Application Infrastructure Management

New Ways
For
New Times

July 2004

Application Infrastructure Management

New Ways for New Times

Introduction

APM Advisors specializes in the Application Performance Management (APM) marketplace. We invest most of our time in emerging solutions that change the rules within the industry and provide the highest value proposition. With this insight we assist our clients in understanding not only what's important today, but also how to plan for tomorrow.

We have found that as problems arise, most people will choose the quickest and cleanest path to resolve it. You get an itch, you scratch it! You get a headache, you take an aspirin! Many tactical problems can be solved with a tactical solution, which makes it a little easier to associate the problem with the cure.

However, the message in this paper is that performance management is a fundamental aspect of the application delivery architecture. Just as you carefully select hardware and software platforms, architect and thoroughly test systems before deployment, the same energy needs to be applied in selecting your approach to Application Infrastructure Management (AIM). As a subset of APM, AIM solutions specifically address the requirement to manage the complexities of n-tier, Web-based application systems. As outlined in this paper, the right AIM solution will improve mean time to repair, lower capital expenditures on infrastructure and management tools, improve end-user service levels, and help you get the sleep you need.

Background

In the mainframe era, business critical application architectures were monolithic and so were the systems management tools—they ran on the mainframe as well. The migration to client/server computing environments created 2-tiered application architectures, typically consisting of fat clients communicating with database servers. This led to the introduction of new management tools that were highly distributed in nature, comprised of server processes, host agent processes, database processes, security processes, and so on. Since the late 1990s, organizations have been moving rapidly to Web-based computing and application architectures have shifted from 2-tier to n-tier, introducing a great deal of additional complexity. This rapid shift has left IT organizations ill equipped to manage their new application infrastructures effectively. The systems management industry has not re-tooled and so IT staff have to use existing products and processes that were designed in a different era, for a different set of infrastructure technologies, and for much simpler application architectures.

How AIM Is Done Today

Before Web applications are rolled into production, IT organizations perform capacity planning and analysis in pre-production environments under simulated load to estimate the IT resources required. These benchmarks provide a little insight, but due to the critical nature of these applications the production environment (servers, network,

and storage) is over-provisioned to assure acceptable performance under peak load conditions.

With the application infrastructure in place, the day-to-day management of the system is divided among several specialized groups. Within each group is a specialist who uses specific tools and processes to perform their tasks, and interacts with adjacent groups to coordinate changes, monitor system performance, diagnose and repair problems, etc.

An interesting aspect of this type of organization structure is that the granularity of specialization and the tools used within each group are closely aligned. For example, the group that builds and maintains server images uses its set of tools. Those images are used by the provisioning group, which loads the proper images on platforms and configures them within the infrastructure using another set of tools. Once in production, the groups that manage web, application and database servers monitor their resources with yet another set of tools.

What's The Problem?

Today's Web application infrastructures are very complex and interdependent. They typically include hundreds or even thousands of different resources and services (server platforms, network devices, storage subsystems, presentation services, application middleware, database services, etc.) that must work together in balance to deliver business application services to end-users. Most organizations lack the end-to-end instrumentation necessary to get a complete view of the health of these hundreds to thousands of resources and services. Certainly, very few have the visibility they need to identify the specific resources that contribute to application service level problems.

Despite the breadth and depth of specialized skills and tools in IT, the application infrastructure is a system of many complex resource dependencies that, unless viewed as a complete system, is difficult to understand. Yet typically, several different infrastructure monitoring products are used to check the status of servers, network, storage resources and software services, each with its own interface, and each storing monitoring data in its own proprietary database. Some organizations have invested in custom integration projects to provide a more complete system view, but this is hard to do and creates ongoing maintenance work. The result is a layer of management complexity on top of already-complex application infrastructures.

When service level problems arise, organizations perform an initial triage to determine what components are the likely culprits and then send in the specialists to diagnose those areas. Unless a problem is isolated within a specific resource or tier, the different staff members print reports, view various monitoring tools, dump log files, get in a room, compare notes and develop a hypothesis as to what the problem is, as well as a recommended solution (Figure 1). Even with all this effort and expertise, the staff team does not always know exactly how the recommended solution will impact application performance, or how it will impact other interdependent applications and resources, or if it will create some other problem, but they do have a sense that it will push things in the right direction.

