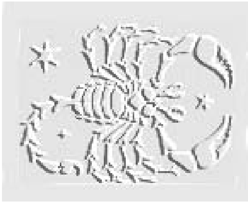
February 2009

Scorpion© – The Third Generation

An Updated Method of Analyzing and Optimizing Corporate IT Systems' Infrastructures



Rick Schoenmann, Certified Senior IT Infrastructure Consultant
John F. Ryan, Senior IT Consultant and Certified IT Specialist

Table of Contents

Preface

This whitepaper is the third generation of a review on IT systems rationalization[1]. The first version of this paper was originally released in June, 2000, ("Scorpion - Simplifying the Corporate IT Infrastructure" Publication GF22-5168) and subsequently revised in January, 2003 ("Scorpion Update - An Evolving Method of Analyzing and Optimizing the Corporate IT Server Infrastructure, Publication GM13-0189). This paper attempts to update the reader on many of the changes in IT rationalization strategies and experiences that have taken place since its last release. It also outlines some of the shifts that have taken place in corporate spending on IT rationalization and efficiency projects and the challenges that face executive decision makers going forward.

Introduction

Over the ten years that the IBM Systems and Technology Group's Lab Services Scorpion Team of consultants has been conducting IT Rationalization Studies using the Scorpion Methodology, few would argue that the nature, shape, complexity and cost of IT service delivery have evolved substantially. Some areas, such as the unit cost of commodity hardware, have seen drastic reductions, which in turn have abetted the practice of large-scale server farms and all the support complexities that come with it. Other areas, such as rising data center power and cooling requirements, have taken on increasingly significant implications, both from a cost as well as a global public relations perspective. While still other areas, such as a corporate wide focus on IT consolidation brought on in part because of the proliferation of server farms mentioned above, have led to the re-energizing and retooling of legacy based systems like the IBM Mainframe into consolidation platforms for workloads traditionally hosted on distributed Intel®[1] and UNIX®[2] platforms. Further, significant advances in hardware and software technologies have led to positive changes in, for example, the number of servers that can be effectively administered by each IT staff member, the rates of average utilization at which Intel and UNIX platforms can be run or the server consolidation and virtualization ratios that can be expected from current state of the art servers.

But whatever the changes have been, small or large, technical or financial, one underlying premise has begun to drive executive decision making. That is the desire to transform IT from a reactive, component level hardware delivery model, to a more agile service oriented architecture that works in partnership with business. In the desire to help their clients move through this level of transformation, IT vendors and service providers have had to alter their approaches to addressing and analyzing customer needs. New products and services have taken root while older products have had to change and adapt, or be left behind. As these products and services have evolved or developed so to have their affects on existing cost models. This has required an evolution and rethinking of those elements that affect the overall cost and complexity of maintaining a corporate wide IT service delivery model and its supporting infrastructure. This paper will attempt to update the reader on how IBM's Scorpion Methodology has evolved to address this changing IT landscape.

Studies have shown that the combined spending on hardware, software, staff, facilities, and services for servers account for as much as 40% of total corporate IT expenditures; and it is growing[3.] Further, we have seen that the ratios of the various types of expense have also

---

[1] We use the term "Intel" throughout this document to refer to any Intel and AMD x86 based systems running Microsoft® Windows® and Linux®

[2] We use the term "UNIX" throughout this document to refer to any Mid-Range systems, and in particular RISC based systems running a variant of UNIX, eg. AIX®, Solaris, HP-UX, OpenVMS, Tru64, etc.

[3] ITG Group, 2001: Value Proposition for e-Infrastructures - Cost/Benefit Case for IBM ~

changed over the last five years, with the majority of costs now shifting toward software, labor and maintenance, and away from hardware acquisition costs. In addition, the escalating costs of electricity generation and delivery, coupled with the increased power demands from servers and storage, make the percentage of costs for "facilities" (power, cooling, HVAC) the fastest growing portion of IT expenditures. In fact, according to a report from ASHRAE, today's energy costs alone for a typical 1U server equal or exceed the initial cost of the server itself (http://www.electronics-cooling.com/articles/2007/feb/a3/).

Cost of ownership analysis is still a hot industry topic among most IT executives and continues to be promoted by all the server hardware vendors. Likewise "Server Consolidation" is still promised as the universal panacea in helping IT organizations cut costs. As such, both these concepts continue to be widely used methods for managing IT cost and complexity, but they are only partial approaches to understanding and optimizing the IT infrastructure.

In addition, now, more that ever, many of the critical success factors for improving shareholder value are very dependent on the company's IT organization and its IT infrastructure. It has become more and more central to the job of the IT organization to contribute to profit growth, not only by cost containment but also by enabling new revenue opportunities through new channels and new services. All of this activity is typically aimed at delivering new competitive advantage to the organization quickly and effectively.

