IDC
Analyze the Future

WHITE PAPER

Proactive Performance Management: Finding that Needle in the Haystack

Sponsored by: ConicIT

Dan Yachin
October 2007

Global Headquarters: 5 Speen Street Framingham, MA 01701 USA P.508.872.8200 F.508.935.4015 www.idc.com

IDC OPINION

As businesses struggle to become more flexible in responding to changing customer needs, they are increasingly relying on IT infrastructure to streamline critical business services. This growing dependency is driving a fundamental change in the way the IT infrastructure is managed. In a dynamic IT environment that supports business-critical applications and services, for example, the inability to address critical application and system performance degradation or downtime in a timely manner may result in operational disruption, customer dissatisfaction, increasing call center expenses, and mounting IT expenses. Hence there is a need to detect performance problems as they occur, rapidly pinpoint the root cause, and fix them before they escalate.

Given the growing complexity of IT environments, it will be difficult to address this need using traditional application performance management approaches. For the most part, the challenge of pinpointing the root causes of problems in a timely manner does not stem from lack of information; with the prevalence of multiple management and monitoring tools for IT infrastructure components and applications, IT departments are flooded with performance data. But due to the lack of means to effectively correlate the different pieces of data gathered from disparate systems, the troubleshooting process is often too lengthy and rigid, particularly because in complex environments problems can rapidly escalate into a maze of multiple symptoms that may or may not appear to be correlated, and are often completely irrelevant to the root cause of the problem. New performance and availability solutions are needed that can complement traditional solutions with proactive capabilities.

METHODOLOGY

IDC developed this white paper using a combination of existing market forecasts and direct, in-depth, primary research. To gain insight into the challenges of applications performance management, and to learn how ConicIT's solutions can help organizations address them, IDC interviewed the company team on the issues of technology, product offerings, and go-to-market strategy.

IN THIS WHITE PAPER

This IDC white paper discusses the need for proactive application performance management solutions to address the growing importance of identifying and resolving performance problems in a timely manner, before they escalate and affect business performance. It discusses traditional approaches and their weaknesses in detecting and addressing performance issues in IT environments that are steadily becoming more complex, and the need for proactive performance management solutions.

SITUATION OVERVIEW

Background

Ensuring application performance and availability is a key issue for organizations. As critical business services and processes are today highly dependent on IT operations, the inability to provide consistent service levels for mission-critical applications can affect the business' overall performance.

According to IDC estimates, 75% of downtime occurrences are caused by poor technology in the network and application infrastructure. As they become increasingly complex, IT organizations that do not undertake incremental infrastructure projects that reduce downtime will see the latter affect their revenue as infrastructure moves from passive monitoring to active, real-time, analysis tied to business innovation, revenue streams, and IT service delivery.

Performance management is a decades-old system management function that began with the early monitoring of hardware device utilization and has evolved to include a variety of measures for operating systems, applications, databases, and other components of today's multi-tiered and distributed architectures, including network elements, Web clients, Web servers, application servers, storage, and so on.

The primary function of performance management software is to measure and report basic performance information and to help IT departments achieve desired operational performance goals. Performance management has always been challenged to keep pace with evolving technology and operating platforms, such as mainframes and open systems (e.g., Unix/Linux and Windows), Web, and wireless. It has also been challenged to keep up with major applications such as ERP, Web, and J2EE and databases such as Oracle, DB2, and SQL Server.

In terms of functional markets, IDC defines performance management as software used for capacity planning, performance data collection, performance tracking, and simulation software, as well as service-level management software when applied to systems and applications. Performance management also includes resource accounting software for resource utilization tracking and reporting.

Performance management software is used by IT departments to both help achieve performance objectives, such as desired end-user response times, and to optimize the utilization of individual hardware and software components to help control costs. Ideally, performance management solutions can help organizations avoid losing business opportunities due to poor performance or the unavailability of IT services, as well as to improve the effectiveness and reduce the cost of troubleshooting process.