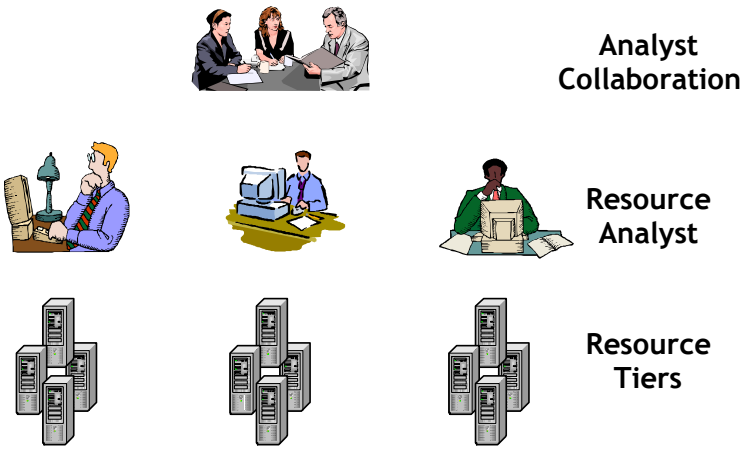

Figure 1. Problem Resolution Requires Many Tools, Processes and Specialists.

This well established 'people, tool, process' approach has a few challenges.

- The data that must be analyzed is stored in different proprietary databases in different formats.
- The resource data is not correlated relative to the other resources or to the applications that use the resources.
- Even if the data were all captured in a single database and correlated to the applications, the hundreds to thousands of variables make the analysis extremely difficult.
- Problem analysis and resolution is done after the problem has occurred, not in real-time as the problem develops.
- It is resource intensive, since it takes people to pull the information, analyze it, collaborate and implement changes.
- If problem resolution requires manual provisioning and reconfiguring of server, network, or storage resources, the time to implement these changes can be significant.
- Once changes are made, the same analysis process needs to occur again to determine the positive and negative effects of the change.

The Business Problem

The lack of a satisfactory AIM solution combined with the complexity of managing today's Web applications and infrastructures have resulted in three main problems:

- High cost of Web application infrastructures
- High degree of management cost and complexity
- Ineffective application service level management

High Cost of Web Application Infrastructures

While each environment varies, the industry average for server utilization rates is between 15% and 20%. Network and storage utilization rates are also very low. Over-provisioning to protect against possible resource failures and spikes in application service demand drives this fundamentally broken equation.

Total Cost of Ownership	$5,084,900
Under utilization cost @ 20%	$4,067,920
Under utilization cost @ 40%	$3,050,940
Under utilization cost @ 60%	$1,627,168

Table 1. TCO of Linux and Windows Servers

According to the Robert Francis Group in 2004, the Total Cost of Ownership of servers averages $62,203 for Linux and $141,193 for Microsoft on an annual basis.

Assuming 25 of each – the cost of under utilizing these resources is significant.

And if you think the current cost model is broken, just take a look at the effect of growth in Figure 2. If demand for Web application services increases by 100% the costs will go up by 500%, assuming 20% server utilization is maintained. This type of under utilization of assets is not only un-supportable; it is unheard of in all other aspects of business.

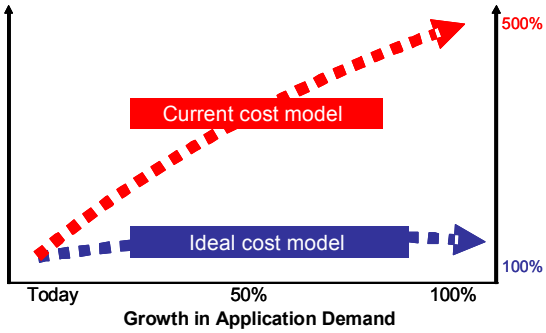

Figure 2. Infrastructure TCO Increases Disproportionately with Application Demand

High Degree of Management Cost & Complexity

As stated earlier, unlike any other time in this industry, the shift to Web applications occurred so quickly that the systems management tools borne out of the client/server evolution are not adequate for n-tiered Web application environments. What has resulted in many companies is a proliferation of very complex Web infrastructures running a greater percentage of the business critical applications, all managed by a surprisingly large number of tools that lack a system orientation, real-time analysis, or any form of collaboration. This means IT organizations must perform custom integration to maximize the benefit from the tools. And this integration work must be continually updated and maintained as the managed infrastructure changes, so a vicious cycle of recurring costs develops.