These changing realities put huge pressures on the standard ways of understanding and solving business problems. In a time of economic uncertainty, increasing costs, skilled labor shortages and heightened competition we need new ways of thinking about financial and technical analysis and the management of the IT infrastructure as well as new server technology solutions. Over the past ten years IBM has developed and, more importantly, continues to refine a unique method that addresses these requirements. This IT infrastructure analysis method balances the best techniques of server and storage infrastructure technical analysis with financial analysis, capital budgeting principles and tools, and new technology solutions. This approach is called the Scorpion method.

Highlights

IBM's IT Rationalization Study, using the Scorpion Methodology, is typically delivered as a comprehensive six to eight week project conducted by experienced IT infrastructure consultants from the IBM Systems and Technology Group's Laboratory Services organization. It investigates the technical and financial aspects of an enterprise's total IT server infrastructure including centralized and remote Intel, UNIX and mainframe server hardware, software, environmental elements like power and space along with the enterprise's service delivery support staff.

A number of techniques are employed to analyze the current server infrastructure, both technical and financial, using analytics and metrics derived from customer provided data as well as data from other reliable sources. These derived rules of thumb and industry assumptions can, when needed, even be used in lieu of available customer data. These industry averages are one area where the IBM Scorpion team works continuously to keep their metrics, ratios and rules of thumb up to date by constantly reviewing and revising these data points against our practical experiences and ongoing industry research.

Data and information leading to insights on topics such as people efficiency, server efficiency and quality of service are also collected and analyzed. Comparisons or informal benchmarks are offered based on past studies and, again, on readily available and reliable third party industry data, which are then included in the form of balanced scorecards.

Incremental and future cost projections provide an insight into the future state costs, and emphasize the importance of the utilization and server performance metrics. The biggest breakthrough, however, is in getting to credible solution design alternative recommendations for tactical and strategic projects that are then used as the baseline for business cases that compare multiple target server scenarios.

The method can, as required, use two different approaches: A full study is appropriate for analyzing large scale server infrastructures, often supporting multi-tier applications and highly cloned infrastructure service areas such as Web, file/print, e-mail and branch applications. This approach assumes the ready availability of detailed server inventory data or a willingness to compile and provide these necessary data. This approach is especially valuable where it is appropriate to more accurately compare and analyze the feasibility of "game changing" IT server strategies or a major technology "paradigm shift."

A shorter "workshop" approach is very effective in situations where an analysis is needed but where extensive and detailed server inventory and cost data is not available. The workshop approach, branded as the Cobra© study, implies a readiness to accept the use of customer validated averages and/or rules of thumb for details such as server original costs and maintenance rates by server class (e.g.: all x86 - 2 socket dual core servers or all UNIX 8 way servers) and utilization rates for entire functional groups of servers, etc. With the workshop approach the study team offers speed and less customer effort and disruption in exchange for a potentially higher margin of error. The findings resulting from both these project approaches enable IT executives to compare and balance alternative technical options and make sustainable server platform and software component decisions.

Method - Scope and Steps

The overall objective of a project can be summarized as defining the current state of the company's IT server infrastructure, describing alternative realistic future states, showing cost projections of the alternatives and recommending a better future state alternative.

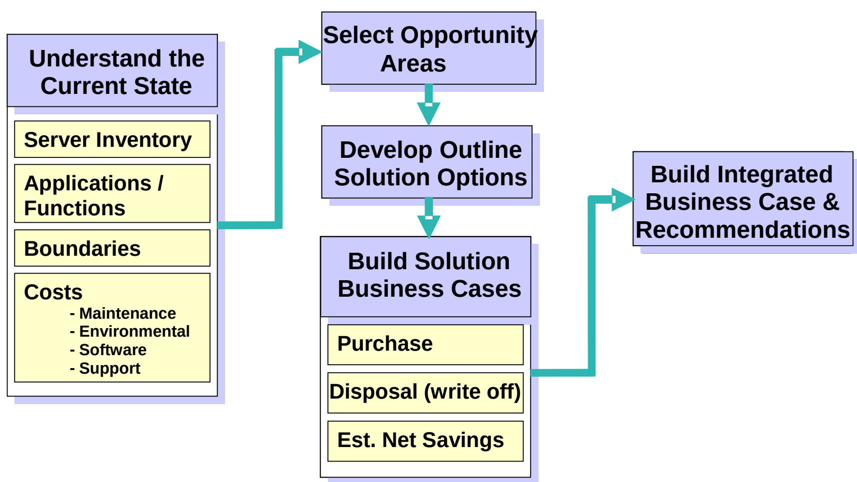


Figure 1: The Scorpion Study Methodology

The tasks followed during a typical project vary in practice but are usually:

1. *Define the study scope with the IT executive sponsor and set expectations.*
2. *Issue data collection tools and templates for collecting server inventory, infrastructure budget and personnel cost data.*
3. *Conduct comprehensive interviews with first and second line IT management and technical staff to validate and complete the server template.*
4. *Analyze the IT budget and personnel costs to identify the current baseline of service delivery costs by server platform and user. These costs typically include hardware, software, staff and facilities costs allocated by the in-scope technology platforms.*