Performance Management Fundamentals

The fundamental measures and concepts underlying performance management are quite simple and are repeated over and over again. They may be summarized as follows:

- **Transactions and Service Times:** Transactions can be viewed as units of work to be performed on a computer system, such as accessing a Web page, performing an online database inquiry, completing a batch job, or getting a stock price quotation. As transactions are processed, they use computing resources on the various infrastructure devices that must be utilized to functionally execute the

transaction. If the number of transactions to be processed or the rate at which transactions are processed rises, the supporting devices can eventually become saturated (i.e., become 100% utilized). Many performance measures, such as response time, are expressed on a "per transaction" basis.

- **Workload:** A workload refers to a group of similar or closely related transactions, such as online database inquiries or stock price quotations. Many performance requirements are expressed in terms of the workloads that need to be processed, such as stock price quotations, and their essential characteristics, such as the volume or number of transactions to be processed, the processing rate in transactions per hour, or the number of users.

- **Throughput:** Throughput refers to the rate at which transactions are processed, and it is usually expressed as the number of transactions completed per hour.

- **Response Time:** Response time represents the time it takes to complete the execution of an individual transaction.

- **Queue Length or Waiting Line Length:** This measure represents a count of the number of transactions waiting for service (or in service) on a particular infrastructure component, such as a server. Queue length is a measure of congestion, as it expresses the number of transactions or other objects waiting to complete on a particular device.

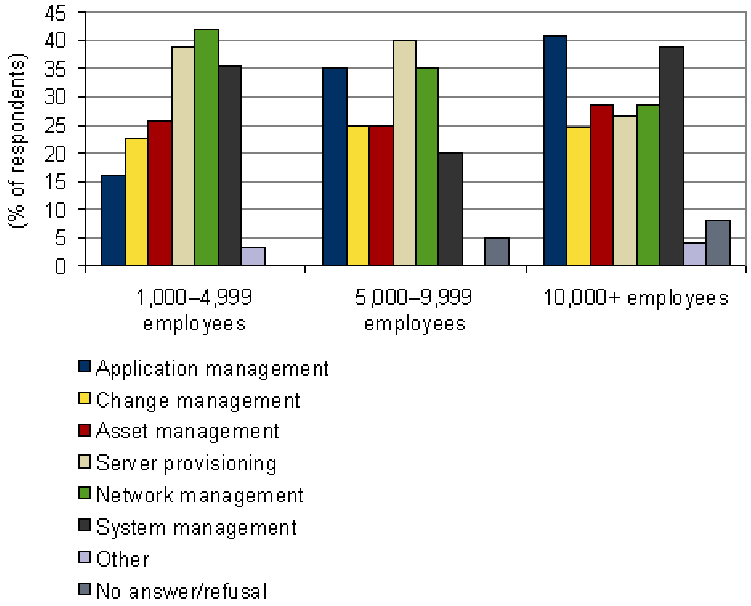
- **Utilization:** This measure determines how busy a device (such as a server) is during a specific time period, expressed as percentage busy. If a server is busy 80% of the time during a particular hour, it is expressed as "80% utilized." Utilization is measured by hardware or software monitors. Many hardware devices can be driven to 100% utilization under heavy loads. Highly utilized devices are often the source of bottlenecks, queue overflow, and contention, which may cause slowdowns in performance (in cases where the utilized device plays an important role in an overall process).

Systems and applications performance (whether current operational conditions, service goals, or future service level objectives) are typically described in terms of these fundamental measures. Indeed, one of the strengths of performance management as a system management discipline is its ability to quantify performance information and to use the performance measures for such activities as bottleneck identification, trending, and "what if" performance modeling.

The strategic importance of addressing application performance and management problems is reflected in the IT spending decisions of organizations. Correspondingly, IDC's 2005 IT Enterprise System Management Software Strategies study, which surveyed 100 senior IT staff about their plans for system management-related project deployment and specific use of applications, found that application management is a top priority for organizations of all sizes. As illustrated in Figure 1 below, the larger the organization, the higher the demand for application management.

