The Case for Server Consolidation: Or Back to the Glass House?

Ellen M. Friedman
Jerry L. Rosenberg
SRM Associates, Ltd.

Abstract

Over the past two decades, Information Technology has witnessed the morphing of computing environments from single large monolithic entities to widely distributed server configurations. The initial promise of less costly, more scalable and flexible platforms has not been realized. To the contrary, many organizations have found that a consolidation of distributed servers back to a more centralized approach offers greater control, standardization, performance and cost savings. This swing back of the pendulum should be considered by all organizations to determine if it is appropriate to redefine their computing environment. This paper addresses the issues involved in server consolidation opportunities and offers recommendations for initiating this effort.

Overview

Corporate Information Technology (IT) agendas are changing once again. Priorities are shifting back from business units, divisions and departments to large-scale organizational systems that support consolidated enterprise requirements. The new goals are business integration and transparency of information across the entire enterprise. Infrastructures are being put in place to support this "new" paradigm. In many organizations, proliferation of servers in the 1980s and early 1990s left legacies of fragmentation and inefficiencies that must now be removed before new challenges can be met.

Integrated systems are too large to run on small servers, and are too critical to entrust to the instabilities and management issues that exist in these environments. The focus of enterprise application deployment must move to larger-scale, more stable platforms. Consolidation is the solution.

The proliferation of servers throughout the enterprise has brought many companies to the point where they have gone beyond obtaining value from this option and are now suffering from its excesses. Too many distributed servers present serious problems in the areas of customer and user service and operational costs. In terms of service delivery, the following list contains some of the more frequently encountered problems:

- Service availability (Outages of significant duration)
- Service response times due to LAN bottlenecks or overload
- Difficulty of developing new cross-application or data services
- Ability to support Web serving based on secure and consistent data
- Uneven data integrity & security

Initially, small server environments were implemented under the guise of savings from reduced server costs. With experience, many organizations have found that the actual operating costs of these systems is

not limited to server cost and total cost to the company is actually higher in the distributed environment. In terms of operational costs, the following problems contribute to higher overall costs:

- Controlling system management costs are more difficult
  - Network management
  - Configuration management
  - Problem and change management
  - Operational management (for both automated & manual operations)
  - Security administration costs
- Requirement for diverse skills and their dispersion in the enterprise
- Hidden costs inside departmental operational costs
- Excess server-based software license costs
- Increased costs due to poor utilization of resources: servers, disks, maintenance.
- Lack of scalability: demand growth results in further server proliferation

Many organizations are looking to consolidation as the solution for addressing changing IT and business requirements. Larger, more adaptable computing platforms, concentrated data availability, reduced network traffic requirements, and single point of system management and support offer significant advantages to mission critical applications.

Benefits

The benefits of consolidation revolve around the reduction of complexity and cost offered by regaining control of the IT infrastructure. The benefits provided by consolidation can be summarized as:

- Reduced costs of delivering service
- Scalability for prompt reaction to changing business needs
- Employment of a more standard, controlled infrastructure offering greater service availability
- Leveraging of investments in existing corporate resources
- Opportunities for new services by enabling or improving interaction between different applications and access to data residing in legacy systems

At a more detailed level, outages and disruptions are minimized, while data is made more accessible and is more effectively managed and protected. Overall technical complexity can be minimized and support staff issues are greatly reduced. Consolidation offers the centralization of the skill pool in a single location. A central pool of technical talent is a more productive and efficient solution since fewer resources will be devoted to managing and supporting a server infrastructure. For many organizations, skilled technical staff to maintain and support the infrastructure is a serious concern. Centralization reduces the overall staffing requirements and allows for global solutions to technical problems.

The single point of control offers a disciplined approach to managing IT resources. Firms that are growing have particularly found that a decentralized mix of distributed systems is too unwieldy to manage. These growing companies also see that consolidation offers the ability to add new workloads and applications at a lower incremental cost and with little or no staff increases. There is an ability to support a larger number of users accessing the same application by running on a larger server than may be cost justified in a distributed environment.

In short, less management is the operational theme here. A distributed system is far more labor-intensive than a centralized model. For each server, costs involve asset management, capacity planning, support strategies, training, security auditing and control, vendor selection, change management, migration strategies, and service agreements. Reducing the number of servers reduces the time and expense it takes to

perform these tasks. Reduction in the number of servers also results in lower maintenance, software license and disk costs. Moreover, the lower costs of communications make remote access more affordable.

