NetQoS
Performance Experts

WHITEPAPER

Performance First:

Performance-Based Network Management Keeps Organizations Functioning at Optimum Levels

IT organizations have spent billions of dollars implementing fault management tools and processes to maximize network availability. While availability management is critical, infrastructure reliability has improved to the point at which 99.9% availability is not uncommon. Given these improvements in device availability, companies are focusing more attention on performance management. By measuring how networked applications perform under normal circumstances, understanding how performance is impacted by infrastructure and application changes, and isolating the sources of above-normal latency, IT organizations can ensure problems are resolved quickly, mitigate risk, and take measured steps to optimize application performance. In this whitepaper, you will learn why this shift is taking place now and how to use a new Performance First model that will enable you to manage your network for application performance.

Most network engineers are very familiar with tools that report statistics on individual components such as links, routers, and servers; these infrastructure monitors have been around for a long time. Newer to the market are performance monitoring appliances that report end-to-end statistics and the end-user experience; these appliances provide a unique and comprehensive view of the enterprise without the need for desktop or server agents. They not only measure how well response time Service Level Agreements (SLAs) are being met, they also help solve a wide variety of problems with solutions that lead to significant reductions in operating costs.

With the increasing importance of networked applications, organizations have begun to shift their approach to network performance management. The first concern for network engineers and operations managers is no longer just device availability, but also how well applications perform over the network. Reviewing the evolution of network management will help explain why this is happening and what it means.

Managing Network Availability

For most of the past two decades network engineers have focused on managing the network for availability. Even today most of the products from network and systems management vendors focus primarily on telling network managers whether or not infrastructure components of the network are working. Is that router up? Is this link down? Does the server need to be rebooted? Historically, network engineers needed to ask these questions. Often routers did malfunction, links did fail, and hardware or operating system failures required servers to be restarted. And when the network failed, operations ground to a halt.

Today, however, most enterprise and service provider networks operate with 99.9 percent uptime or better. What drove the change? One reason is that technology has become more reliable in recent years. In addition, more and more mission-critical applications today are networked, and managers have realized that network downtime can be excruciatingly expensive to the business. For instance, in a brokerage firm, application downtime could cost as much as $6 million dollars per hour. This high penalty has made it worthwhile for network managers to invest in redundant servers, hot-swappable components, and other hardware to guarantee uptime.

Managing Network and Application Performance

While the reliability of networks has improved, performance issues have increased dramatically. The biggest influence driving this trend is the now common user expectation of a ubiquitous and instantaneous network. Other trends that have increased the volume and complexity of traffic and the need to monitor how well applications are delivered over the WAN include:

» **Data center consolidation.**

Enterprises and government agencies alike are migrating applications and data to central locations to save real estate, infrastructure, power and personnel costs, and improve manageability.

» **Increased number of remote users.**

Branch offices and telecommuters are proliferating as companies grow, merge, and expand globally. And the exponential growth of e-commerce transactions continues unabated.

» **The rise of voice and video traffic.**

Voice and video traffic are increasing rapidly and both rely heavily on the quality and consistency of network delivery.

» **Legacy applications.**

"Chatty" applications designed to run over local area networks often do not perform well when deployed over the WAN.

» **Software as a service.**

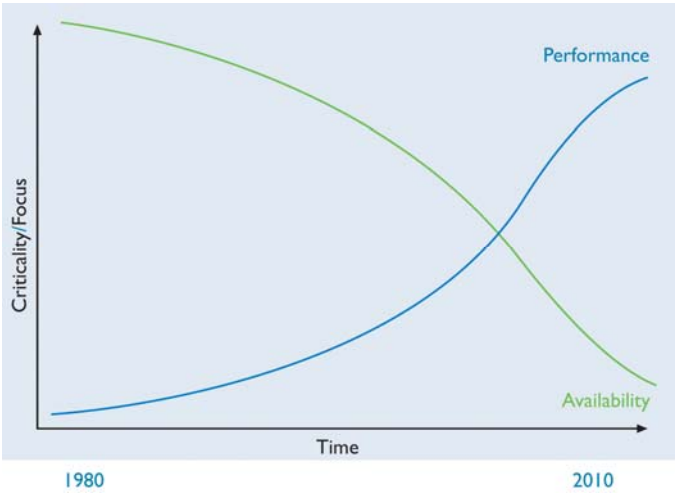
More organizations are choosing applications, such as Salesforce.com, that are delivered over the Internet as a service, in place of internally-hosted, client-server applications.