FIGURE 1

Planned Infrastructure Project Deployment by Company Size



Source: IDC, 2007

Performance Management Perspectives

While the need to address application performance problems has never been more important, the challenges involved in doing so are becoming more daunting. The increasing difficulty of addressing this issue in an effective manner stems from the nature of traditional performance management tools and practices, whereby organizations address performance problems in a reactive manner.

Performance management is typically focused on one of several perspectives from which IT departments may view the processing of applications. An overall performance management strategy may require the incorporation of multiple perspectives to determine if performance goals are being met.

Major performance perspectives are summarized below:

- **Infrastructure Element Perspective:** The "Silo" approach. The most common approach to performance management is to focus on the individual infrastructure elements, according to the "silo" or technology stack to which they belong. Performance is thus monitored and managed in terms of categories such as network elements, application servers, databases, Web servers, storage, and so on. This typically aligns the performance management function with the domain experts in each major infrastructure stack and improves the management of each domain. But since a transaction may well require service from a variety of elements in order to be completed, the element-centric approach does not guarantee that transactions will meet their overall throughput or response time

objectives. This is particularly true for distributed, multi-tiered applications and for applications with Web components.

- **End-to-End or Transaction Perspective:** Here the focus is on the overall performance of a transaction as seen from the end-user or "end-to-end" viewpoint. This means that measures such as throughput and response time apply to the completion of a transaction, not just to the behavior of a particular infrastructure element. While by no means a new concept, end-to-end performance management has greatly increased in importance with the growth of Web-based and Web-enabled applications. In particular, monitoring solutions that send "synthetic" transactions over the Web to test the performance of Web sites have become more important.

- **Application-Specific Focus:** Another common emphasis is on the performance and "health" of a specific application, such as financial transactions, ERP, transportation, or healthcare. This involves a number of activities, such as defining the application (i.e., what infrastructure elements make up the application "tiers"), defining key transactions, and often obtaining application-specific measures such as response times for "imbedded" application functions.

- **Business Focus:** Another increasingly important focus is managing performance to deliver service objectives for key business applications, business units, or lines of business. Ultimately, this relates performance management to delivering business operations, and may often directly affect revenue generation, as in the case of online financial transactions. Indeed, some performance management software can relate revenue generated per hour to the underlying performance behavior of the enabling Web site.

- **Service-Level Objectives and Agreements:** One of the major uses of performance management software is in the area of defining and helping to enforce service-level agreements. Service-level objectives are generally expressed quantitatively, such as a desired response time for a class of transactions during specified loading conditions (usually peaks). Service-level agreements are written documents that specify what the service objectives are as well as the penalties (usually financial) incurred for failing to meet them (or, sometimes, what premiums may be paid for exceeding required standards). Performance management software is used to monitor and report the service levels actually achieved and often to report exceptions and "violations" if actual performance is out of bounds (or "out of compliance") with a required service objective.

The Need for Speed

While the abovementioned application performance methods allow organizations to monitor key performance indicators, identifying and resolving performance issues in real time or at least near real time remains a significant challenge.

