Choosing a Monitoring System for Your IT Infrastructure?

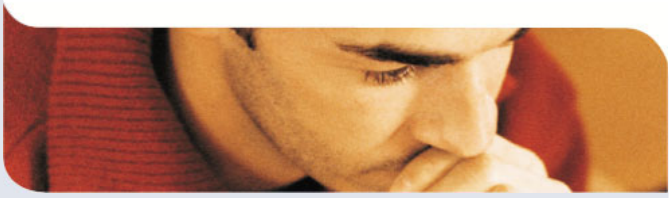
What Should Your Key Considerations Be?

Restricted Rights Legend

Copyright

Introduction

This document is intended for those considering to purchase monitoring software to maintain and manage their IT infrastructures. First of all, congratulations! You are about to make a very important decision – one that could potentially give you peace of mind, ensure you have free time to pursue more productive activities than fight fires late at night daily and a decision that can bring real value to your company, by ensuring the reliable operation of the infrastructure which manages your business critical services.



I have tested my applications thoroughly. Why do I still need monitoring?

Functional testing is a development time activity. Once your applications are deployed in production, various scenarios can occur. First, the real workload may not match the assumptions made during development. Second, administration errors could occur, resulting in application performance deterioration. Third, failure/malfunctioning of one application or network device may impact all the other components of the infrastructure. Hence, monitoring of production environments is a necessity, not an option.

As you know or will soon find out, there are literally hundreds of monitoring solutions out there in the market. Which ones are right for you? Many organizations take months assessing different monitoring software and yet find it difficult to differentiate between these products, since most use the same jargon. Proactive monitoring, root-cause analysis, service-oriented user views, customizable reports, auto-correction, event correlation, etc. are buzz words that most vendors use to describe their products.

This document describes different ways in which you can compare and contrast different monitoring solutions and how you can decide for yourself what are the key factors that you need to consider in deciding on a monitoring solution for your infrastructure.

Decide What You Need :

Stress Testing? Optimization? Monitoring? Diagnosis? Administration?

Before you even begin, you need to be clear on what your objectives are - Why are you looking for a monitoring solution? What do you wish to monitor? What do you expect to achieve? Many IT professionals confuse between stress testing tools, optimization tools, monitoring tools, expert diagnosis tools, and administration tools. Let us first clarify the differences.

- **Stress testing tools** are often used in IT infrastructures to get an idea of the performance that can be expected from the infrastructure under load. Stress testing tools are usually used in pre-production environments to find out what the expected capacity of the infrastructure services being deployed will be. Since they function by generating synthetic load (emulating hundreds of users), stress testing tools are rarely used in production environments as they can actually create an adverse impact on the target infrastructure.

- Like stress testing tools, **optimization tools** are also deployed in development/staging environments. These tools are used mainly by developers to understand how software modules of the applications can be optimized to deliver better performance under load. Examples of optimization tools are database tools that pin-point SQL queries that are not efficiently defined, Java tools that help identify code snippets that are not efficiently written, etc. Since they are designed to go to a great level of detail within a specific application, optimization tools are not ideal for production environments where the overheads of any external monitoring/tuning activity needs to be minimal.

- **Monitoring tools** are mainly deployed for continuous monitoring of production environments. Since they are deployed in production environments and are expected to be operating 24*7, these tools have to be designed to have minimal impact on the target infrastructure - hence, their CPU, memory, and disk footprint needs to be low. Network bandwidth usage should also be limited to a minimum. Monitoring solutions are primarily deployed so that administrators can be alerted as soon as problem conditions are detected, and they can take corrective action as soon as possible. Since they are continuously tracking the health of the infrastructure, monitoring solutions are also critical for proactive monitoring – i.e., to provide indicators of problem situations well before they become critical enough for users to notice.