5. *Analyze the server data including function, operational status, line of business ownership, location and applications run, etc. to identify groups of similar servers. This also requires an understanding of current server utilization levels, service availability, outage impacts and contingency requirements.*
6. *Develop brief and informal server "scorecards" that give useful insights into people efficiency, server utilization levels, quality of service delivery and actual current costs.*
7. *Identify outline solution area recommendations to get to a future simpler, more cost effective IT infrastructure. This involves the identification of specific technical solution areas and tactical projects.*
8. *Build comprehensive cost-benefit business cases for the "future state" based on annual and multi-year cost projections for each of the technology areas for which recommendations for optimization were developed.*
9. *Present the final deliverable to the senior IT executive staff, typically in the form of a two hour presentation.*

Data Collection

The basic input data required for this type of project fall into the following four categories:

Infrastructure Budget and Costs:

The study team provides data collection templates and guidance to the IT financial staff who collect infrastructure annual cost data for each major technology platform (e.g. UNIX, mainframe, and Intel servers) and by cost category (e.g. hardware, software, staff, power etc.)

Actual budget data is used wherever possible and is validated during interviews with the IT finance group. This cost data enables the modeling of the actual current cost of service delivery for each technology platform. Further, these data form the foundation for the cost-benefit business cases that will be offered in support of each of our recommendations.

Since the study team concentrates only on costs that are variable and directly proportional to the number of in-scope servers in the infrastructure, these costs are often very different from those currently used by the company for its ongoing IT management, charge backs and control functions. One of the major value-add elements of this method is the creation of this financial model based on hard dollar, recurring and manageable costs for a client to validate and use. This approach provides invaluable insights into often hidden or "gray area" costs that should fairly be apportioned to the specific platforms in the company's service delivery cost model. This is most often true with respect to the costs of personnel.

- **Servers**
  - Server name
  - Primary function
  - Installed location
  - Vendor machine/model
  - Date installed
  - Operating system
  - Memory and disk
  - CP type & quantity
- **Quality of Service**
  - Server outage history
  - Server utilization stats
  - Business criticality
  - Outage Impact
  - Time to crisis
  - Service levels
  - Backup or DR Plan
- **Applications**
  - Application name
  - Application version
  - Registered users
  - Concurrent users
  - Annual growth %
  - Platform affinities
  - In-house or package
  - Criticality rating
- **Software Stack**
  - Operating system
  - Middleware
  - Database
  - Monitoring tools/agents
  - Admin tools/agents
  - Tape backup agent
  - Security tools
- **Financial**
  - Hardware depreciation
  - Hardware maintenance
  - Software depreciation
  - Software maintenance
  - Facilities costs
    - Power
    - Floor space
  - Staffing costs and # 's
    - Admin/Operators
    - Technical Engineering
    - Middleware Support
    - Applications Support
    - IT Staff/Management

Figure 2: Sample Server Data

Normally the IT organization collects the requested staffing data using HR records, organization charts and telephone directory. Interviews are used to validate the number of support people required to deliver service on each server platform. This data includes employees, contractors and any external services provided and often includes only a portion of a staff member's cost based on the proportion of their time spent directly contributing to the overall cost of service delivery. The study team will then allocate the staff costs to the appropriate in-scope servers using a weighted allocation technique. By judging the complexity of each server based on, for example, its function or the application it supports we apportion the staff costs in such a way that a more "complex" server gets a higher allocation than a simple server would get. This yields a much more accurate estimate of any potential savings shown in the business cases.

This approach provides invaluable insights into the often difficult to determine personnel costs that should be apportioned to the specific platforms in the company's service delivery cost model. All data and proportions are, of course, validated during interviews with service delivery and technical support IT executives. This data is also used to produce a series of metrics (e.g. numbers of servers per full time equivalent (FTE)) that the study team may use to compare or benchmark the customer's operations against other study results and the study teams' experience. These insights can help identify whether the client is, for example, broadly people efficient or not. Experience shows that if this people and cost data isn't easy to get, IT costs are likely to be somewhat out of control!

Server Data:

We begin the physical forensic analysis of the IT server landscape by utilizing a detailed spreadsheet that serves as the primary data repository and which forms the foundation of our model and engagement workbook. This would include all of the collected or assumed data on the in-scope UNIX, Intel, and legacy servers to be studied. This spreadsheet provides the analysis team with information including the server name, server vendor, machine type or model, processor speed in GHz, number of core processors, location, function (e.g. Web server, File server, Application server, Database server, etc.), operational status (e.g. production, failover, test, development), operating system version, database, level of criticality, in-service date, and so on. This server data is collected or aggregated from several sources, by the client, and can typically be the most time consuming task in the entire study engagement.

Server Utilization data:

If available, the client is asked to provide sample utilization data for the servers being examined. Most useful is a plot over a representative 24-hour period of server processor utilization every 10-15 minutes.