Why Now?

The industry trend in the late 80's and early 90's was toward decentralization and departmental control. This was primarily in response to large application backlogs and perceived unresponsiveness to business problems. Lower acquisition costs and relatively lower labor costs led to decentralized computing and a proliferation of many different technologies. Over time, however, many firms experienced unexpected drawbacks including reduced IT efficiency, higher cost-of-ownership, lack of standards, increased support staffs, scalability problems, etc. The push towards client-server computing seems to have reached its peak in the mid-1990s and now the pendulum appears to be swinging back. Improvements in the price performance of today's Unix servers, along with improved scalability, multiprocessing and high availability, is driving the trend towards centralized control with all the systems management and staffing benefits that come with it.

Infoworld magazine (9/15/97) presents a David Letterman style Top 10 list for whether your firm is currently a good candidate for server consolidation.

It's time to consolidate when:

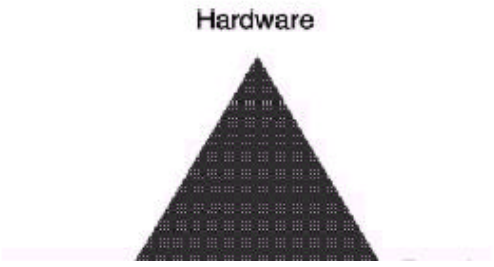
1. You start losing track of your servers.
2. Your hardware is having seizures over scalability.
3. You have more systems administrators than you have users.
4. You're running 15 different operating systems.
5. You don't know if you are in compliance with all of your software licenses.
6. Capacity planning is a synonym for buying more servers.
7. Department managers routinely purchase and install their own servers (but leave management to you).
8. Utilization rates for more than half your servers are in the single digits.
9. Physical security of every server could only be accomplished by Star Trek-like shields (rather than locking the door to a single room).
10. And the No. 10 sign that your company is a prime candidate for consolidation: You spend more money on server upkeep than the U.S. government owes.

If you find that you are answering YES to several of these issues, it is time to seriously consider consolidation.

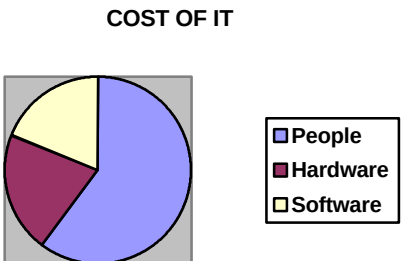
Cost Considerations

Server consolidation offers the opportunity to

- Recentralize the IT effort
- Merge workloads for cost and manageability
- Combine servers into a common system



ıajor components:



- Hardware includes the cost of equipment and maintenance as well as facilities expenses.
- Software includes licenses and maintenance
- People costs include salaries, benefits, education, office space, etc.

It should come as no surprise that the people costs have been growing rapidly while hardware and software costs have been going down steadily. The major savings from server consolidation are derived from significant reductions in people costs.

In the past, many IT organizations have made attempts at cost reduction which have been less than successful. Often the plan involves addressing only one of the cost areas and ignoring the other two.

- Minimizing hardware costs by limiting the capacity provided is one such technique. Unfortunately, this almost always results in an increase in people costs. Too much time is spent managing the system rather than getting proper work done. This approach has been very prevalent in PC environments.
- Minimizing software costs is often attempted by keeping back-level versions of software and only upgrading when absolutely necessary. This approach also increases people costs by requiring excessive monitoring and it does not allow for technology exploitation available in newer releases.
- Minimizing people costs usually revolves around the imposition of very strict standards, such as single vendor solutions. This solution often limits flexibility for users of the system and can lead to higher hardware costs.

The effective solution should review all cost components to arrive at a balanced resolution. Consolidation offers the potential for balancing these cost issues.

Cost Analysis

81 percent of the IT managers that had recently completed a consolidation project were pleased to find that their efforts had produced significant cost reductions. According to a Gartner Group study (K-LAN-308: Dave Cappucio; February 27, 1995), consolidating six small servers into a pair of larger machines provides a substantial total cost of ownership savings of 35%-40%, primarily in internal support costs.