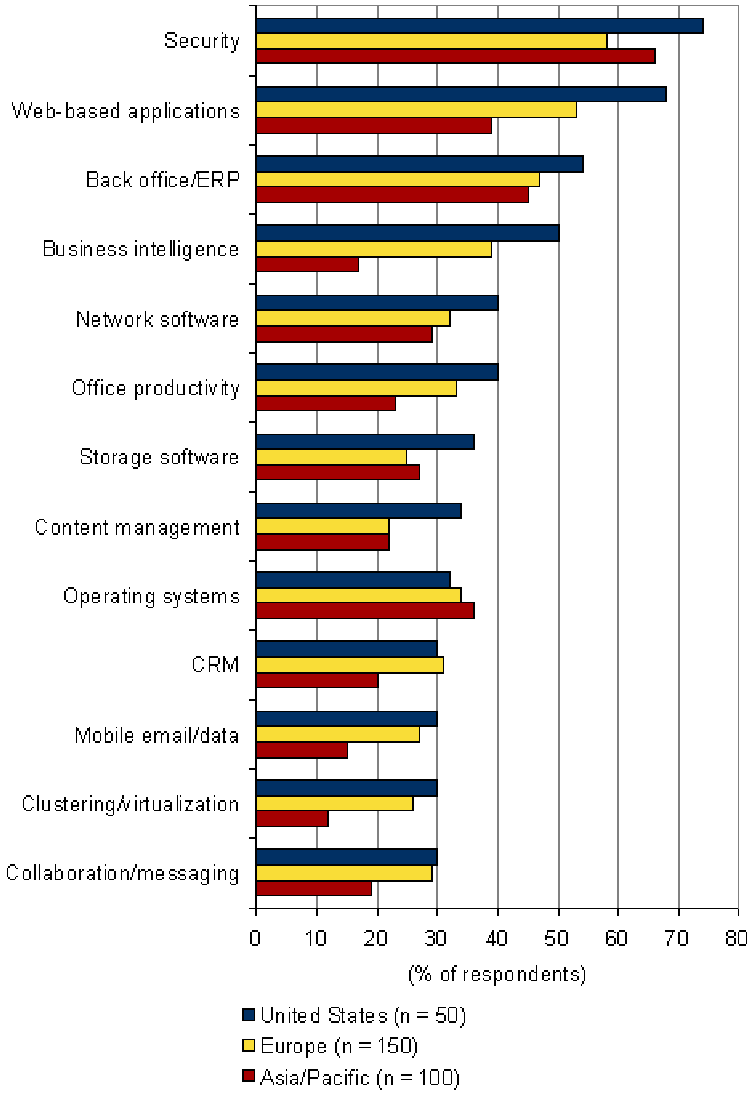
The need for timely resolution of performance problems has always been important. In recent years, however, with the move towards dynamic IT (which IDC defines as a model for creating a high-performance IT capability that can support the rapid pace of business change), the speed factor has become crucial. For the "dynamic enterprise", addressing application performance problems in a reactive manner might result in significant disruption to business operations due to the lengthy troubleshooting

process, in which the troubled application cannot serve the different parties that normally use it.

The importance of rapid detection and resolution of performance problems is highlighted by the fact that a growing number of organizations in different vertical industries are moving to provide business-critical services through customer-facing Web-based applications, which are connected to back-office systems. This trend is a key finding to emerge from the IDC 2006 Global Market Watch Survey, which analyzed IT budget trends and priorities survey. The study found that in 2006, 68% of large firms in the U.S. planned to invest in Web-based applications, and more than half in Europe also planned to do so (see Figure 2 below).

FIGURE 2

Planned Software Investments by Region, 2006



Source: IDC, 2007

When dealing with customer-facing Web-based applications, time plays a critical role as any minute of downtime immediately has an impact on revenue due to the inability, for example, of receiving customer orders. Other consequences of continuous downtime of these applications may include increasing customer dissatisfaction and potential churn, damage to the firm's reputation, and a decline in stock price.

Continuous application underperformance may also affect other "internal" aspects of the business. For example, downtime of mission-critical enterprise applications, resulting in the inability to perform routine tasks, could have an impact on the organization's productivity. In addition, for IT departments, tracking the root cause of a problem and restoring the application back to its normal status using traditional tools can be both time- and resource-intensive, thus affecting their productivity and ability to address other problems.

Too Much Information, Too Little Correlation

But why do organizations find themselves repeatedly chasing their own tails, trying to figure out where the problem is? The inability to track and address performance problems in a timely manner does not necessarily stem from the lack of adequate performance monitoring tools and statistics. Quite the opposite: IT departments often use a mixture of the abovementioned different performance management solutions, which provide them with multiple statistics.

