts when existing performance trends suggest that response time commitments will not be met. An SLA-based response time monitoring solution that delivers on these capabilities can help make the difference between experiencing costly response time breaches and attaining 100% SLA compliance.

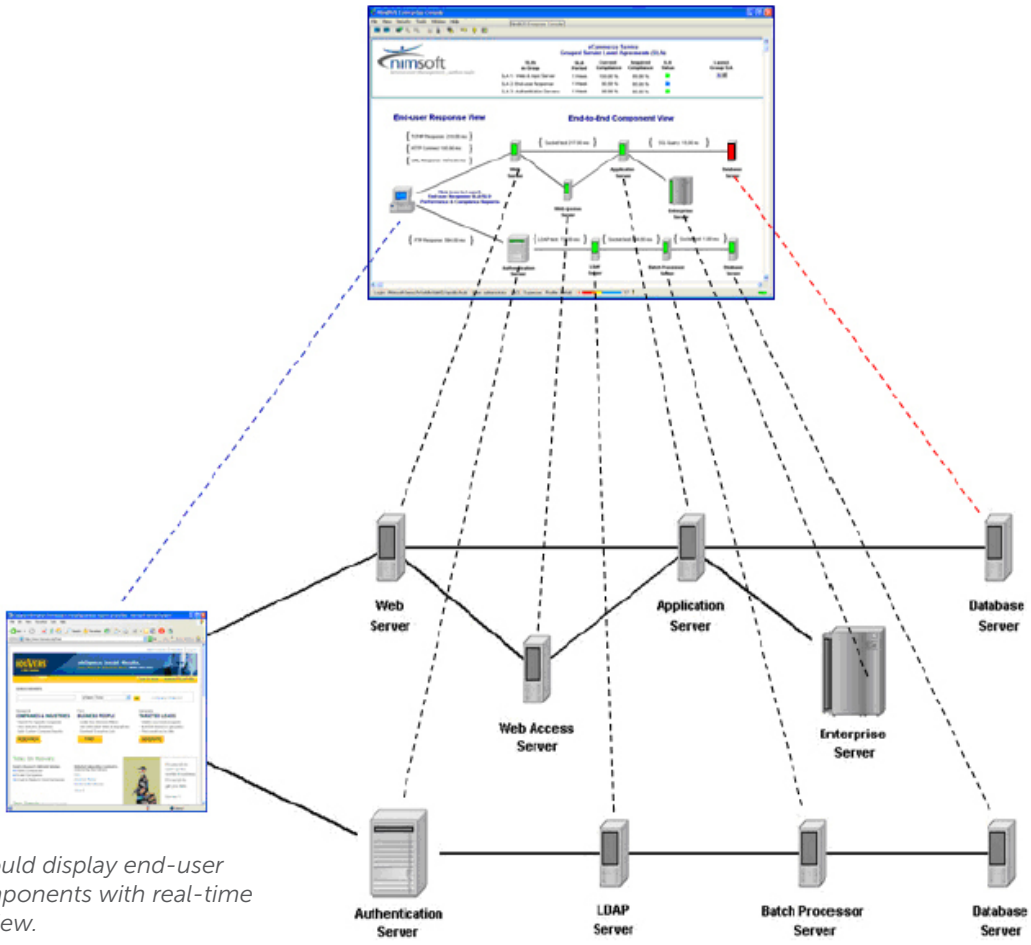


Figure to the right: Dashboards should display end-user and end-to-end infrastructure components with real-time alarm status indicators in a single view.

Distributed Architecture

In selecting a monitoring solution, stay away from market offerings that only support response time monitoring from a centralized management server. Typically, these offerings can only be deployed within the data center—inherently limiting their effectiveness in offering a realistic view of actual user response times. Rather, look for solutions that employ lightweight response time monitoring probes that can be deployed strategically throughout the business infrastructure. By deploying robots and probes on remote machines, it becomes possible to create "virtual users" that can be used to collect response time data by simulating application services and user transactions. These probes can be deployed wherever actual users and customers are when they interface with corporate services—whether they are in branch offices, business units, partner sites, or homes.

Additionally, when monitoring client/server applications, it is important that, whenever possible, solutions use native communication protocols to simulate end user activity and transactions. For example, to monitor Citrix application response times, look for solutions that use the Citrix ICA communication protocol to perform testing. This provides the most realistic picture of what a real Citrix client would experience. Other important application environments to consider when looking for native monitoring connectivity include SAP R/3, Lotus Notes, Microsoft Exchange, IBM WebSphere, and BEA WebLogic.