The Gartner study took into account that servers required the following internal support disciplines that accounted for the majority of the savings:

- Storage Management
- Backup
- Administration
- Hardware Activity
- Trouble Shooting
- Performance Tuning
- Monitoring

The costing analysis for these servers arrived at $57k for a 600 server distributed environment. Assuming a consolidation to 200 nodes onto 2 SP central servers costs were reduced to $23K. This amounts to 60% reduction in the cost of system support. In addition, the overall quality of backup, trouble-shooting, performance tuning and monitoring were felt to be low in the distributed environment and significantly improved after consolidation.

The benefits revolve around the need to regain control of the IT infrastructure. This control includes the management of costs. Specifically, hardware costs, software licenses, facilities requirements, cost of

outages, manpower requirements, and the loss of business leverage caused by the lack of complete data in a timely fashion should all be considered in evaluating consolidation.

In a second analysis (found in "Consolidating UNIX Systems onto OS/390", IBM Redbook SG24-2090-00), the costs of multiple UNIX servers and a single centralized system were compared. In this case, 36 UNIX servers, each running an independent Sybase application, were consolidated onto a 36-node RS/6000 SP. The system management capabilities of the SP allowed the user to reduce systems administration staff. This led to a monthly reduction in support labor cost of 60%. It should be noted that in the 36-server environment, support labor accounted for half of the expense. The hardware/software depreciation was virtual the same in both cases (The SP was actually 3% lower) and there was a 27% reduction in maintenance costs on the SP. The overall cost savings on the SP were 35%.

In addition, the following observed savings were not included in the analysis:

- Utilizing the load leveler to drive up the effective utilization of the processors reduced the required number of nodes.
- The user did not have to buy system management tools for the SP.
- There were reduced networking costs due to the internal network on the SP when using the high performance switch.
- People costs go up over time. Consolidation offers more protection from this escalation.

The key to both of these studies is that the cost of administrative support (people) is significantly reduced by server consolidation efforts. The people costs are already the major component of IT budgets and it is likely that they will grow in intensity while hardware/software costs will diminish. Therefore, consolidation offers a considerable cost savings potential while affording a more robust, integrated platform.

Cost Justification versus Value Justification

While some specific cost savings can be and have been readily demonstrated, comparing the total cost of computing for distributed versus consolidated servers sometimes proves very difficult. User departments remain focused upon achieving organization objectives, not in collecting apportioned costs of personnel involved in part-time support. In fact, customers considering consolidation frequently discover that costs of part-time support are not even recognized, much less gathered into a composite total. The very inability to determine costs of distributed server computing becomes a compelling argument for consolidation into a professionally managed central site. Cost accounting and chargeback, common components in centralized implementations, are noticeably absent or minimally employed in distributed server environments. That is, while many customers believe that substantial system management personnel cost savings certainly do accrue with consolidation, they recognize the futility in attempting to accurately ascertain those savings. Instead, customers might better focus on the added benefits gained from a consolidated environment relating to the other technical issues. The case can be readily made even if the cost of service is the same.

General Sizing Considerations

Each of the various classes of servers has different architectural characteristics, making some more suitable for satisfying business requirements than others. These classes of servers range from small PCs to midrange processors, to large mainframes supporting millions of mission critical on-line transactions per day across the enterprise. PCs and many UNIX implementations tend to be optimized for processor intensive work. S/390 platforms are more suited to data intensive work and have a sophisticated I/O

subsystem that can support very high data rates. Additionally, many commercial packages such as SAP are available to be ported on multiple hardware platforms. In considering server consolidation, one must not loose sight of business requirements while choosing a platform which can both satisfy current needs as well future ones.

In planning for systems it is most important to take a longer-term view of how business will grow over time and choose a platform that can scale accordingly. Still one needs to be concerned with solutions that best fit the current application's requirements but also to consider possible future benefits, such as integration with other applications, user productivity, and timeliness of information. The long term plan should include a 5 year total cost projection for running the application, provide solutions which are best for the application as well as the entire enterprise, and consider options which facilitate data to be readily access by key corporate decision makers, optimizing DSS and e-commerce applications.

The top 10 application issues to consider in evaluating server consolidation options and platforms:

1. Data Accessibility/Data Sharing- need for consolidated business views of data such as DSS and Data Warehousing
2. Data security and integrity requirements
3. Ease of integration with other enterprise applications and growth
4. Performance Expectations
5. Reliability and Availability requirements
6. Manageability- System management ease and cost
7. Availability of technical personnel
8. Scalability
9. Recovery and Contingency considerations- Backup requirements and failure/recovery operational and time considerations.
10. Ability to exploit new technology.