  Typically, these solutions are expected to operate with minimal human intervention. Although they are mainly intended for use in production environments, monitoring tools also find use in pre-production environments. For example, they can be very effectively used in conjunction with stress testing solutions. While a stress-testing tool can provide an indicator of the capability of the target infrastructure, the monitoring solution can provide a guidance on where the bottlenecks are in the target infrastructure. Since the bottlenecks usually limit the capacity of the infrastructure, identifying and correcting these bottlenecks can significantly enhance the capabilities and value of the target environment.

- Unlike monitoring tools that are deployed 24*7 monitoring, expert **diagnosis tools** are deployed after problems are detected. These tools are often specialized solutions that require experts to look at the data being collected to interpret what the cause of a problem may be. Since the overhead of these tools may be high, and since they often require experts to handle them, diagnosis tools, like network sniffers, are usually turned on when required and not enabled for 24*7 operation.

- **Administration tools** are usually solutions specific to each network device/application in the target infrastructure, and they are often provided by the vendor of the device/application.

These tools are intended to allow the device/application to be configured for optimal operation. For example, with an application server administration tool, an administrator can set up the size of the database connection pool required to handle the expected load, to tune the number of server threads needed to handle the workload, tune the actions that will be taken when error conditions are detected by the application, etc.

The table below provides a quick comparison between the different types of tools.

	Use in preproduction	Use in production	Handle configuration & administration	Target users
Stress-testing tools	✓	×	×	Application developers developing applications; Experts deploying the solution in production
Optimization tools	✓	×	×	Application developers developing applications
Monitoring tools	✓	✓	×	Operators running the production environment; Experts deploying the solution in production
Diagnosis tools	✓	✓	×	Experts troubleshooting a problem
Administration tools	✓	✓	✓	Experts setting up the environment Experts maintaining the production environment

What Do You Want the Monitoring Solution To Do?

Of all the tools you are considering, first determine which ones are the monitoring solutions that can be deployed 24*7 on production environments without adversely impacting their performance. Once you have narrowed down your choice, consider why you are looking to add monitoring for your infrastructure.

Network Monitoring?

If you are interested in monitoring the network, there are many specialized network monitoring solutions. Simple solutions monitor network availability (mainly ping) and SNMP variables to report on traffic levels in different locations of the network. The more complete solutions offer in-depth topology views showing the network interconnections, making it easier for administrators to identify where the problem areas of the network are. Some specialized solutions also offer in-depth monitoring of traffic flows - e.g., who are the top talkers in the network?, which protocols are taking up most of the network bandwidth? etc.

System Monitoring?

System monitoring solutions mainly focus on monitoring the operating system of the servers in the network - i.e., their CPU, memory, disk activity and usage. These solutions also monitor the individual application processes running on the servers. The more complete solutions offer wider coverage of operating systems - across Microsoft Windows and Unix variants. Many vendor solutions offer different tools for Windows and Unix environments, since these toolsets have evolved by merging multiple independent solutions into one. Having to deal with a different user interface model for each operating system can be a burden for the administrator. Hence, look for solutions that offer a consistent look-and-feel across heterogeneous operating system implementations.

Application Monitoring?

Since no two application platforms are similar, application monitoring solutions must have specialized monitoring for each different application. Again, **look for solutions that offer a consistent user interface for monitoring different applications**, so that your operations staff does not have to deal with different, diverse interfaces.

IT Managers Want Automation

An Enterprise Management Associates Survey indicated that one in four respondents to the survey said that when they evaluated management products, automation capabilities played a critical role in their final decisions.

Network World, 05/2003

Many monitoring solutions require specialized knowledge modules or smart plugins for each application. These solutions offer very little deployment flexibility since the knowledge modules or smart plugins are specific to each application, in re-using these modules. For example, if you buy a module to monitor an Oracle database on Solaris, you are going to have to buy another module if you decide to move the database to Microsoft SQL on Windows 2003, and not receive any credit for the Oracle/Solaris plug-in that you had previously purchased. The industry is leaning towards monitoring technology that is not licensed specific to individual applications or operating systems. Hence, **look for monitoring solutions that have universal monitoring technology** - that can be deployed with a single agent instead of many application-specific modules.