» **More complex applications including Service-Oriented Architectures (SOA).**

Distributed architectures and the use of Web Services as a means to develop reusable software for rapid application delivery and easier maintenance introduce network traffic between the various application tiers and infrastructure components. Enterprise applications may have ten or more networked tiers with each tier representing a possible failure point or bottleneck.

The convergence of increasing WAN use with improved device availability is leading network engineers to put performance first when it comes to managing their complex networks. (See Figure 1) They see greater returns by shifting their focus from fault management— which is largely under control—to performance-based management, focusing on how the network is affecting service delivery.

Figure 1



Over the past two decades, efforts by IT organizations and infrastructure vendors have reduced infrastructure downtime to a minimum. The most important trend in wide area networking today is the growing need to manage application performance.

The Case for a New Management Approach

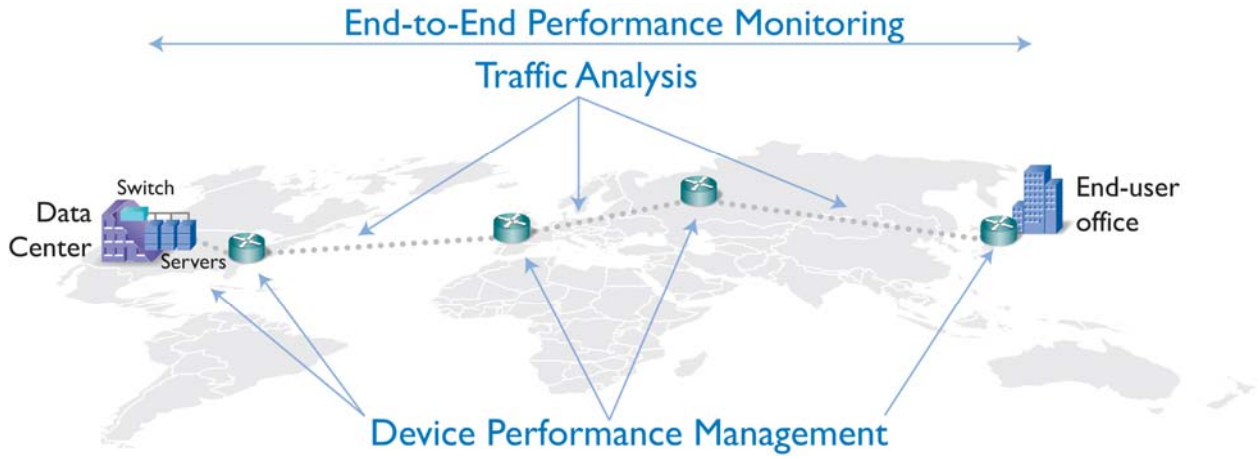
A recent Yankee Group report—aptly titled "Performance is the New Mandate for Network Management"—refers to a study on enterprise application management. The study found enterprises report an average productivity drop of 14 percent when experiencing performance degradations in Oracle, SAP, PeopleSoft, Siebel, and custom .NET and J2EE applications. The same study found that almost 70 percent of network managers said they are now responsible for isolating the cause of application performance issues, meaning that their purview now extends beyond the network infrastructure. Clearly "Performance First" is becoming the new standard for network management.

Other validations of the Performance First paradigm appear in the ever-growing number of white papers, analyst reports, and trade magazine articles that tout headlines such as: "The New Network Challenge: Application Performance Management," The Yankee Group, or "Adapting IT Performance Management: A Business Imperative," O'Donnell Research, and "Network Managers Getting Apps Savvy," Network World Magazine.