Server Consolidation Opportunities

Many organizations have experienced increases in cost and operational complexity as the number of servers they employ to support their business needs has proliferated across the enterprise. It is much more difficult to meet service levels, availability requirements and general system needs as applications grow in a decentralized, distributed environment. In order to reduce costs and gain control of the infrastructure, many organizations are consolidating their servers into a more manageable centralized environment. In considering a consolidation, the options range from one requiring minimal perturbation and restructuring to a full scale application and database redesign. The consolidation stages are:

- ***Physical consolidation- Creating a Glass House- centralize all servers in one location***
- ***Operational Consolidation- create one logical machine and consolidate servers into a single system:***
    - Moving the distributed servers onto separate nodes of a clustered multi-node UNIX server e.g, an IBM RS/6000 SP or Sequent Server.
    - Moving the distributed servers onto separate partitions of an S/390 system or a Sun Enterprise 10000 system configured with Dynamic System Domains (allowing it to be configured as logically divided into multiple servers)
- ***Application Integration- Consolidating the application to run as a single large application- requiring restructuring of the application and database***
    - A restructured application and DBMS such as DB2 on an IBM G5 wiithS/390 (e.g., converting SAP application to use DB2)
    - Run multiple different applications under S/390.

Server Environments and Consolidation Benefits and Opportunities

The overall benefits for consolidating any distributed application environment are:

- Reduction in operational costs
- Better availability of service
- Improved systems management
- Better version control and software distribution.
- Possible performance improvements

In addition, there are benefits of and opportunities for consolidation in each of the following server environments:

1. **Consolidating multiple copies of the same application which is replicated on multiple distributed servers**
2. **Consolidating a single application which is distributed across multiple servers e.g., 3-Tiered Implementation**
3. **Consolidation options and benefits for Multiple Distributed Applications**
4. **Consolidating LAN Servers- e.g. print and back-up**

1. Consolidating multiple copies of the same application which is replicated on multiple distributed servers

Multiple copies or instances of an application may be required if the application and or database cannot scale on a single platform. The options as with any server consolidation are:

- Centralize the servers
- Consolidate operations on a single partitioned server,
- Re-architect to run the application on a single system image (DBMS and application changes may be required)

When we consolidate multiple instances of the same application, which execute across multiple servers into a single centralized location, the following savings due to economies of scale and greater flexibility can be achieved:

- Reduction in system management costs and improvements to process due to more simplistic management.
- Reduction or elimination of user department operational costs (e.g., shared facilities in a centralized environment)
- Possible improvements in physical security- (do not need to secure multiple small environments)
- Improved recoverability after an outage that can be achieved due to both more available personnel and better hardware such as faster tapes to perform restores.

The next step of operational consolidation brings with it additional benefits over merely physically moving the servers to a single location. With a multi-node cluster such as an IBM RS/6000 SP platform, the following savings can also be realized:

- Possible reduction in software licenses
- Improved availability over single server solutions- or at least automated reliability and availability achieved via the RAS hardware as opposed to application based solutions.
- Reduction in systems costs including operational, maintenance, number of systems to manage over and above what can be achieved by merely centralizing servers.
- Improved service times due to reduction in inter-system communication.

2. Consolidating a single application which is distributed across multiple servers e.g., 3-Tiered Implementation

There are minimal differences between consolidating a single application which is replicated across multiple servers (i.e, same application in its entirety is replicated in multiple environments e.g., London, New York and the Far East) and one which is distributed across multiple servers (such as a 3-tiered implementation or separate application and database servers). Therefore, all of the same benefits cited above can be achieved for multiple instances of the same application can be achieved for true distributed applications.

The only difference between the replicated application and distributed application is the actual application design and architecture. One example, illustrating the differences in design to consider is partitioning the application by regional area (such as different sales regions). In the replicated case, each regional area would log on to a particular server and only back-end marketing programs (as necessary) would bring data together from each independent server. In the distributed case, all regions would log onto a central server, the DBMS would be partitioned (by region) and architected as a distributed DBMS. The application would appear as a single entity to the end user.

The most common example of a distributed application is a typical Client Server, 3-tiered implementation consisting of the Presentation Layer (i.e., GUI), Application Layer and the Data Layer. These layers can be spread across multiple platforms, with standard interfaces used to communicate between the layers. This is the typical client server design, however, it does not require that each layer executes on a different platform. In fact, there could be savings realized by minimizing intersystem communication between portions of the application.