Breadth of application coverage is also an important criterion. The monitoring solution must cover a majority of applications in your infrastructure. Since there are so many different application technologies available, you have to expect that there is probably no solution that covers all your applications. Given this scenario, **consider what the effort will be required to customize the monitoring solution to monitor your new applications**. Almost universally, monitoring solution vendors claim that their solutions are extensible. But then, what is the effort involved to add new monitoring capability in? Can you integrate pre-built scripts into the monitoring solution? Do you have to write Java or C++ code to add new monitoring capabilities? Do you need to build a new SNMP agent just to add monitoring for an application log file? Can you get new monitoring capability (say for custom applications) integrated into the monitoring solution in minutes or days or months?

Service Monitoring?

It is universally acknowledged today that monitoring systems have to be aligned to lines of business. A recent Network Computing Survey found over 92% of IT managers were focused on aligning their operations with the business goals. Aligning IT to business goals means that IT administrators need to be concerned about the availability, performance, and usage of services end-to-end, not just about individual network, servers, or applications. That is, service monitoring is what is required, not network, server, or application monitoring.

To offer maximum value, **a service monitoring solution must be able to monitor the performance of the service end-to-end**. That is, it should have the ability either by active simulation of user requests or by passively observing the real user activity to measure service availability and performance. The measurement must be done end-to-end - i.e., involving all the tiers of the infrastructure. Many solutions monitor the individual applications and network devices independently - i.e., check if the Oracle server is responding, whether the web server is up, etc. Often times, checking the health of individual applications or network devices is not sufficient. A true end-to-end perspective can be obtained only if the monitoring solution observes the service in the same manner that a user of the service experiences it.

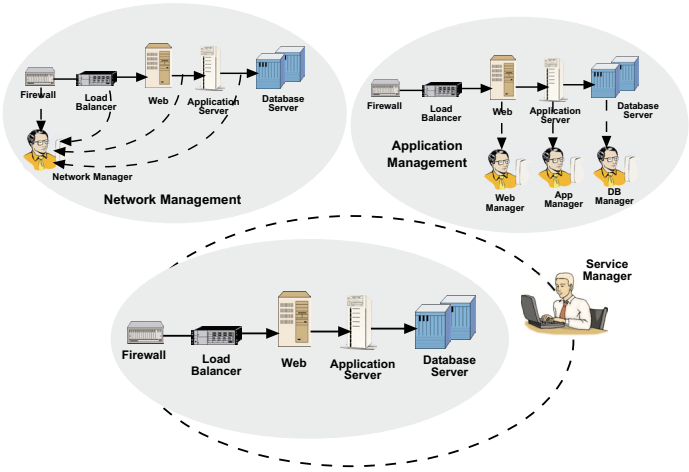


Figure - 1: Monitoring services not silos

The Challenge of Compliance

Monitoring tools also help with compliance to emerging auditing standards and practices. A Pricewaterhouse Cooper's survey of IT executives found that compliance to Sarbanes-Oxley puts the greatest strain on corporate resources. 84% of respondendts called Sarbanes-Oxley a challenge. Of those, 49% described it as a "major challenge." Only 15% of respondents said their reporting policies and procedures are fully automated.

Sarbanes-Oxley Compliance Journal, 10/2004

Secondly, **the essence of service monitoring is not just to get an integrated view of the components involved in supporting the service**. Many monitoring solutions offer the ability to group applications and network devices into services, so that administrators can look at the state of all the components of the service. **True service monitoring involves understanding the inter-dependencies between applications and network devices and using the knowledge of these inter-dependencies to determine where the bottlenecks involved in supporting the service may lie**.

How to be Proactive?