A New Focus for Network Management: Performance First

The Performance First paradigm inverts the traditional, bottom-up device monitoring approach and begins with top-down visibility into overall performance of applications running over the network. In this approach, infrastructure availability and utilization are no longer the sole gauges of network health. After all, it makes little sense to focus 100 percent of network management efforts on the 0.1 percent or less of the time there are hardware or software infrastructure failures. Fault management is necessary, but not sufficient.

Figure 2



A Performance First approach begins by measuring end-user response times, then analyzing traffic flows and device performance to optimize the network and troubleshoot problems.

The Performance First approach is driven by the fundamental purpose of the network infrastructure—to transport data from one end of the system to the other as rapidly as possible. The more efficiently data flows at the transport layer the better the application performance. This end-to-end response time measurement is the best measure to use when deciding how to optimize the network, plan new infrastructure rollouts and upgrades, and identify the severity and pervasiveness of problems. This approach recognizes that between the limits of the network and application infrastructure being "up" or "down" there is a wide spectrum of performance variation. It is not uncommon for availability status indicators in the Network Operations Center (NOC) to be all green while the help desk phones are ringing off the hook with users complaining about slow response times.

Focusing on the performance of key applications running over the network and identifying where there is opportunity for improvement enables IT organizations to focus on the most important issues: making informed infrastructure investments to support business demands, delivering consistent, acceptable end-user response times, and resolving problems that impact the business faster. IT organizations that successfully make the transition to a Performance First approach typically receive high marks from the business lines they serve.

The Performance First paradigm requires monitoring three key metrics: (1) end-to-end performance (2) traffic flows, and (3) device availability and utilization, as illustrated in Figure 2.

Three Critical Components of Network Performance Management

End-to-end Performance Monitoring: To track, measure, and analyze application performance for all user transactions from end to end for insight into the end user experience and the source of any latency issues.

Traffic Analysis: To visualize and analyze the composition of network traffic on specific links. This yields the information needed to redirect or reprioritize application traffic, or add capacity.

Device Performance Management: To poll network infrastructure components to isolate the source of problems such as a busy router or a server memory leak so that corrective action can be taken.

End-to-end performance monitoring gauges how well the network is delivering services to the end user and provides the best overall view of what is happening on the network. Whether troubleshooting a bottleneck, monitoring a new application rollout, or upgrading the network infrastructure, a Performance First management approach starts by understanding and base-lining end-to-end performance. What is normal performance for an application or a group of users? How does "normal" change during busy and off-peak business cycles? Which applications and users are experiencing poor performance? Where is the increased latency occurring? Is it in the network, server or application itself? What impact did the new application have on other application response times? Did the infrastructure upgrade deliver the performance boost expected?

With end-to-end performance metrics captured and the source of latency isolated, further analysis is much more focused. Traffic analysis enables network engineers to understand the composition of traffic on specific links where latency is higher than normal or expected. This yields the information needed to redirect or reprioritize application traffic, or add capacity.

If the source of latency is isolated to an infrastructure component—a busy router or a server memory leak, for instance—network managers need device performance management capabilities to poll the device in question and pinpoint the reason so that corrective action can be taken.

If latency cannot be attributed to the network or the server infrastructure and can be shown to be isolated in the application itself, the network team is armed with the proof that will eliminate the typical finger-pointing between IT infrastructure and application teams.

To understand clearly the sequence and unique role of each of these three metrics, consider the following analogy. Managing network traffic is similar to managing vehicular traffic in a metropolitan area. An end-to-end performance view of the network is similar to the view a helicopter has of traffic in a city. The helicopter pilot can observe generally how well traffic is flowing and where there is congestion.

However, he can't always tell if the problem is because of a slippery road surface, a stalled automobile, a bad traffic light, or an unusually high number of slow-moving trucks from a nearby construction site. Closer inspection of traffic elements and road conditions is required to determine the cause.