Additional benefits can be accrued by operationally consolidating the distributed application to execute on a multi-node cluster such as Sun Enterprise Server 10000. With Dynamic System Domains, each of the distributed components can execute in its own domain on the same platform and improvements in response time can be achieved by reducing communication between the applications resulting in faster, more reliable and more responsive service. There are opportunities to consolidate both the application server and the database server in a three-tier client/server application with consequent improvements in reliability and performance.

3. Consolidating multiple instances of different applications which execute on multiple servers Consolidation options and benefits for Multiple Distributed Applications

An example of opportunities to be exploited when consolidating multiple distributed applications can be seen in an example from the financial industry. A Portfolio Management System consists of Equities, Bonds, Multiple Currencies and other instruments. Often these applications have been architected to run on separate servers with SQL*Net and accesses and NFS usage to link the systems together. Ultimately, there is a need to integrate the back end pieces for viewing by the account executive and to facilitate Positions and Options as well as for statement creation. As in the three-tiered application example, savings can be achieved by minimizing any intersystem communication. But additional benefits could be accrued through opportunities for additional business as a result of integrating these applications. Opportunities for e-commerce and electronic trading may also be realized by integrating this application with the trading applications.

The consolidation of multiple Web servers in an organization is another example that can be considered for understanding the opportunities and benefits of moving multiple distributed applications to a centralized location either on the same distributed servers or to operate on a different single server platform such as a Multi-node cluster. In this case, the closer proximity of the servers can facilitate improved communications and thereby result in improved response times and higher availability of service. Web Serving/Electronic commerce can also be facilitated based on secure access to data residing in Legacy databases.

The multi-node cluster architecture as well as the Enterprise Sun Servers and S/390 provide high availability that cannot be accomplished in the distributed server environment. The smaller servers such as

Sun SPARC 4000's do not offer high availability so achieving this level of service requires application and software based disaster recovery planning with additional servers for hot back-up. In addition, these larger platforms such as IBM SP, S/390 and Enterprise 10000 server offer greater scalability particularly for single applications.

As in the previous option, the benefits for moving to a centralized location include:

- Reduction in systems management costs including maintenance costs as well as improved systems management
- Reduction or elimination of user department costs
- Improved availability of service (fewer or no outages, and shorter outages should they occur)
- Improved recoverability after an outage
- Improved physical security
- Potential for improved service times to the end user due to a reduction of the time to communicate between clients and servers in single location

The move to a multi-node cluster provides the following additional benefits:

- Reduction in software licenses
- Reduction in operator costs (fewer footprints to manage)
- Reduction in number of systems to manage
- Opportunity for new services by enabling or improving interaction between different applications, or between new applications and data in existing systems.

Finally the move to S/390 or Sun Enterprise Server 10000 with Dynamic Systems Domains (dynamic or logical partitioning) allows the deployment of multi-tiered applications on a single server, greatly reducing management complexity and overall cost of ownership. Improvements in access to operational data as well as opportunity for new services can be realized on these integrated highly scalable platforms.

4. Consolidation of LAN Servers - e.g. print and back-up

Centralizing servers offers the benefits associated with lower operational costs and the opportunity for improved systems management that can be accomplished with a pool of skilled staff in a single location. The further consolidation to fewer servers can improve the service offerings to the end user on a remote LAN.

The benefits of consolidating to either a Clustered machine or separate partitions on a single machine for this type of application include:

- More reliable backups performed by central data center.
  - The use of higher speed disks and tape for backup and recovery.
  - With central backups, the equipment and support costs are reduced, while the reliability of the service is increased
- Improved systems management by administering LAN environment from a central site
- Enabling effective distribution of software to multiple LAN servers from a single location
- Improved physical security
- Ability to better manage systems storage – increase overall capacity as required in a more cost effective manner
- Allow users of centralized applications to print files to LAN printers
- Reduction in cost and complexity of print environments by supporting enhanced print stream transforms and enhanced host-to – LAN print services.
- Improved scalability over multiple distributed servers

Server Sizing Considerations in Consolidation Efforts- The Enterprise Server Requirements

Business imperatives such as the force of e-commerce and e-business, and the need to align I/T with business goals without comprise are driving the selection process toward enterprise class servers. Server consolidation is a necessity for businesses to remain competitive for analysis and DSS systems. Combining multiple instances of "data marts' into a single server entity such as S/390 is necessary in order to:

- Reduce the effort and time in consolidating varying views of critical databases distributed across the enterprise to able efficient Business Intelligence processing
- Improve response time of DSS based queries
- Enable efficient processing of a wide spectrum of concurrent query types (I/O intensive vs. CPU intensive).