Clearly, the emphasis of any monitoring solution is on being proactive – i.e., alerting administrators before users notice problems. There are two key considerations that determine whether a monitoring solution is proactive or not:

- **The depth and breadth of metrics being collected by the monitoring solution**: Many administrators over-emphasize the importance of monitoring response times. Clearly, response time monitoring is important since it gives an indicator of the end-user experience. However, monitoring response times should not be the end-goal in itself. It is well known that the variation in response time with load is exponential – i.e., response time stays low as load increases initially, however, at a certain point, the response time starts to increase dramatically with increase in load(see Figure 2). Hence, monitoring the degree of load can be a better proactive indicator than monitoring the response time itself!

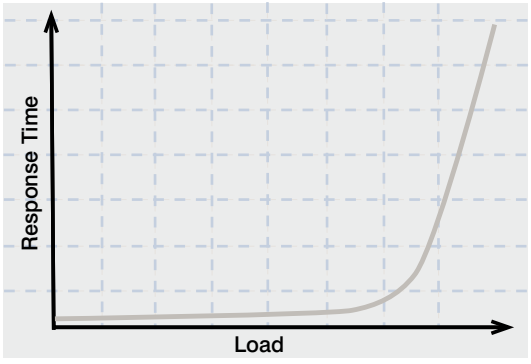


Figure - 2: Exponential growth of response time with load

There are many metrics that are better proactive indicators of impending problems than response time. For example, monitoring the system run queue can provide an indicator of contention for CPU resources, while an increase in the memory scan rate can be an indicator of low memory resources. Likewise, queue lengths and packet drop frequency monitored at a router can be a useful indicator of congestion at the router. Figure 3 shows an example where queue drops initially started occurring on 5/13, when they were left unattended. The problem got worse on 5/28 when the queue drops started impacting the end-user response time.

Hence, when choosing a monitoring solution, consider the breadth of metrics being collected by the system from the network, system, and applications. Solutions that monitor more proactive indicators of performance should be chosen ahead of solutions that just monitor response times.

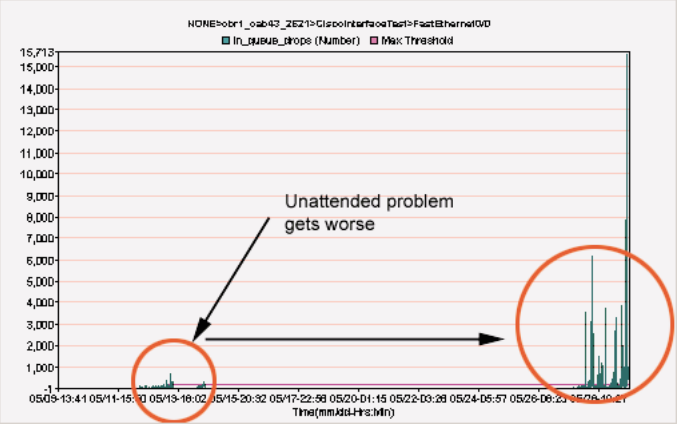


Figure - 3: A proactive network metric - queue drops on a router

- **The ability to automatically baseline performance**: Most monitoring solutions allow users to set fixed thresholds on metrics they monitor. Whenever a threshold is violated, an alert is generated. Fixed thresholds work well for metrics such as response time for which the acceptable threshold is set by business requirements (e.g., users will not accept a response time greater than 8 secs). There are other metrics for which thresholds cannot be fixed. For example, consider the number of users connecting to a web server. This metric is time-varying – more users during the middle of the day than during the early morning or late evening hours. Administrators usually do not have an idea on what thresholds to set for these metrics. In fact, they may not even care to monitor these metrics until they experience a problem. A monitoring system that can compute time varying baselines for these metrics would be ideal for such metrics. The baselines can be computed based on ongoing observations of the system, i.e., using past history, and alerts generated whenever the time varying threshold is violated. See Figure 4 for an example. The green line in the graph represents the number of users connected to a web server. Three days of data is plotted on the graph. The purple line represents the auto-computed baseline. Notice that the baseline matches the raw data in its variation with time of day.