Ineffective Application Service Level Management

The lack of visibility and control of end-to-end resources results in Web application service levels that are not satisfactory, and are in fact, worse than most organizations realize. For example, a recent study of millions of end user transactions across many different organizations showed that "U.S. Web site pages and transactions … exhibited an average availability of 96%, well below the 99+% availability sought by IT managers." And "…37% of the time site pages and transactions did not exhibit performance levels considered acceptable." All this while business results are more dependent on 24x7 Web application availability and performance for customer and partner interactions, as well as for staff productivity. The business costs of application service level problems can be enormous, and are rising. See Tables 2 and 3.

Industry	*Operation*	*Avg. Cost Per Downtime Hour*
Financial	Brokerageoperations	$6,450,000
Financial	Credit card	$2,600,000
Media	Pay per view	$150,000
Retail	Home TV shopping	$113,000
Retail	Catalog sales	$90,000
Transportation	Airline reservations	$89,000
Media	Teleticket sales	$69,000
Transportation	Package shipping	$28,000
Financial	ATM fees	$14,000

Table 2. Cost of Downtime by Industry

Application	*Avg. Cost Per Downtime Minute*
ERP	$13,000
Supply Chain Mgmt	$11,000
E-Commerce	$10,000
Internet Banking	$7,000
Customer Service Center	$3,700
Electronic Funds Transfer	$3,500
Messaging	$1,000

Source: Dataquest Contingency Planning Research

Table 3. Cost of Downtime by Application

Tactical approaches to application performance management are costing IT organizations dearly. By continuing the vicious cycle of proliferating point management tools, custom integration scripts and processes, they are "putting lipstick on a pig" and introducing business risk. By taking the time to invest in a strategic, robust AIM architecture, they can dramatically reduce the cost and complexity of managing today's Web-based applications, deliver the service quality required by their end user constituents, and increase the return on their IT assets.

Application Infrastructure Management Requirements

Taking a step back and looking from the outside in, it is important to develop a unified, system perspective of the application infrastructure. While made up a many elements and their sub-components, it is in fact, a system within the system that delivers business services. Therefore, the core APM services of developing *information, resolving* problems and *controlling* resources at a system level are the goals. This implies an integrated management approach that is simplified, centralized, and highly automated; one that can be broken down into the following requirements:

- *Real-time monitoring of Web application service levels*
  The baseline *information* has to be an understanding of service levels, in the context of end-users and resources.

- *Mapping of resources to application service levels*
  It is not enough to just monitor how infrastructure resources are performing. An AIM solution should also identify which resources are used by specific applications and pinpoint resource conditions that impact service levels.

- *Business-driven policy specification*
  A complete AIM solution needs to align goals within the context of the business that depends upon the application services. As contention for resources is inevitable, the requirement is to accommodate business critical services over less important services by providing elasticity within the available resources.

- *Centralized, normalized database*
  Performance and resource *information* must then be normalized, centralized, and stored in a single database to enable sophisticated computing analysis to be run against it. This allows the specialists to focus on *resolving* a problem rather than correlating information.

- *Instrumentation of the end-to-end infrastructure*
  Rich instrumentation in the server, storage and network infrastructure is required, as well as in the software services across the different application tiers, to provide the *information* needed for intelligent problem and resource utilization analysis.

- *Industry-standard architecture*
  The AIM solution needs to easily integrate with other system management services throughout the entire IT system.

- *Real-time provisioning of infrastructure resources*
  To maximize resource utilization while being able to accommodate spikes in load, the AIM solution should integrate resource *control* capabilities, including server and network provisioning capabilities, to enable real-time, or near real-time, alignment of supply with demand.

- *Rapid 'time to value'*
  The time from initial installation to recognized value has to be (and can be) measured in days rather than months. Automated capabilities such as instrumentation deployment, development of resource-to-application associations, base-lining of service levels, and determination of resource normal operating ranges are key to obtaining immediate value.

- *Highly adaptive management architecture*
  Web applications are inherently dynamic environments where changes in application services and infrastructure are the norm. The AIM solution needs to adapt with the infrastructure it manages. It also needs to be insulated from configuration changes, infrastructure failures, and extreme load conditions so that it is available when needed most.

- *Federated intelligence to prevent routine infrastructure problems*
  Many of the routine problems that plague IT organizations and bring down applications, like running out of disk space on a volume, should be automatically identified and corrected. These can be dealt with via federated, rules-based intelligence.

- *Real-time performance resolution*
  With a clear picture of resource interdependencies across the application infrastructure, an AIM solution should analyze the complex cause and effect conditions in relation to service levels. As service level conditions fluctuate, the correlated resource information obtained across the system can quickly lead to problem *resolution*.