On a network, an end-to-end view of performance provides the first piece of information required for effective management: Is application performance normal? Is it better or worse? State-of-the-art monitoring technology can also determine where latency is abnormal, by timing packet flows across the network and server infrastructure, and through intelligent base-lining to determine a typical performance profile. With this information, it is possible to determine quickly which user locations and applications are affected, which WAN links are bottlenecked, or which server in a particular tier of the application infrastructure is performing poorly.

If the cause of congestion is determined to be a network issue, traffic analysis can be used to drill down to view details on the links in question. Monitoring network traffic flows is roughly analogous to observing vehicular traffic on a stretch of road passing through a specific intersection, providing specific information about the type and volume of traffic for problem resolution and planning. Analyzing network flows on an individual link determines how much traffic is HTTP, SAP, or streaming audio for example, and also provides a measure of volume by protocol, host, and conversation.

Without the end-to-end performance (helicopter) view to direct the focus of diagnostic efforts, traffic analysis on a link-by-link basis in a very large network with thousands of links is at best, inefficient. Furthermore, attempting traffic analysis by placing probes at each router is an expensive and ongoing management burden, similar to trying to locate a traffic slowdown somewhere in the city by sending 500 cars to look at each of the 5,000 roads to find the problem. This approach may eventually work, but it is a huge waste of resources and time.
NetFlow/IP Flow Information Export (IPFIX) technology, embedded in today's enterprise routers and switches from the major vendors, makes traffic flow analysis possible without expensive probe deployments. By utilizing the existing router and switch infrastructure, users can harvest and analyze NetFlow data to provide comprehensive flow volume and composition data for each link in the enterprise. This may be used for troubleshooting, capacity planning, and measuring the impact of changes.

Device-level performance, the third metric used in the Performance First management approach, is helpful for capacity planning and fault management (a broken traffic light). Many performance details of each device can be checked. SNMP device-level statistics indicate the percent busy for a server or router processor, memory utilization, packet errors, discard rates, and so on.

The Benefits of the Performance First Paradigm

The Performance First approach helps IT organizations optimize the delivery of critical application services to end users, mitigate the risks from change, and make the most efficient use of resources. The Performance First approach provides ways to:

- **Prove the performance of applications running over the network.**
  Too often IT managers are unable to provide objective measurements of performance against application SLAs. They have no way of knowing how well their service provider is meeting its performance targets. Frequent user complaints may be unfounded, but can't be disproved.
- **Deliver consistent application performance and measure it.**
  You can't manage what you don't measure. Without real-time visibility into end user response times, traffic flows, and infrastructure health, it's impossible to manage application performance proactively.
- **Mitigate the risks from planned changes and unexpected events.**
  How many times each week in an IT group is the question asked, "What changed?" More application outages and 'brown-outs' are caused by planned changes with unintended results than any other cause. It is critical that the impact of these changes is discovered and isolated immediately.
- **Make more informed infrastructure investments.**
  When infrastructure managers make uninformed upgrade decisions, the cost can be high. Often the anticipated results don't materialize, ROI is negative, and performance problems persist. Infrastructure utilization metrics alone are inadequate; knowing the exact source of latency and the composition of network traffic can often present alternatives to expensive upgrades, such as changing QoS rules or rescheduling data-intensive applications such as backups.
- **Work collaboratively and more effectively.**
  Performance management in large WANs requires laser-like precision because in today's distributed network applications there are thousands of hardware and software elements affecting end user response times. Network managers need tools that give them real-time global visibility and historical information to optimize the network infrastructure for application performance and work with peer groups to plan for changes.
- **Troubleshoot problems faster to reduce MTTR.**
  When performance problems do arise, what is the typical process to resolve it? Who gets involved? How long does it take on average to fix problems? How much of this time is spent in finger-pointing between IT functional groups? Knowing the source of availability and performance problems means the right technicians may be assigned immediately, and having detailed diagnostic information means they can fix problems quickly.