But the lack of means to correlate the data collected from the different resources employed by a given transaction or a process (e.g., application servers, databases, mail servers, gateways, and others), and the inability to distinguish between irrelevant "noisy" data and actual precise and usable information, lead to difficulties in obtaining an integrated and consistent view of the application performance. As a result, IT departments often become aware of a problem only when disgruntled customers or employees are calling the help desk to complain. In this situation, the symptoms are being treated after the fact, when the damage has potentially already been done. Furthermore, without sufficient correlation, the process of troubleshooting performance problems might be cumbersome, involving multiple, lengthy tests (often based on a "runbook" manual) to determine the exact location of the problem.

The growing complexity of application infrastructure is making it difficult to obtain a correlated view of application performance, which could be utilized for timely detection and resolutions of problems. While the traditional "silo" approach for performance management served well in mainframe and client/server eras, as systems were relatively centralized and stable, the more the IT infrastructure becomes distributed, the more the effectiveness of this approach is reduced.

The typical application infrastructure today is a complex multi-tier environment composed of network equipment, Web servers, application servers, database servers, and back-end storage, spanning multiple technologies, architectures, platforms, and applications types (e.g., homegrown, legacy, third-party, and new applications). In this environment, applications are no longer isolated entities but rather distributed sets of components residing on different systems. Furthermore, as organizations are moving towards large-scale, enterprise-wide Service Oriented Architectures (SOA), the number of machine-to-machine, cross-application interactions is expected to increase dramatically in the coming years. In the SOA environment, applications are turning into highly distributed and fragmented entities – and ensuring the ongoing performance of these applications requires the addressing of severe technical difficulties.

Another cause for complexity is the increasing adoption of virtualization technologies. IDC believes that protection against unplanned application downtime will become increasingly important in the virtualized x86 server world as more business-critical and mission-critical workloads, hosted within VMs, come online. The requirements of these important workloads will largely be the same as they were when hosted on physical servers.

The highly distributed, dynamic nature of IT environments today is setting significant challenges for application performance management. In fact, the more complex the IT infrastructure becomes, the more frequently application failures occur. At the same time, finding and treating the causes of the failure is becoming increasingly difficult, because in complex environments a problem may escalate within seconds into a maze of multiple symptoms across multiple points, which may appear to have no correlation to one another. In this situation, drilling down to identify the root cause may not obtain meaningful data, as the problem could be transient or could change over time. IT departments are also familiar with a situation in which the monitoring tools show that the assets under their monitoring are working well, but end users are still complaining of performance problems. This occurs because in complex environments, problems tend to not to be isolated to a specific component but can rather span multiple silos.

The Need for Proactive Application Performance Management

The growing complexity of the application infrastructure is highlighting the drawbacks of traditional tools and approaches towards identifying and addressing performance issues. In addition, the option of addressing performance issues through manual processes is becoming practically impossible due to the ever-growing complexity and number of application and IT elements interdependencies, while internally developed tools for automating these processes are often too expensive and hard to maintain. Another common approach towards improving performance is simply throwing more hardware resources. In addition to the high costs involved in purchasing and managing additional hardware, there are many cases in which this approach is not effective: For example, adding more CPU power would not improve performance if the problem stems from I/O or memory access bottlenecks or database locking.

As organizations rely on high availability and performance of mission-critical applications to maintain their competitiveness, they are realizing the need to become more proactive in addressing this issue. To do so, several key challenges that need to be addressed.

Adopting a proactive performance management approach does not necessarily require the replacement of existing monitoring and management tools. On the contrary: as mentioned, the main problem that IT departments are faced with in regards to performance management is finding that needle in the haystack – i.e. the root cause of a problem that has escalated into a maze of multiple, apparently uncorrelated symptoms – from a multitude of data generated by different monitoring tools. Thus the key to proactive performance management is filtering out this "noise", which may mask the information that is relevant for solving a specific problem.