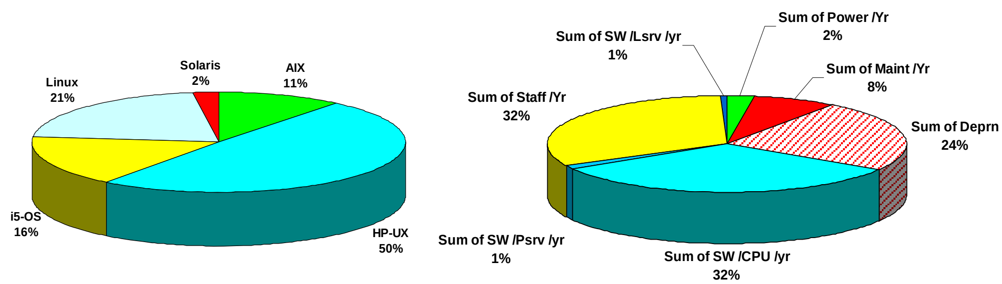
Operational Data and Information:

These data and information come primarily from the interviews conducted with the key executives, line management and key support and operations staff. During our two to three day on site interviews the study team will attempt to gather as much useful information and insight into, for example, managements' strategic goals and primary long and short term objectives, the important issues confronting the organization, their feelings about the infrastructure's strengths and weaknesses, backup and recovery plans and procedures, "availability" data including, for example, service hours, hours of scheduled outage or a recap of unscheduled outages. These insights and information provide a deeper understanding of the reality of the current technological state and help us formulate better conclusions about the environment and, therefore, more effective recommendations about a better alternative future path.

Data Analysis

Financial Analysis:

Once the IT infrastructure budget data collection is complete, the project team builds a model of the IT expenditure by allocating the major costs to each technology platform or service.



Operating System	Power /Yr	Maint /Yr	Admin Staff /Yr	SW /Phy srv / Yr	SW /Lsrv /yr	SW /CPU/ Yr	Deprn/Yr	Total Cost/Yr
AIX	51,088	104,650	358,269	0	6,091	766,393	272,788	1,559,280
HP-UX	201,305	903,086	1,240,919	0	18,273	3,509,732	2,462,817	8,336,132
i5-OS	3,532	153,472	1,782,500	13,734		0	488,499	2,441,737
Linux	109,956	0	1,451,752	149,722	54,819	741,600	478,436	2,986,284
Solaris	15,453	16,100	169,060	0		50,552	45,648	296,812
Grand Total	$381,333	$1,177,308	$5,002,500	$163,456	$79,183	$5,068,277	$3,748,188	$15,620,246

Figure 3: Sample of Financial Analysis

This modeling enables a comparative analysis of the major cost components of each major server group and an informal "benchmark" of a client's costs compared with information from industry sources and other studies. This analysis provides a very useful objective verification of the determined costs of each major platform or application group and often uncovers important costs which are not currently allocated to these technology/application areas. Financial performance and efficiency ratio analyses of overall IT support operations are also developed at this point, as well as "what if" and other pro-forma analyses. The current base costs identified in the financial analysis forms the foundation for constructing business cases and estimating cost savings and the other projections developed in the business cases offered in support of each of our solution recommendations.

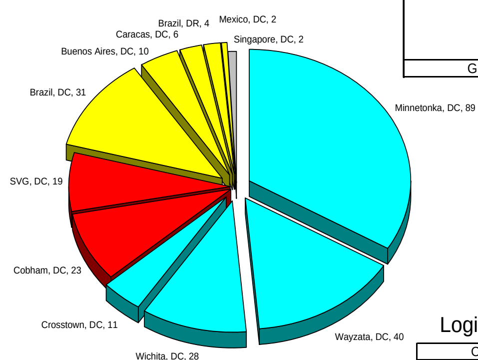
Technical Analysis

There are typically hundreds of UNIX servers and thousands of Intel servers in any major enterprise. One of the first steps is to place these servers into a number of broad major categories and further subcategories, for example: by location.

Server Images by Location, Platform/OS, Vendor Type

Logical Server images by Vendor in 3 major US datacenters

Vendor	Minnetonka, MN	Wayzata, MN	Wichita, KS	Grand Total
Sun	9	2		11
Dell	52	17	7	76
HP	40	23	2	65
IBM	13	3	19	35
Grand Total	114	45	28	187



Graph of WW Logical Server Images
by Town for 12 Datacenters

This pie chart comprises 265 of the 391 total logical server images in-scope. The remaining 126 servers, distributed over an additional 70 sites, were excluded from solution analysis because there were insufficient numbers to justify consolidation or because they were slated for decommissioning, i.e., 63 mail servers going to Exchange

TOTAL IN-SCOPE
Logical Server Image Count

391

Logical Server images by O.S. type in 3 US datacenters

O/S	Minnetonka, MN	Wayzata, MN	Wichita, KS	Grand Total
AIX	7	3	13	23
HP-UX	40	23	2	65
i5-OS	5		6	11
RH7.3	1			1
RHEL2.1			1	1
RHEL3	34	15	4	53
RHEL4	18	2	2	22
Solaris	9	2		11
Grand Total	114	45	28	187