All sizing considerations are dependent upon the application requirements, database requirements and architecture of the platform on which the system is based. There are many standard benchmarks that can be used to compare processor power. These include tpc-A, tpc-C as well as specific Oracle and SAP industry benchmarks. They provide a means of comparing the theoretical performance of different UNIX vendors. These benchmarks can be used to get an initial sizing of the processor, memory and disk requirements of specific applications. But caution should be taken as true commercial systems will not always achieve the performance results of these benchmarks.

The tpc-C benchmark is the most commonly used but again is based on an idealized workload in which there are only OLTP transactions, with a steady state arrival rate and uniform, well-behaved I/O. As such, there are no bottlenecks imposed by "workload or data skew".

How, then, does one handle the consolidation of different types of work onto a single box? UNIX systems are generally designed where one application/database is dedicated to a particular server. In order to support workload growth the choices are:

- SMP growth
- Distributed databases with requests shipped between servers
- Clusters of servers

Accomplishing the growth in an SMP environment requires partitioning of the database and the application supporting it. We still need to consider that such environments still run most efficiently at 40-60% utilization. On an Enterprise Server 10000 with Dynamic Partitioning, these figures can probably increase substantially. On a S/390, processor utilization can peak at 90-95%. The primary difference between Sun Enterprise Server and S/390 is that while S/390 can achieve automatic workload and system balancing, the Sun system requires operator intervention to reconfigure the partitions (e.g., from daytime load requirements to night time batch support).

In our previous discussion of consolidation options, we indicated that the SMP environment offered significant savings over the decentralized distributed environment. The performance savings could be achieved as inter-system communication and function/request shipping could be accomplished locally rather than across the WAN.

The advantage of the S/390 environment is that workload growth can be accomplished without partitioning the database and application, though total scalability will be dependent upon the application architecture and database design.

S/390s are most effective at handling heterogeneous, I/O intensive workloads, with large working sets. UNIX servers however, are optimally suited for parallel, homogeneous workloads which are CPU intensive rather than I/O intensive and have small working sets. The Sun E10000 with dynamic partitioning enables multiple disparate workloads to be executed on the same platform by logically isolating the work into separate domains. Since it is still a UNIX based system, the workload management is best accomplished via this logical separation, as compared with S/390 which can automatically handle workload management and service delivery.

S/390 can handle such disparate workloads up to very high utilizations with acceptable response times, whereas UNIX systems start to give unacceptable response times at much lower utilizations. The result is that UNIX systems are very good at well-defined, processor-intensive single applications, whereas S/390 servers are better at running large, mixed workload systems.

The following chart taken from the S/390 homepage demonstrates typical utilizations of S/390 and UNIX systems. Note that the usable capacity multiplier is approximately a 3-fold increase for UNIX over S/390 based on how these systems run. The example they site is for typical SCSI disk devices rather than RAID technology. In typical UNIX implementations, RAID performance benefits are only achieved for applications with high Read/Write ratios.

Table 1.
S/390 Utilization vs. UNIX

- Usable Capacity

Utilization	**UNIX peak**	**S/390 Peak**	**UNIX Average**	**S390 Average**	**Usable Capacity Multiplier**
CPU utilization at sub-second response time	50-60%	100%	20-30%	65-75%	3X
Disk capacity utilization			20-30%	60-75%	3X

The authors of the S/390 homepage (Ref. http://www.s390.ibm.com/marketing/gf22509.html) substantiate these findings by quoting Mr. Steve MacKay, Chief Technical Officer of Sun Microsystems, as quoted in Investors Daily in March, 1999. In the interview Mr. MacKay indicated that " Peak performance is an area that Sun is working on, and one of the advantages of mainframe over UNIX is that mainframes are capable of running at 85-95% of capacity. UNIX servers run at 20%- 30% of peak load".

To further illustrate the disk I/O performance benefits of S/390 over UNIX, the article gives an example of disk requirements for a typical high-end 700 GB database. This example uses the same 2.8-3X usable capacity multiplier as in the previous table. As such, a server consolidation on S/390 can achieve greater cost savings for disk than in standard UNIX implementations.