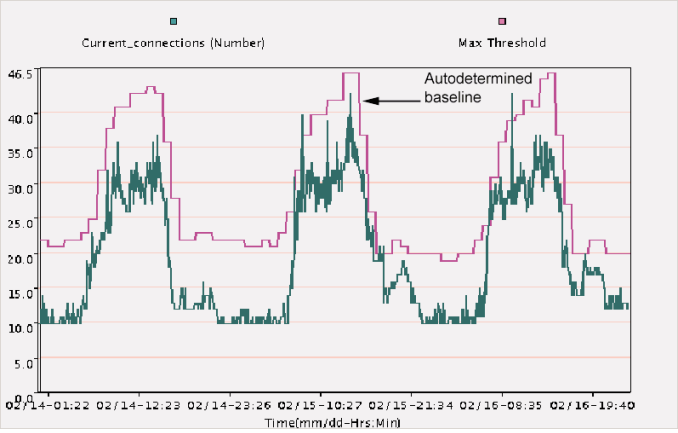


Figure - 4 : Auto-baselining performance metrics

IT Infrastructure Library – Best Practices - What are Administrators Focusing on?

ITIL is a set of best practices that organizations must follow to keep their critical infrastructures working at peak performance. A recent survey by Forrester Research shows that administrators rated problem management and service-level management as the top two management areas in importance. "What's most important is being able to describe a service in meaningful terms to the user, discover all elements needed to deliver the service, measure service quality and deal with exceptions and breakdowns," Thomas Mendel, principal analyst at Forrester, wrote in a research paper.

Network World, 07/2005

The key advantages of auto-baselining are:

- Makes it very simple to setup and configure a monitoring solution
- Saves on hours of consulting time
- Very little reconfiguration needed if the environment changes the monitoring solution is self-learning

When comparing monitoring solutions, look for solutions that embed intelligence to reduce the time and effort that need to be spent configuring and maintaining the monitoring solution.

Correlation and Root-Cause Diagnosis - The Holy Grail of Monitoring

"Root-cause diagnosis" and "Correlation" are probably the most abused words in the monitoring and management field. Clearly, effective correlation and Root-cause diagnosis capabilities are a key for effective monitoring and management. The more intelligence that is built into the software, the lesser human effort that is needed in diagnosing a problem, and the faster the diagnosis can be.

Almost every monitoring solution claims to do correlation and root-cause diagnosis. What is the difference between these tools then?

Correlation is not to be confused with state propagation. In some cases, a service is represented as a grouping of related applications and network devices, and the state of a service reflects the state of its sub-components (e.g, a service is faulty if any of its sub-components is faulty). In this case, the state of the service is merely a representation of the combined state of its sub-components. Likewise, suppression of events from multiple sources based on thresholds is also not correlation.

Importance of Monitoring & Management

A 2004 Yankee Group survey showed that 60% to 70% of current IT budgets goes toward maintaining systems, and less than 30% of the budget goes to new application development.

Yankee Group Report, 03/2004

True correlation and root-cause analysis go hand in hand - correlation involves analyzing the state of a service and its sub-components to identify where the cause of a problem lies. By differentiating between cause and effects of a problem, a monitoring solution can allow administrators to focus on the cause of problems rather than being distracted by their effects.

Some tools provide the means for human administrators to "visually correlate" between data from different points in the infrastructure. In this case, the tools themselves have no intelligence. Often a great deal of expertise is needed to analyze the data and isolate problem conditions. Moreover, the manual correlation effort is time consuming and laborious. Look for tools that can automate the correlation and diagnosis process, so that very little human intervention is necessary.

Many solutions embed rule-based correlation engines. To effectively use these tools, you will need to build elaborate correlation rules. Essentially, these correlation rules are if-then-else conditions that cover all the possible event conditions that can occur in your infrastructure. Crafting the correlation rules for your infrastructure can be an elaborate process, requiring expert knowledge and involving months of consulting hours. Furthermore, if your infrastructure were to change ever so slightly, you will need to re-architect the correlation rules.