Flexible, Easy Deployment and Modification

Look for monitoring solutions that offer extremely flexible deployment options. This deployment flexibility will allow you to get response time visibility from practically any perspective you want. Any long term solution should also offer open APIs, a flexible architecture, and broad third-party integration, ensuring that you can easily adapt to other management tools and monitoring processes over time.

End User Response Monitoring: Use Cases

With the right solution in place, a broad range of organizations and infrastructures can be monitored to gain vital insights into the user experience:

- Branch offices. Businesses with branch offices may deploy response-time monitoring on designated 'virtual end-user' workstations at branch offices, and track application connectivity and round trip response time between those offices and the data center.
- Distributed environments. In a distributed application environment, response time probes can be deployed directly on Web servers or application servers to monitor the response times to and from servers hosting common application sub-services, for example database queries, directory lookups, and user login authorization.
- Managed service providers. For MSPs hosting customers' Web servers and services, response time monitoring can be deployed in strategic regional, national, or global locations to gain visibility of the experiences of end users who are using Web-based services from these remote locations.
- E-commerce. For businesses with e-commerce and other Web-based services, response time monitoring can be deployed in various geographic locations to monitor Web service response time from a consumer's point of view.
- Online business-to-business applications. For Web-based business-to-business services, response times can be measured among strategic business end points to ensure satisfactory application response times between business-to-business applications and transactions.

Search for monitoring products that offer an efficient code base and robust functionality that enable you to manage more systems, more users, and more data. Look for the following features, which will significantly improve your ability to broaden monitoring visibility, while minimizing monitoring expense:

- 24x7 monitoring automation. Find a solution that automatically plots performance statistics and measures them against user-defined thresholds. When thresholds are exceeded, you should receive alerts automatically, preferably through a variety of notification options—including pager, cell phone, and email.

- Fast deployment. Look for products that can realistically be deployed in days or weeks, compared with the lengthy, multi-month deployment times of the legacy monitoring solutions.
- Configuration flexibility. Search for monitoring solutions that have the sophistication and flexibility to be configured for different monitoring scenarios, including "one-to-many" monitoring configurations in which one remote probe is used to monitor many services, and "many-to-one" configurations in which many remote probes are used to test one service.
- Quick customization. Look for intuitive, yet powerful performance monitoring that can easily adapt to the unique, and fast changing, demands of your business.

Broad Infrastructure Support

The broader the infrastructure coverage a monitoring solution provides, the more benefits you stand to gain from the solution. While application environments may consist of common, commercially accessible software and hardware components, the collective whole represents application interdependencies that are unique to your organization. In some cases, these interdependencies can be straightforward, in others, highly complex. Look for solutions that can easily provide user insight into any potential scenario.

To be effective and practical, you need a monitoring solution that is equipped to address a heterogeneous environment. Further, look for customization capabilities that can be configured to record the timing of any potential transaction performed by the user. This should include monitoring, alerting, and reporting on each end-user transaction and the total end-to-end application response time.

With the right multi-platform monitoring tool in place, platform-specific variances can be masked from administrators, presenting a consistent monitoring interface for the entire infrastructure. Ultimately, this broad infrastructure support makes it possible to much more effectively and proactively manage the infrastructure and better ensure optimal service levels for users.

Virtualization Support

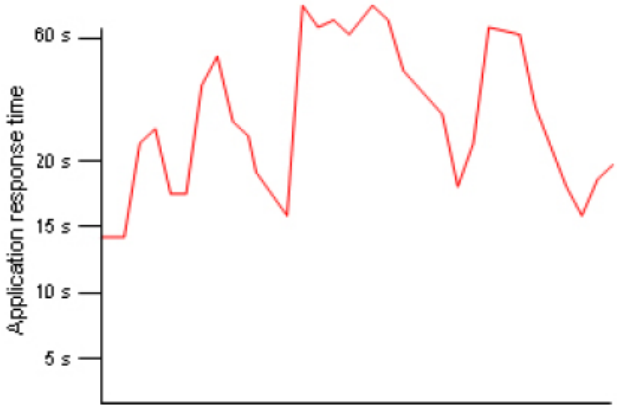
Organizations around the world, in every industry, and of every size have adopted virtualization technologies like VMware to gain greater resource flexibility and utilization. Yet in spite of all its benefits, virtualization presents significant challenges from a monitoring perspective. When virtualization gets implemented, an entirely new layer of "moving parts" gets added to the mix, and dramatically increases the complexity of monitoring servers and the infrastructure upon which they rely.