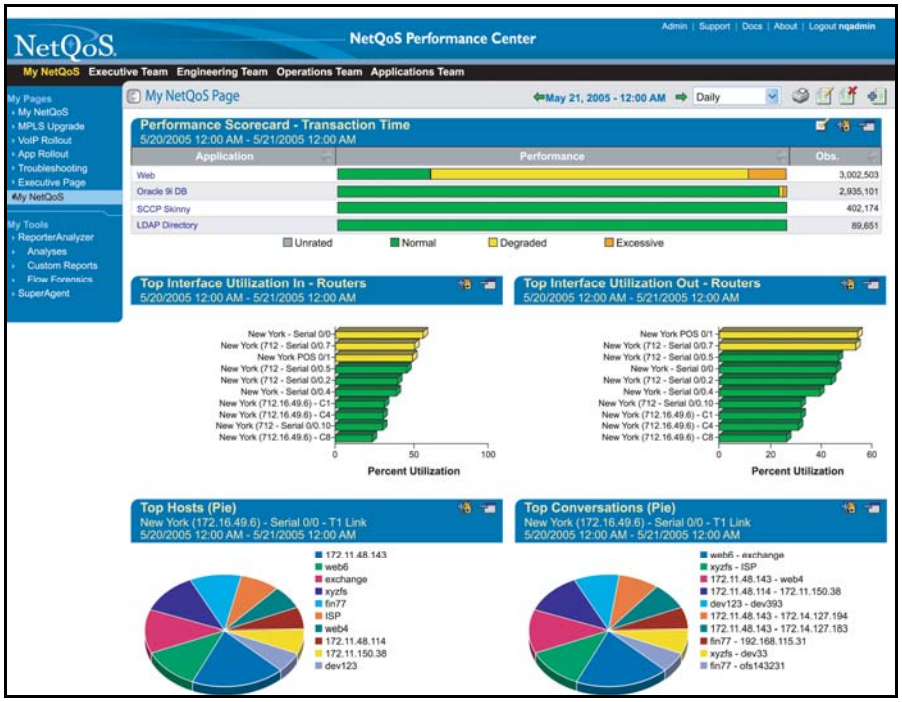
NetQoS Performance Center

Recognizing the absence of management tools to support the Performance First management paradigm, NetQoS set out at the turn of the millennium to fill the void. Since that time, it has introduced products to address each of the three key metrics described above, namely: end-to-end performance, traffic flow analysis, and device performance. Each is designed to scale to support the world's largest and most complex networks.

Together, they leverage industry standard instrumentation to provide network managers the global visibility needed to monitor performance SLAs, troubleshoot problems, and plan for growth, all without the use of desktop or server agents.

The NetQoS Performance Center is a management portal that integrates three NetQoS products, allowing network managers and IT executives to view the role-specific metrics they need with real-time data and historical trends in a single Web-based dashboard. See Figure 3.

Figure 3



The NetQoS Performance Center integrates performance metrics from SuperAgent, ReporterAnalyzer, and NetVoyant.

To understand the full value the NetQoS Performance Center offers, it is important to understand the contributions of each product module.

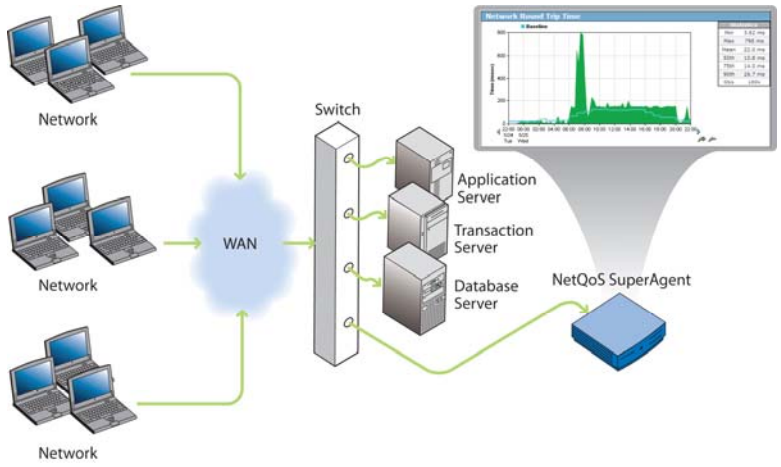
End-to-End Performance Module – SuperAgent

The most useful measure in establishing how end-to-end performance appears to the user is response time. NetQoS' end-to-end performance module—SuperAgent—monitors all the TCP application packets from the network into the data center and out again. This provides a way to measure network round trip time, server response time, data transfer time, and much more.