Recently, there have been many solutions that do rules-free correlation. One popular approach is to use past history of problems

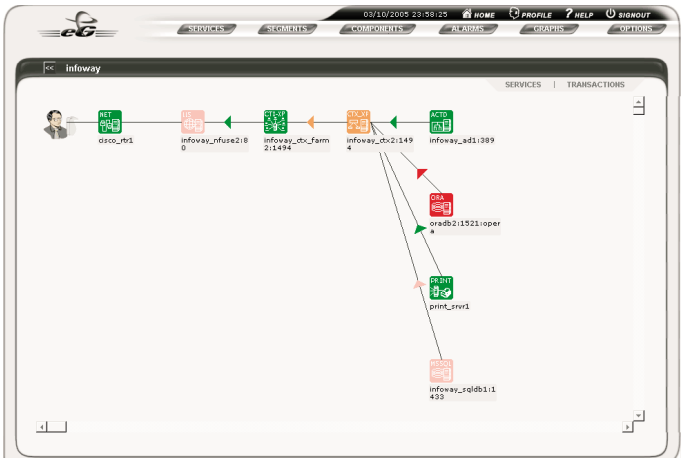


Figure - 5 : Service topology facilitates problem isolation

to perform correlation. For instance, when a problem occurs, the monitoring solution can keep track of all the symptoms it has seen and what the ultimate deduction of the root-cause was. If the same set of symptoms were to reoccur, the correlation engine would then be able to quickly provide an indicator of the possible root-cause (kind of what a doctor does when he sees multiple patients with the same symptoms). While this approach has merit for handling network events, the same problems rarely repeat when you get to the higher layers of the infrastructure.

Ideally, correlation must be performed within the layers involved in supporting an application (e.g., if the network is down and an application is not reachable, then the network error must be given higher priority over the application error). Since most services are multi-tiered, with inter-dependencies between the applications and network devices supporting the service, correlation must be performed across the different tiers of the infrastructure.

One caveat - in many cases, the root-cause diagnosis/correlation engine may be licensed separate from the base product, and in many cases, this capability alone can be more expensive than the rest of the monitoring solution!

The Need for Infrastructure Triage

With the dramatic growth in complexity of IT infrastructures, it has virtually become impossible to have a single monitoring solution that is appropriate for all possible conditions and all the infrastructure components. Hence, we believe that IT managers should explore a two pronged approach to root-cause analysis. The first step involves using a solution that is capable of identifying a problem with a business service and then performing triage of the problem down to a specific domain - i.e., network? database? application? Citrix? Additional, domain-specific tools can be used for the next level of diagnosis within a domain.

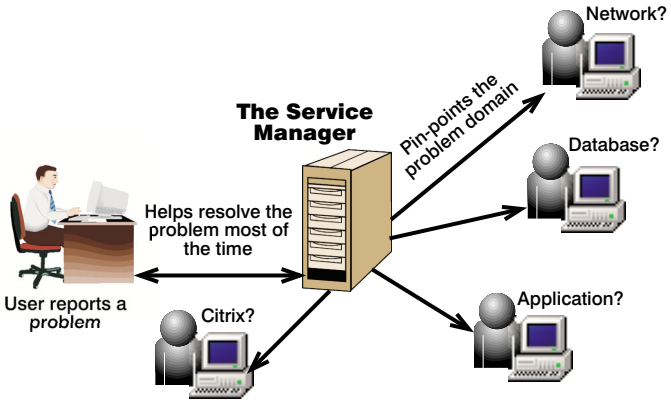


Figure - 6 : Infrastructure triage using a Service Manager

It is safe to say that no IT organization today is over-staffed, and most IT administrators are being called upon to do more within a shorter time period. Furthermore, as IT infrastructures get more complex, the few experts that are around are being pulled into more troubleshooting and fire-fighting activities and they have less time for new venture and more productive tasks. A service triage solution should be simple to use, so that it is not essential to have expert administrators use this solution. Lesser skilled service operators should be able to use the solution, so that they can be alerted to problem conditions sooner, and they can be able to determine which domain the problem relates to. In doing

so, this solution allows service operators to address and solve many infrastructure issues by themselves without requiring expert assistance. Furthermore, even if a service operator cannot solve a problem, he/she can at least get the relevant experts involved in the troubleshooting process. By ensuring that only the appropriate experts are involved in troubleshooting a problem, such a triage solution can allow IT operations to be streamlined for effective and efficient operation.