A Strategic Choice

Most organizations have a pretty good idea that they have many of the problems outlined earlier and desire the capabilities listed above that define Application Infrastructure Management. But due to operational structures and historic remedies many "itches are scratched."

The introduction of more tools just adds to the fundamental complexity challenge already in place. And it drives up capital and operational expenses due to the acquisition costs, the associated personnel costs, integration costs, and the time spent in collaborative analysis between peer specialists.

The Reactive, Tactical Path

While every environment is unique and there will always be a need for various application, database, network and process management tools, it is well worth understanding how the proliferation of tools continues to escalate operational costs. To illustrate this, see Table 4.

You would need to license many point management products to gain all the services described in the earlier *AIM Requirements* section. While the capital acquisition costs will vary significantly, they highlight the point. Taking a conservative 4x multiplier to cover training, installation, deployment, ongoing support and maintenance, the baseline TCO begins to expose some significant cost to the organization.

Management Tool	Capital Cost	Annual TCO
Web Server	$50,000	$200,000
Database Server	$50,000	$200,000
Application Server	$50,000	$200,000
Service Level	$25,000	$100,000
Server Configuration &	$50,000	$200,000
Application Route Cause	$50,000	$200,000
Totals	$275,000	$1,100,00

Table 4. TCO of Point Management Products

A critical cost, and typically the highest within any IT organization, is that of the technical specialists. While standard operations costs are factored into the TCO of the management products in Table 4, the time these specialists spend analyzing information at the resource or element level and then collaborating with other peer specialists must be accounted for. The more tools they have to use, the more time they must spend on independent analysis and collaboration to resolve problems. Therefore, the number of problems and time analysts need to spend in these 'fire-fights' the more it becomes clear that this uncoordinated management approach is unsupportable.

The Proactive, Strategic Path

As discussed in the introduction, one of the most compelling aspects of AIM solutions is the strong ROI they deliver. The kicker though is that some of the most innovative solutions in the marketplace have an architecture that enables the consolidation of functionality. As more functions are consolidated into a single management solution the 'cost to value' model improves, from both a capital and TCO perspective.

Like all markets, things evolve and an understanding of the unique problems that occur within an n-tiered application infrastructure is no different. Initially, point products or tools are introduced to 'scratch the itch' and they are rapidly introduced into the marketplace because the 'problem – solution' relationship is clearly understood, and there is someone whose very job depends upon that solution. As the marketplace matures, there is a better understanding of the system requirements and innovation begins to displace tactical tools with strategic solutions.

The development of the AIM marketplace is no different. The consolidation of functionality and the associated ROI multiplier associated with the new class of AIM solutions define the marketplace.

The early adopters of these solutions are the companies who stand the most to lose from down time or poor performance of web applications, and who currently invest heavily in management staff, tools and custom integration. The most compelling factor is the performance *resolution* value an AIM product can deliver. If a brokerage company can lose over $6,000,000 because they are down for one hour, it is pretty clear that a solution that can quickly isolate the problem condition and potentially fix it has high value.

The AIM value proposition is delivered by the shift from element to system management and the granular instrumentation of the entire application infrastructure. As depicted in Figure 3, an AIM solution provides the experts with highly integrated yet detailed information across the infrastructure, intelligent analysis of the cause and effect relationships between resources and application service levels, and resource controls to affect required changes in near real-time, enabling rapid resolution of problems and insight into the trends that require proactive attention.

Figure 3.
An AIM Solution Centralizes, Integrates and Automates
Information, Problem Resolution and Control

Conclusion

The shift from mainframe, through client-server, to n-tiered web based application models has been challenging for any IT organization. The ability to extend application services across an enterprise, to partners and even directly to customers is a powerful enabler for any business. Along with this capability is the responsibility to ensure that those applications are truly available, to each and every user their use is intended for.

This availability is precisely what Application Infrastructure Management solutions help IT organizations achieve. They provide insight across a system to monitor availability, assist in quickly identifying conditions that impede availability, and provide proactive services to ensure that critical business services obtain the necessary resources when demand exceeds supply.

Extending these services into the application infrastructure should be one of the priorities for IT, because if the infrastructure isn't stable the system isn't stable. Until the introduction of AIM solutions, the approach to supporting those systems was 'scratch an itch' or buy a tool.

With intelligent instrumentation and intelligent real-time analysis of information obtained from within the application infrastructure, IT can dramatically improve application availability, improve resource efficiency, and reduce operational costs. While scratching always provides temporary relief the real goal is to itch much less often. It's a sign of health.