A proactive performance management solution should be able to gather and correlate performance data from various systems, and create a real end-to-end view of an application across the different underlying components (application server, Web server, database server, and so forth). In addition, automating the process of

identifying irregularities is essential, as even a highly skilled and experienced IT staff may sometimes overlook minor performance anomalies that can indicate an evolving problem. Hence there is also a need to maintain historical statistics of performance data from which a baseline of normal behavior can be established in order to automatically detect and provide alerts on performance deviations in real time.

CONICIT PROACTIVE PERFORMANCE MANAGEMENT SOLUTION

ConicIT was founded in 2005 by Yoram Kariv, a serial entrepreneur in the systems management field. Kariv is former co-founder of Precise Software, a developer of performance management solutions, which was acquired by Veritas (before its acquisition by Symantec) in 2002 for $537 million.

ConicIT develops a software platform that automates the process of identifying and predicting system and application performance problems. Based on a non-intrusive, agent-less architecture, the product analyzes and correlates data from various sources, including existing management and monitoring tools, to create an integrated view of the entire application or system. Using proprietary mathematical models the product identifies behavioral patterns of critical logical and physical computing resources and the relations between them over time to establish a baseline of normal behavior. Using a rules engine, the product can detect and provide alerts on performance deviations, discrepancies, and other indicators of abnormal behavior in real time.

All monitoring statistics are recorded in a central database, allowing for further data mining and statistical analysis that is aimed at pinpointing the root cause of different types of problems, including source code flaws, hardware faults the may lead to application failures, network bottlenecks, and others. With this information, IT departments can take accurate measures to fix problems and thus shorten the troubleshooting process. The product can then be used to detect the reoccurrence of similar problems in real time and prevent them from escalating. In addition, it continuously analyzes the behavioral characteristics of applications and systems in order to fine-tune its ability to predict "new" types of events.

The data provided by the ConicIT product is presented in common Web-based protocols and can be accessed and viewed from any machine on the network through a standard Web browser. Currently, the product provides support for mainframe and IBM WebSphere environments. The company is planning to expand its support for Oracle databases and BEA WebLogic environments.

CHALLENGES/OPPORTUNITIES

The advantages of proactively addressing application and system performance problems can be considerable. For the most part, it can help IT departments to significantly reduce the time to resolve problems, thus minimizing the impacts of application downtime or performance degradation. The rapid detection of a problem and its source also enables delegating the resolution to the appropriate IT group, thus eliminating the common blame game and reducing the resources allocated to the resolution process.

Proactive application performance management is becoming increasingly important for organizations that rely on their IT infrastructure to achieve business agility and

maintain their competitiveness. Demand for solutions such as ConicIT's is thus likely to heighten in the coming years, as organizations seek to move away from traditional, reactive performance management approaches.

The main challenge for ConicIT is standing out in a market that has been dominated by a small number of performance and availability management powerhouses. On the same note, as many organizations have already made significant investments in performance monitoring tools, the company may face lengthy sale cycles in which it will be required to demonstrate significant added value to existing solutions.

Given the market conditions, ConicIT should pursue a strategy of building partnerships with prominent IT management vendors as well as system integrators, outsourcing companies, and IT service providers. On the technology side, ConicIT is also facing the challenge of building support to various popular hardware and software platforms in order to be able to manage heterogeneous environments.

CONCLUSION

As organizations struggle to become more dynamic and real-time oriented, they are increasingly relying on IT infrastructures to streamline critical business services. In this environment, the inability to address critical application and system performance degradation or downtime in a timely manner may compromise the organization's competitiveness, and may also have a direct financial impact. With the growing complexity of the IT environment, however, addressing performance problems involves some significant challenges that for the most part cannot be fully addressed by traditional approaches. Organizations increasingly need to take a proactive approach to detecting and resolving problems before they escalate.