Figure 4: Sample Technical Analysis: Servers by Location

This is a very important step towards developing a server solution strategy for any enterprise. Each of the major categories may have very different technical solutions and cost saving profiles. Practical experience in many projects of this type point to using the following simple categories of servers at this initial stage in the analysis:

Large enterprise-wide applications: It is often easy to spot this class of servers and their associated applications. These servers are often distributed across hundreds of locations, for example servers in retail bank branches and retail stores. They may also be concentrated into tens of regional locations (for example, regional call centers), or into two or three trading rooms. Normally one such major application will have hundreds of associated servers.

Increasingly these may be multi-tier client-server applications with a mixture of Windows, UNIX and mainframe functionality.

IT Infrastructure: Depending on the industry or organization this can often be 25%-50% of the server population and includes those servers associated with a variety of non-application functions in the enterprise. These can usefully be broken down into further sub-categories, for example:

a. e-mail servers
b. File/print/fax serving
c. Systems management
d. Network servers (e.g. PDC, WINS, DNS, Gateways...)
e. Security servers (e.g. firewalls, authentication…)

Medium applications: These are important, usually business-critical, applications that are typically centralized and have less than 20 production servers. This category includes, for example: ERP, CRM and data warehousing. Of course, in addition to these production (and failover) servers, there are usually many others including pre-production, test, QA, development and training servers which can easily double or triple this server count. These applications are very often UNIX operating system based applications.

Small applications: This category includes all applications with five or fewer production servers. This is typically an application server and a database server, or a single combo server. Again, there are usually additional failover, test, and development servers. These small applications are UNIX or Windows based applications.

Other/Unknown: In practice there are always unknown servers. This category should always be less than 10%, and ideally less than 5% of the total server count.

Other analyses are done on the server infrastructure with the aim of both helping the study team understand the environment as well as to offer valuable insight to the customer of their own server infrastructure.

We estimate that 31% of the installed 340 servers reviewed are 5 years old or older, while nearly 60 of the servers are 7+ years old. These systems may require less labor to support them, but they represent a potential exposure to the firm's long term growth and eventual move towards an On Demand infrastructure.

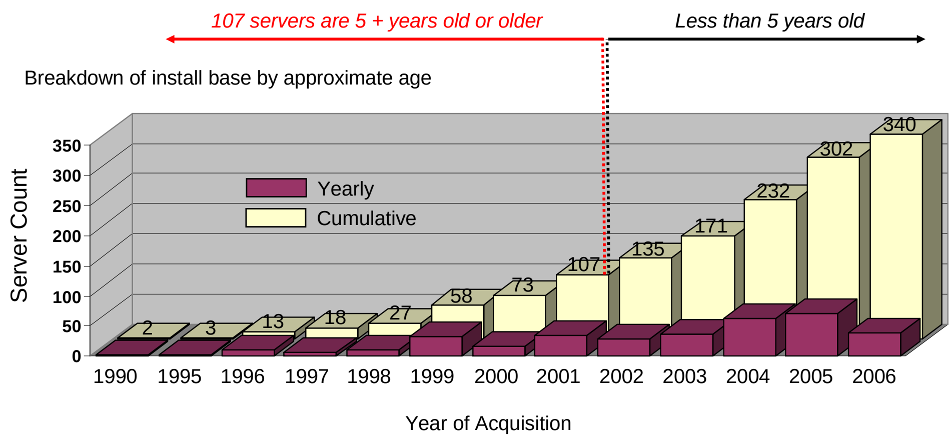


Figure 5: Sample Analysis of Server Ages

Solution Design

The next major step is to generate recommendations for alternative future states for each of the solution areas, such as Web servers, database or infrastructure servers, or a specific group of application servers. These recommendations target the current areas that may show poor service quality, higher than average costs of service delivery or are the foundation for a longer term strategic shift in the IT business and its associated technical model. There are a variety of options available for architecting these recommendations. They most naturally breakdown into: virtualization, encapsulation, server technology and system redesigns as detailed in the following chart:

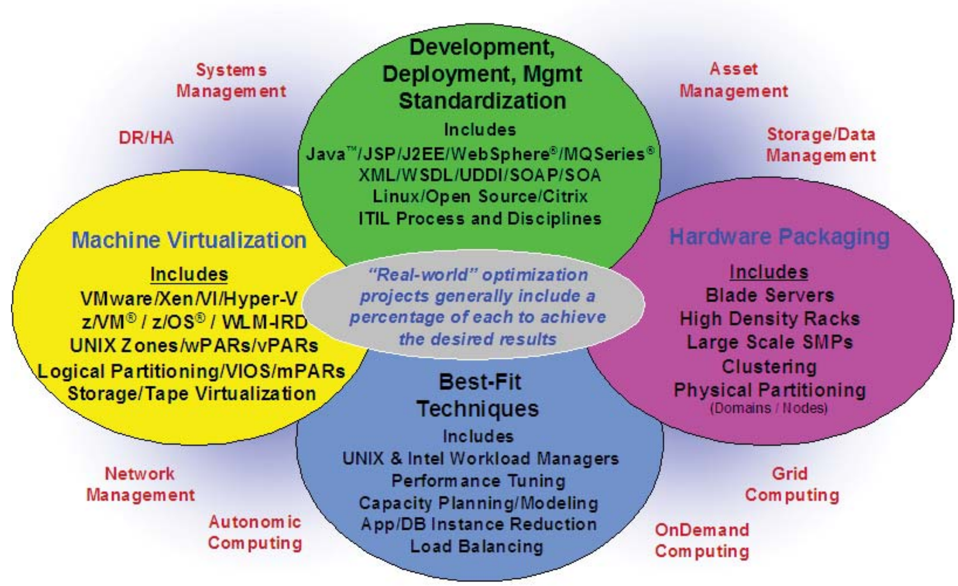


Figure 6: Approaches to IT Server Optimization Tools and Techniques

Realistic alternative configuration recommendations are developed based on the technical options, along with the hardware, software and people cost data. The realities of the specific situation are factored in, including actual server utilization and actual staff efficiency. The solution parameters then form the baseline for the actual business cases analysis. The business cases use classic cost/benefit analysis to compare the alternative future state with the current "continue as usual" case. These business cases can and often include multi-year projections, ROI and break-even analysis and can also include, where appropriate, "what-if" analyses that explore the business case impact of, for example, different utilization levels, different people efficiency ratios or different hardware configurations.

Sizing	Current	Alt.Case	Change	
Server Type				
Total CPUs	8,584	3,312	-61%	
Used CPUs			0%	
#Logical Servers	3670.0	3670.0	0%	
#Physical Servers	3670.0	178.0	-95%	
Ave.Log.Srv RIP	885.6	407.9	-54%	
Total RIP Capacity	3,250,128.3	1,496,818.3	-54%	
Total RIP Workload	482,259.0	482,259.0	0%	
Ave CPU %Utilization	14.84%	32.22%	117%	
Annual Operating Costs (AOC)				
Staff Cost Code				
Software Cost Code				
Software Cost /CPU				
Software Cost /Lsrv				
Software Cost /Psrv				
Software M&S	7,328,350.00	3,164,313.00	-57%	
Hardware Maint	3,668,700.00	367,514.00	-90%	
Space	0.00	0.00		
Power	3,344,146.00	317,490.00	-91%	**1.0** :1 Log.SCON ratio
Staff Cost	16,941,944.00	13,553,557.00	-20%	**20.6** :1 Phy.SCON ratio
			0%	
Total AOC	**31,283,140.00**	**17,402,874.00**	**-44%**	**13,880,266** est.potential saving /yr
One Time Costs (OTC)				
Software Purchase	0.00	598,260.00		
Hardware Purchase	0.00	6,289,026.00		
Total OTC	**0.00**	**6,887,286.00**		**6,887,286** Net Cash Investment
Write-Off	0.00	0.00		
				0yr 6m Payback Period
5 Year Projection				
OTC + 5x AOC	**156,415,700.00**	**93,084,142.00**	**-40%**	**63,331,558** 5yr saving

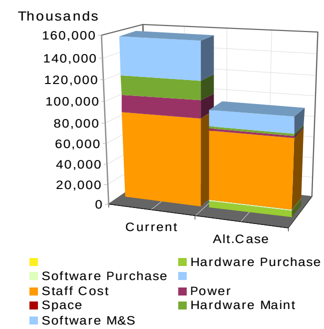


Figure 7: Example of a Scorpion Study Summary Business Case

This sample business case shows the current servers in the column headed 'Current' and offers a recommendation on an alternative in the columns headed 'Alt Case'. The top section of the table contains the detailed information on the servers, their composition, their relative power/performance ratings and rates of utilization. In the second section we explore the current and alternative Annual Operating Costs or AOC; following below that are the One Time Costs or OTC's that represent the required investment in new hardware and, if appropriate, new software necessary to transition the infrastructure to the alternate, optimized state. Finally in the bottom section of the table are the multi-year projections of the current "business as usual" costs and the alternative optimized costs. On the right side of the business case are various cost and benefit analyses, for example: savings per annum, consolidation ratios, estimated payback period and net multi-year savings.

Facilities and Environmentals

A report released by the Department of Energy in 2006 stated that the then current projections anticipate overall U.S. energy demands to increase by more than one-third by 2030, with electricity demand alone rising by more than 40 percent[4] More specifically, another study done by the Environmental Protection Agency[5] states that under current efficiency trends, national energy consumption by servers and data centers could nearly double over the next five years.