Cost Considerations

A critical factor in choosing a monitoring solution is the total cost of ownership. A key component of this is the initial investment - how much does it have to be? The cost of the monitoring solution is only one component of this. A rule of thumb is that implementation costs of many monitoring solutions can be six to eight times the cost of these solutions. When considering a monitoring solution, consider how long it will take to implement the monitoring solution, what hardware and software you need to install this solution, how well can the monitoring solution be used across locations, and what kind of personnel do you need to effectively utilize the solution. While these factors govern the total cost of a monitoring solution, consider how long it will take before you can effectively utilize this solution and how much you can save (by reducing man power, enhancing efficiency, and improving customer satisfaction through better service uptimes). These factors govern the return on investment from deploying a monitoring solution. Look for solutions that can provide return on investment in months. It will be even better if such a solution can also streamline your operations, so that you can be on a fast track for compliance to industry standards such as ITIL and Sarbanes Oxley.

Consideration must also be given to how flexible the solution is. No solution can meet all your requirements. How much time, effort, and cost will be involved in tuning the solution to meet your needs? For example, can you add monitoring for your custom application in minutes or do you need to hire a consultant who spends days to develop an SNMP agent to monitor the new application.

Getting Under the Hood of Monitoring Solutions - What you should not believe !

- Unix and Windows environments are different. I need different monitoring tools for these environments.
- Even if i have a single console, the look and feel of the Windows and Unix monitoring tools have to be different.
- When choosing a monitoring solution, I have to decide whether I need agent-based and agentless monitoring.
- Every application is different. So I need different modules for monitoring each application!
- Service monitoring involves placing icons representing all the applications and network devices of a service on the same page!
- If I monitor the availability of individual applications involved in supporting a service, i am monitoring the availability of the entire service!
- Response time monitoring is a proactive indicator of service performance.
- Correlation or Root-cause analysis is the process of displaying the status of multiple devices and applications on a single console and allowing the user to deduce where the root-cause of problems might lie.
- Correlation has to be done by defining rules.
- The monitoring solution can be easily extended (by writing SNMP agents !!!).
- The cost of a product is the main contributor to the total cost of ownership of the solution.

Conclusion

This document has highlighted many of the key considerations that you have to keep in mind when choosing a monitoring solution. The table below summarizes the key decision criteria.

Summary of Decision Criteria for Choosing a Monitoring Solution

Decision criteria	What to look for
Addresses your immediate needs?	Stress testing vs. Monitoring vs. Diagnosis vs. Optimization
Has breadth of coverage?	Support current and future applications/devices; Easily extensible
Is flexible and easy to deploy?	Reuse licenses? Single vs. multiple modules - one for each application
Is easy to use?	Consistent interface across platforms; Short learning curve
Is proactive?	Goes beyond response time monitoring; Auto-baselines your infrastructure
Offers automatic root-cause diagnosis capability?	Don't need experts to use the tool; Actionable alerts not raw data
Is cost effective?	Total cost of ownership - product cost + consulting cost + on-going maintenance

SINGAPORE

eG Innovations Pte Ltd
33A Tanjong Pagar Road
Singapore 088456
Ph : (65) 6423 0928
Fax : (65) 6423 1744

USA

eG Innovations, Inc.
33, Wood Ave, South
Suite 600, Iselin
New Jersey 08830
USA
Ph: (866) 526 6700

INDIA

eG Innovations Pvt Ltd
2, Murali Street, Mahalingapuram
Chennai 600 034
India
Ph : (91) 44 2817 2801 / 2817 2799
Fax : (91) 44 28179041

Email : sales@eginnovations.com Web : www.eginnovations.com