That's why, if you have a virtualized environment, it's vital that you have capabilities for monitoring user response times, in addition to having a solution that can monitor all virtualized and non virtualized systems in your infrastructure—including servers, hosts, applications, databases, networking services, and network devices. It is only with this broad visibility that you can effectively understand what's happening from a user's perspective, and effectively track both virtualized and non virtualized systems in troubleshooting and addressing any issues that may arise.

Three Steps for Leveraging Response Time Monitoring to Improve Service Levels

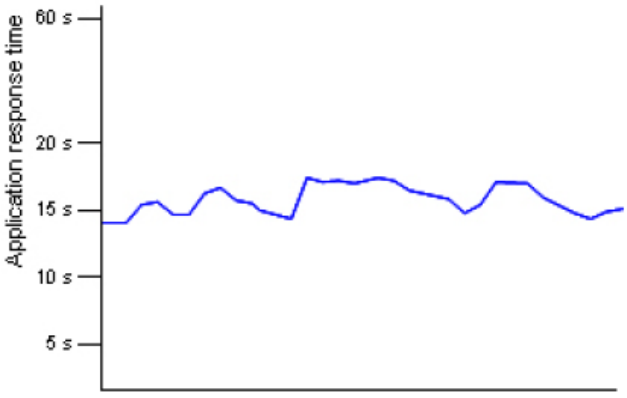
Visibility of application response times supports continuous improvement initiatives. The following graphs illustrate the progressive quality improvements that can be realized through increased service level control. Armed with this control, you can not only realize immediate benefits in reliability and performance, but leverage those insights into even more gains in the long term.

Step 1: Gain visibility into application response times and end-user service quality



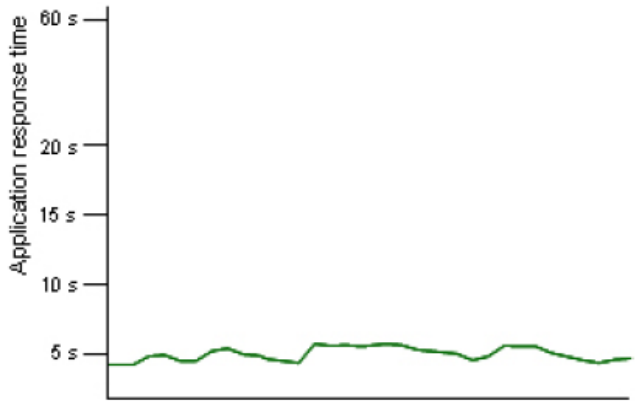
Real-time and historical trend visibility exposes dramatic, business-impacting performance variations.

Step 2: Apply visibility to improving user service quality.



Real-time and historical trend visibility enables early detection and correction of degrading conditions, which results in more consistent performance.

Step 3: Exploit visibility to deliver user service quality that meets and exceeds business demands.



Real-time and historical trend visibility provides a view with which to monitor how various adjustments affect service levels.

Conclusion

User response monitoring represents a vital missing piece in many organizations' monitoring perspective. Yet to fully leverage the value of user response monitoring, you need to be able to track not only end user response times, but to get a holistic view of your infrastructure. It is only with this holistic infrastructure view that you can both identify when users are experiencing delays or outages, and quickly pinpoint what the source of the issue is, wherever that issue may reside. Armed with this comprehensive view of the infrastructure and user response, you can very effectively start to improve service levels—and business productivity.

About Nimsoft

Nimsoft provides integrated, modern IT management solutions for more than 1000 enterprise and service provider customers globally, including 1&1, CDW, SoftLayer, Sur La Table, TriNet, and Virgin America. The company's Nimsoft Unified Manager is an industry-leading solution which helps organizations easily monitor and manage IT services in increasingly complex business environments. Nimsoft products integrate with existing solutions at any point from the data center to the cloud. For more information, visit www.nimsoft.com.

North America Headquarters
U.S. toll free:
1 877 SLA MGMT (752 6468) 1 408 796 3400

Email: info@nimsoft.com
Web: www.nimsoft.com

United Kingdom
+44 (0) 845 456 7091

Norway & Northern Europe
+47 22 62 71 60

Germany
+49 (89) 90405-170

Australia
+61 (0)2 9236 7216

Brazil
+5511 5503 6243

Mexico City
+52 (55) 5387 5406

Singapore
+65 64328600

New Delhi
+(91 11) 6656 6667

Mumbai
+(91 22) 66413800