Energy and Climate	Current	Alt.Case.3	Alt.Case.2	Alt.Case	Change	Difference	
avg RackU / Server	1.9	4.0	4.0	4.0	110%	2.1	
Total RackU	1,981.0	100.0	156.0	156.0	-92%	-1,825.0	
30U Racks	66.0	3.3	5.2	5.2	-92%	-60.8	
Total kW	266	24	38	38	-86%	-228	
Adjusted kWh /Year	3,499,053	317,154	494,760	494,760	-86%	-3,004,293	Green saving equivalent to
Heat BTU /Hour	608,494	55,154	86,040	86,040	-86%	-522,454	1,292 Tonnes CO_2 /Year
CO_2 tonnes /Year	1,505	136	213	213	-86%	-1,292	or 4,273 Trees; or 515 Cars
Carbon tonnes /Year	411	37	58	58	-86%	-353	
RIPs /m2	7,144	46,375	37,100	37,100	419%	29,956	
RIPs /Watt	3.5	12.8	10.3	10.3	189%	6.7	
RIPs /BTU /Hour	1.6	5.6	4.5	4.5	1.9	2.9	
RIPS / tonne CO_2	627	2,267	1,814	1,814	189%	1,187	
W /m2	2,015	3,618	3,618	3,618	80%	1,603	

Figure 8: Sample Environmental Analysis

In response to the renewed interest in environmental costs, our analysis not only includes the comparison of current versus alternative costs for power and cooling and, where appropriate, space, but we've also added an extended analysis comparing our recommended alternative state to the current state in areas such as: total kilowatts consumed, total square footage occupied (expressed in "rack units") as well as carbon offsets and other extended environmental analyses. These are all aimed at helping our customers explore evolving to a "greener" technological footprint.

Indeed, Scorpion IT Rationalization studies completed over the last two to three years have shown a marked increase in customer focus on the facilities portion of the IT cost equation. We have seen that five years ago, facility's costs typically accounted for only 3 to 5% of the total cost of the IT infrastructure. These costs now account for upwards of 8 to 10%. Although this doubling in the operational expenditures for power, cooling and space are somewhat alarming, their source of escalation may be a surprise to some corporate executives.

---

[4] "The National Action Plan for Energy Efficiency Report", July 2006
[5] "Report to Congress on Server and Data Center Energy Efficiency" Public Law 109-431, August 2, 2007

The rise in the operational expense for facilities costs, and in particular power and cooling has little to do with the cost to generate and distribute that power. In fact, the cost for a kilowatt-hour of electricity has risen very little over the last 5 to 10 years, and certainly does not account for the doubling affect that many IT shops have seen. The true cause for such a dramatic increase has its roots in the seemingly unchecked growth in the proliferation of higher density servers per rack and blade packaged servers. These creatively designed and packaged systems, are allowing customers, to put more processing power in to an ever increasingly smaller space. The resulting affect has been a drastic rise in the facilities electrical energy consumption to both operate and cool these systems.

All of the above not withstanding and while the operational costs for power and cooling are still an important factor in any financial business case analysis, these costs can still take a back seat to the potentially monstrous capital expenses (CapEx) and costs associated with the building of new facilities, infrastructure build-out or expansion required to accommodate these rises in power and cooling requirements.

An example of this was evident in a recent study completed for a large urban healthcare organization in the US. An initial cost analysis of the primary datacenter for this company revealed a 30% per annum increase in the power requirements for the servers and storage.

Again, most of this rise was attributed to the rapid growth in server units rather than the actual per unit cost of electricity. However, this increase still only accounted for approximately 9% of the overall yearly IT infrastructure operational cost.

As it happened, the greatest potential cost impact associated with this drastic rise in power consumption, had to do with the constrained power handling and distribution capabilities within the datacenter. With the uninterruptable power supplies (UPS's) nearing their maximum rated capacity, and with the datacenter in question unable to accommodate an upgrade in power input from the providing utility company, the only alternative was to expand its present facility at an estimated capital cost of nearly $25 Million dollars, a figure which dwarfed any concerns about their operational costs discussed above.

The solution in this case was to virtualize the x86 and UNIX server portfolios using a phased, but aggressive approach, thus allowing the present datacenter to meet its power needs through 2012, ample time for it's planned relocation to a new facility within that same timeframe. This certainly had the positive affect of lowering the overall power (and space) impacts on annual operational costs but the primary driver in moving forward was in avoiding the unaffordable capital costs of datacenter expansion.

As a result of these shifts and their implications on IT costs and in part because of their success in putting greater amounts of computing into smaller spaces, many hardware vendors have been challenged by their customers to develop systems which consume less power coupled with their advances in hardware and operating system software technology that result in never before capabilities for workload consolidation and image virtualization. This further underlines the importance of the Scorpion Methodology as an analytic approach to understanding the cost impacts of power, cooling, and space in operating a datacenter and the IT infrastructure within.