Because of the I/O differential, or usable capacity differences , they site a 2TB requirement for S/390 as compared with a 2.8 TB requirement for UNIX. This is shown in the Table below. With 2 DB instances the costs and disk requirements are correspondingly doubled.

Table 2. Usable Capacity
Comparing Typical Disk Requirements
(700 GB database)

	Capacity required by UNIX at 25% Utilization	Capacity Required by S/390 at 70% Utilization	Usable Capacity Multiplier
1 DB Instance	2.8 TB	1 TB	2.8 X
2 DB Instances	5.6 TB	2 TB	2.8 X

The discussion of benefits of S/390 over UNIX, even for Sun E10000 class processors, indicates that multiple DB instances may be required to satisfy performance requirements of multiple disparate workloads such as DSS, Web serving and data mining applications.

The purpose of the S/390 discussion was to illustrate how different system architectures can ultimately impact server sizing and cost. Consolidating work on the Sun Enterprise E10000 Class machine will still achieve disk savings in terms of system files, libraries, work and sort space requirements. Perhaps the disk farm cannot run at the same utilization levels now as S/390, but improvements in technology and RAID implementations are forthcoming.

Performance Measurement Data to Collect for Sizing Consolidation Platform

In order to adequately determine server platform and capacity requirements, data must be gathered from each of the applications and servers which are candidates for consolidation. The data gathered must include application and system design requirements, as well as service level requirements to insure that the platform selected can satisfy all business needs. Graphs of current and average peak utilization statistics for each of the workloads can be used to gain an understanding of current system demands. In addition, current chargeback or cost allocation figures should be collected for use in the financial comparisons between alternative platforms.

- System and Application Design Information
  - Application dependencies and integration requirements
  - Ease of migration
  - Platform dependencies
  - RAS, security issues
- Performance Data
  - System and application: processor utilization, disk capacity and utilization, storage (memory, DB cache requirements)
  - Network utilization, bandwidth, protocols
  - Device characteristics: Processor Class, speed etc., Disk types and characteristics including use of RAID
  - DBMS statistics
    - Concurrent users
    - Query types
    - Size of raw database
    - Data/Log/Temp tables sizes and mapping to devices

    - Back-up requirements
- Application Service Level Requirements
    - Availability
    - Response times for workload types: OLTP, DSS, Queries and Batch turnaround time
- Workload Characteristics
    - Database
    - OLTP vs. DSS
    - Load statistics: average and peak
- Costing/Chargeback Algorithms
    - Personnel costs
    - Processor/Disk
    - Etc.

Benchmarking Results- Large Enterprise Servers

The following Chart is from SAP homepage and the Sun homepage and serves to illustrate the load levels that can be achieved on a single consolidated platform. It gives technical credence that these large servers can in fact achieve the requisite performance and service required. The example is based on SAP/R3 benchmark results and contrasts performance on S/390 and the Sun Enterprise servers.

Table 3.
SAP Benchmark Results

- S/390 vs. Sun Starfire

System	SD Users	# DB Instances	CPUs	% Utilization	Date	MIPS
IBM S/390 RY5	8,000 SD (3-way Sysplex)	30	332 MHz	97%	7/98	1380
Sun Starfire 336	14,400 SD	64	336 MHz	82%	8/98	2552

The Sun system could support 4,400 benchmark users with average dialog response times of .97 seconds per transaction and throughput of 4,725,000 dialog steps/hour. The IBM system was a 3-way Parallel Sysplex of IBM S/390 Model 9672 (RY5 G4 CMOS 10Way processors). The CPU utilization of the database server was measured at 97%, overall throughput was measured at 829,000 processes order line items per hour and over 2.8 million dialog steps per hour. Throughput would most likely increase significantly on the G5.

The SUN system can scale to 64 CPUs on a single system, whereas the S/390 can be used to cluster up to 32 parallel sysplexes each containing 10 processors. The difference is that more capacity is achievable on a single Sun Enterprise Server but more total capacity and throughput can be achieved in S/390 sysplex implementation.