SuperAgent is the first product in its class to capture this information passively without the use of server or desktop agents. Before SuperAgent, enterprises had to deploy software on application servers and end-user machines to get a view of application performance. For enterprises with hundreds of applications and tens of thousands of users, it was a management nightmare.

With SuperAgent, a single device deployed in the data center can report application performance for all users at all locations. The SuperAgent appliance is attached to a switch mirror or span port, or a network tap near the server farm. This location provides a way to effectively inspect all TCP packet headers and calculate end-to-end response time metrics based on the source and destination IDs. Latency for each network hop, application server tier, and application component can also be determined. Most applications today have multiple tiers, and incoming client requests are routed to be processed, piece-by-piece, by a combination of different servers such as a transaction, application and database server. This means the only place to effectively inspect and troubleshoot multi-tiered application transactions is from the data center. See Figure 4.

Figure 4



The data center is the only measurement point to capture end-to-end response time and the latency for each hop in the application infrastructure. SuperAgent taps into the switch mirror port to detect all TCP application packets.

SuperAgent measures the latency in the application, server, and network components against normal performance baselines that are generated automatically by SuperAgent. Baselines are established for each hour of the day, each day of the week, and each week of the month. In this way, normal fluctuations between peak and off-peak transaction loads (nights, weekends, end of month, end of quarter, etc.) are accommodated. Alerts may be triggered when baseline performance is violated so that network managers know immediately, 24 hours a day, if end user response times are acceptable, and if not, where latency is abnormal (application, server, or network) so that further diagnosis may be done.

Similarly, this information may be used very effectively to prepare for new application roll-outs, additional users, and other changes. The Performance First management approach enables IT staff to be much more proactive in managing the network for application performance.

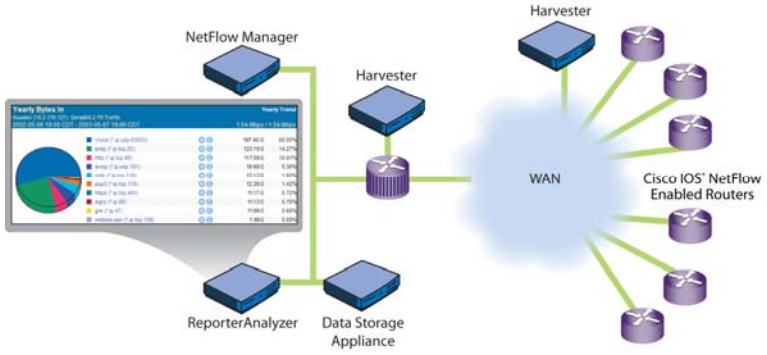
Network Traffic Flow Analysis - ReporterAnalyzer

Having visibility into the composition of traffic—including which users and applications are consuming bandwidth—on every link in the network gives network engineers the information they need to plan effectively for new applications and additional users, and make smart decisions on router configurations, bandwidth allocations, and upgrades. And if end user response time degradation is caused by abnormal latency in a network segment as determined by SuperAgent, analysis of that segments' traffic composition can quickly identify the source. NetQoS' traffic flow analysis module—ReporterAnalyzer—can identify a remote branch office backing up its server during the daytime, users downloading music files, a denial of service attack, or a fast-spreading virus.