Conclusion

The Scorpion Methodology has been used in hundreds of situations in over 20 countries worldwide in the last few years. It has been extremely well received by many senior IT executives and has helped bring significant and highly useful insights and effective cost and benefit data and insight to strategic decision making in the most modern corporate IT organizations.

One of the many lessons learned during the course of these many projects concerns direct and indirect costs. In almost all projects there has been a significant difference in the "real" costs of service delivery, identified by the study, and the service delivery cost model the organization uses for accounting, user charge backs and ongoing management and control. The accounting and cost management methods used by any company are usually unique to that company and have evolved over time. They often stem from specific individuals within the organization and the history, practicalities and politics of that IT organization and its user base. The much more direct "costs of service delivery" that these Scorpion study projects identify are often surprising and occasionally offer an organization a first time look at the real, manageable and variable costs involved and, very often, have very profound implications. These truly manageable costs of service delivery usually need to include all direct and indirect costs, regardless of the organization of which, for example, an FTE is part or the cost center in which an expense is shown or even on whose budget that cost may appear. Indeed, sometimes, the primary objective of one of these projects is to estimate, as closely as possible, the actual total cost of delivering service for each application group or server technology platform compared with the planned budgeted expense.

In addition to the IT Systems Rationalization or "Scorpion" Study and the IT Optimization Workshop or "Cobra" Study discussed in this paper, the IBM Systems and Technology Group's Laboratory Services IT Infrastructure Optimization and Virtualization Services team offers a number of additional infrastructure optimization studies. These offerings include the following:

- *Virtualization Rapid Assessments (VRA) - These highly targeted assessments assist our customers in advancing their current virtualization implementations to higher levels of effectiveness.*
- *Systems Green Study – This study is designed to help our customers' IT organizations more fully analyze their datacenter power, cooling and space issues.*
- *Comparative Total Cost of Computing Study or "Polaris" Study – This study is designed to analyze and compare a customer's cost of computing for selected, usually mission critical, applications across their legacy, open and distributed technology platforms.*
- *Information Infrastructure Storage Optimization Study and Information Infrastructure Storage Optimization Workshop - These projects are designed to help analyze and optimize a customer's storage infrastructure*
- *Storage Energy Analysis – This engagement is designed to help our clients analyze the power, cooling and space impacts of their storage subsystems.*

Clearly, it has been shown over many of hundreds of occasions and across a myriad of customer industries that the findings produced as a result of these projects can be extremely valuable to IT decision makers. Regardless of whether the participating companies are small or medium sized or global in scope, the resulting findings, conclusions, recommendations, and cost-benefit business cases from Scorpion studies can be key inputs to any corporate IT optimization project. Further, these assessments and their accompanying insights have been proven to help senior IT executives considerably in their efforts to develop better and more balanced IT strategies and reduce their total IT infrastructure costs.

Rick Schoenmann is a certified IT infrastructure consultant with the IBM Scorpion team. He has over 30 years of experience in the IT industry and has played a senior role in the delivery of many significant IT consulting projects. Rick can be reached at raschoen@us.ibm.com.

John F. Ryan is a senior IT consultant with the IBM Scorpion team and is a Certified IT Specialist with IBM and the Open Group©. Over the last ten years, he has contributed significantly to the definition, execution and development of IBM's portfolio of IT rationalization and consolidation tools and analysis methods. He can be reached at jfryan1@us.ibm.com.

Copyright IBM Corporation 2009
IBM Systems and Technology Group
Route 100
Somers, New York 10589
U.S.A.
Produced in the United States of America,
02/2009
All Rights Reserved

IBM, IBM logo, AIX, MQSeries, WebSphere, z/OS and z/VM are trademarks or registered trademarks of the International Business Machines Corporation.

Adobe, the Adobe logo, PostScript, and the PostScript logo are either registered trademarks or trademarks of Adobe Systems Incorporated in the United States, and/or other countries.

Cell Broadband Engine is a trademark of Sony Computer Entertainment, Inc. in the United States, other countries, or both and is used under license therefrom.

InfiniBand and InfiniBand Trade Association are registered trademarks of the InfiniBand Trade Association. Java and all Java-based trademarks are trademarks of Sun Microsystems, Inc. in the United States, other countries, or both.

Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

Intel, Intel logo, Intel Inside, Intel Inside logo, Intel Centrino, Intel Centrino logo, Celeron, Intel Xeon, Intel SpeedStep, Itanium, and Pentium are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

UNIX is a registered trademark of The Open Group in the United States and other countries.

Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.

ITIL is a registered trademark, and a registered community trademark of the Office of Government Commerce, and is registered in the U.S. Patent and Trademark Office.

IT Infrastructure Library is a registered trademark of the Central Computer and Telecommunications Agency, which is now part of the Office of Government Commerce.

All statements regarding IBM's future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only.

Performance is in Internal Throughput Rate (ITR) ratio based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput that any user will experience will vary depending upon considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve throughput improvements equivalent to the performance ratios stated here.

ZSW03062-USEN-00