The following chart contrasts performance in online transaction processing performance (OLTP) as measured by the TPC-C benchmarks. Again, these benchmarks can be used for comparative purposes to gain an understanding of how various hardware vendors scale to support large OLTP environments. These benchmarks as with others demonstrate the effectiveness of consolidating multiple servers onto large single system platforms. From the Sun homepage comparisons of tpmC performance can be found comparing the Sun E10000 with other multi-node clusters include Sequent, HP , and IBM SP Model. The following table is extracted from the Sun web site The performance for the IBM SP system was measured over a year earlier than the newer Sun and Sequent systems. (Ref: http://www.sun.com/servers/highend/10000)

Table 4

- TPmC Benchmark results

Source: Sun Home page

TPC-C System	TPmC	Database	Availability Date	Database Instances
SUN E10000 (Starfire 64-way)	115,395.73	Oracle 8I Enterprise Edition 8.1.5	8-22-1999	1
Sequent NUMA-Q 2000 (64-way)	93,9000.85	Oracle 8 Enterprise Edition 8.0.4	6-15-1999	16
IBM RISC System 6000 SP Model 309(c/s)	57,053.80	Oracle 8 Enterprise Edition 8.0.4	4-30-1998	8

Summary on Server Selection

In selecting a server, your number one priority should be to select the server that best meets your business requirements--now and into the future. One needs to take a long- term view in selecting a server platform. Ongoing operational costs and the flexibility to adapt to new requirements are key to the total cost and business value of that application over the years.

The general costing considerations that must be included in any evaluation of server consolidation and platform selection include:

- An identification of server costs and back-up/recovery requirements
  - Support requirements in distributed vs. centralized environments will be significantly reduced.
  - The amount of time needed for back-ups and cost of labor will be reduced in a centralized model and reliability of service will increase.
- Processor Utilization can be improved with centralized servers.
  - Many distributed servers are often under-utilized and there may be savings in combining multiple workloads.
  - It may also be feasible to combine workloads on a single system if their peak demand times for service do not coincide (e.g., back-ups and batch processing vs. OLTP work)
- Software licensing costs are typically reduced when the number of footprints are reduced.
- Network requirements and costs will increase as a result of centralization because users will be further away from the application.

In summary, one must select a platform which delivers the lowest total cost over the life of the application while addressing the following :

- Application service levels, reliability, availability, recoverability, and security requirements.
- Provides the best solution for the entire company
- Satisfies all of the application and system design requirements
- Satisfies all of the application data requirements
- Minimizes system management requirements and overhead
- More automated operation
- Ability to integrate new and different applications
- Ability to manage growth and scalability

We have already demonstrated that large, single system servers can meet the demands for large SAP, OLTP, and DSS environments. An examination of the various hardware vendor homepages demonstrate their abilities to satisfy these requirements and third-party benchmarking organizations such as TPC-C can substantiate these findings. The issue is therefore to determine what applications can be consolidated together, and to select a platform that satisfies business requirements while minimizing costs.

In order to satisfy today's business climate which includes rapid adaptability for data warehousing, DSS systems and e-commerce implementations, the issues of RAS are a critical consideration particularly in server consolidation. Therefore the platform choices can be restricted to include only those satisfying RAS, and that also minimize system management and operational impacts. Once the choices are narrowed down, the final issues are to ensure that all application system design, data and inter-system dependencies can be satisfied. Given multiple platforms satisfying all of the requisite criteria, the final decision can ultimately be made by minimizing total cost.

Summary

Distributed computing environments, which emerged in the 1980's as the panacea for computing problems, have not performed up to expectations. Consolidated computing environments offer higher availability, better response times, easier integration, improved security and more flexibility than current distributed environments. In addition, benefits can be realized from the centralization of staff, maintenance, licenses, and system management. Given the current equipment, products and integration options available, it is appropriate for organizations to reexamine their current computing implementations and evaluate the feasibility and opportunities of consolidation.

It is likely that the analysis will lead to a change in the direction of corporate computing strategies to a more centralized model than has been recently employed. Implementing this "new" paradigm will contribute to a control of IT resources that has been lost in the move to decentralization.

References

1. InfoWorld Magazine, "Its time to Consolidate", September 15, 1997
2. Gartner Group Study, K-LAN-308:Dave Cappucio: February 27, 1995
3. Consolidating UNIX Systems onto OS/390, SG24-2090
4. Selecting a Server, The Value of S/390, SG24-4812-01
5. Sun Homepage, http://www.sun.com/servers/highend/10000
6. IBM S/390 Homepage, . http://www.s390.ibm.com/marketing/gf22509.html
7. Brian L. Wong , Configuration and Capacity Planning for Solaris Servers