ReporterAnalyzer is the first product to harvest and report on enterprise-wide Cisco IOS® NetFlow data (see Figure 5). Today ReporterAnalyzer also harvests flow data from enterprise routers and switches manufactured by Nortel, Juniper, and others that have adopted the emerging IPFIX standard. By leveraging the flow data instrumentation in these devices, NetQoS makes it possible to deliver global visibility into every flow across the network without deploying probes.

ReporterAnalyzer is unique in its ability to report on 100 percent of flow traffic for the entire network and access up to one year of enterprise-wide data for trending and detailed planning. Some NetQoS competitors offer Top-10 reporting solutions, reporting traffic on only the most-used ten applications on an interface while omitting hundreds or even thousands of other applications. ReporterAnalyzer monitors the distinct performance of various virtual circuits—discrete data paths of varying speeds on the same physical link. This makes it possible to troubleshoot Multi-Protocol Label Switching (MPLS) networks and Virtual Private Networks (VPN), which fence off certain portions of the network.

Figure 5



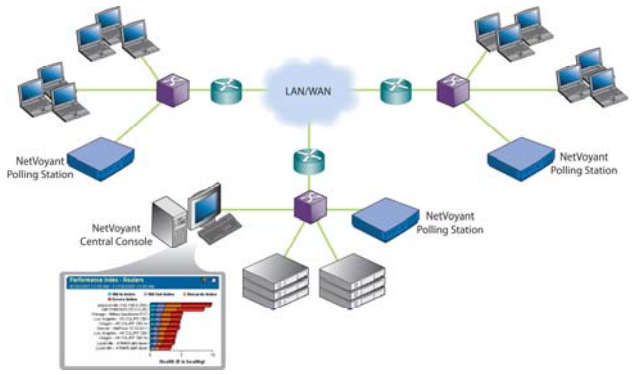
ReporterAnalyzer leverages native instrumentation, such as Cisco IOS® NetFlow, to monitor flows for every host, protocol, and conversation across the network.

Device Performance - NetVoyant

NetQoS' device performance module—NetVoyant—is the third product module of the NetQoS Performance Center. It provides SNMP-based performance metrics for network infrastructure, devices, and services. It polls WAN and LAN components throughout the infrastructure to collect statistics such as CPU and memory utilization, component or service availability, packet discards and errors, and information about FCENS/BCENS, used to address network congestion during peak usage periods. NetVoyant provides the ability to import and report on custom Management Information Bases (MIBs).

NetVoyant helps network engineers and operations managers solve problems quickly by singling out the causes of device performance issues. NetVoyant helps IT staff manage network and server infrastructure capacity by comparing current performance values with normal baselines, which it calculates automatically.

Figure 6



NetVoyant polls infrastructure components and reports on their health.

Summary

IT organizations can no longer manage networks in isolation from the applications they support. Traditionally IT staff built their network management practices around infrastructure availability and fault management. Today most networks are available more than 99 percent of the time and increasing user expectations for fast, trouble-free networked applications requires a shift from a device-centric to a performance-centric focus.

NetQoS is the only company that delivers a complete suite of products, included in the NetQoS Performance Center, for applying the Performance First management approach. NetQoS products are used to manage large enterprise networks, including a majority of the world's 50 largest companies. Representative customers include: Avnet; Barclays Global Investors; Bed, Bath, & Beyond; Chevron; Deutsche Telekom; NASA, Schlumberger; Turner Broadcasting Systems; and Verizon Communications.

About NetQoS

NetQoS is the fastest growing network performance management products and services provider. NetQoS has enabled hundreds of the world's largest organizations to take a Performance First approach to network management—the new vanguard in ensuring optimal application delivery across the WAN. By focusing on the performance of key applications running over the network and identifying where there is opportunity for improvement, IT organizations can make more informed infrastructure investments and resolve problems that impact the business. Today, NetQoS is the only vendor that can provide global visibility for the world's largest enterprises into all key metrics necessary to take a Performance First management approach. More information is available at **www.netqos.com**.

NetQoS, Inc.
e. info@netqos.com
p. 512.407.9443
t. 877.835.9575
f. 512.407.8629
www.netqos.com

WP rev3 20060804