Cost and Scalability in Vertical and Horizontal Architectures

Implications for Database and Application Layer Deployments

Technical White Paper — Tom Atwood, September 2004

Table of Contents

Chapter 1

Executive Overview

Over the past several years, many have drawn attention to the increasing presence of small horizontally-scalable servers in data center deployments, occasionally going so far as to predict the demise of larger vertically-scalable systems. In fact, industry data shows that organizations continue to deploy a range of different server platforms. Database- and application- tier deployment in the modern data center include both vertical and horizontal architectures, along with diagonal architectures that combine the strengths of both to meet the needs of different applications. While horizontally-scalable architectures clearly have an essential role in the data center, vertically- and diagonally-scalable systems are often better suited to certain tasks.

When real-world considerations such as database software licensing costs, flexibility, and resource utilization are taken into account, vertically-scalable configurations are often cost-competitive on a per-processor basis. For large databases, vertically-scalable architectures can provide higher levels of scalability than large numbers of horizontally-scalable systems with distributed databases, and vertically-scalable servers are often easier and less expensive to manage. Consolidation, better resource management, and improved utilization can also make vertically-scalable architectures just as affordable to acquire as horizontally-scalable systems. As a result, organizations are free to choose solutions based on the merits of the architecture, the needs of the organization, and the particular requirements of the application.

This paper describes the characteristics of vertical, horizontal, and diagonal architectures along with system-level attributes that can affect application performance and availability. Cost analysis of several vertical, horizontal, and diagonal database and application layer scenarios examines the very real impact of system utilization on cost of acquisition. An cost-of-acquisition analysis is also provided for database layer systems, comparing vertically-scalable systems with Oracle 9i/10g against horizontally-scalable systems with the Oracle 9i/10g RAC distributed database option.

Sun understands that modern IT infrastructure must be tightly aligned with the needs of the business and that total cost of ownership (TCO) goes well beyond initial purchase cost. Rather than advising an adherence to one particular architecture, Sun provides a broad range of 32-bit and 64-bit servers that can be deployed in vertical, horizontal, or diagonal architectures. Sun's approach helps enable a focus on the individual business requirements, giving customers access to new and innovative technologies to protect their considerable investments in hardware, software, and people. Flexible upgrade options make Sun Fire™ servers a smart investment for the long term, as IT infrastructure evolves to match the needs of the business it serves.

Chapter 2

IT Infrastructure Drives Business Results

In most organizations today, information technology (IT) infrastructure must contribute strategically and directly to the needs of the business. IT departments can help provide a competitive advantage by deploying new applications that can lead to increased revenue and improved organizational efficiencies. However, like any other investment, IT infrastructure must be cost effective and show rapid return on investment (ROI). As a result, IT infrastructure choices are increasingly scrutinized to make sure they meet current and future application requirements as well as being economically feasible. In this environment, deploying the right architecture is more important than ever.

Choosing the Right Architecture for the Workload

Complexity in the modern data center drives up costs and reduces organizational agility and each deployment must be carefully analyzed to understand whether it increases or decreases complexity. Just as every organization is unique, system architectures must be carefully selected to fit the requirements of the application *and* the environment. Customer requirements such as service level agreements (SLAs) must be considered as well as the business- or mission-criticality of the application. Any new deployment must also be evaluated with an eye toward future growth, considering potential changes, upgrades, and updates, as well as ongoing availability requirements.

Once these considerations are in hand, organizations can begin to weigh the options available to them in terms of server architecture. Most hope to maximize performance, scalability, availability, and utilization, with the ultimate goal of maximizing competitiveness. At the same time they will strive to minimize complexity, management costs, licensing costs, and acquisition costs. In addition to getting architecture right the first time, solutions must be flexible to allow for change and adaptation over time in concert with changing business conditions and business priorities.

When evaluating system architecture, it is often useful to examine the cost associated with a given amount of computational work as well as evaluating how easily systems will be able to scale over time (in terms of both performance and cost). With vertically-scalable system architectures, capacity can be increased by adding (or upgrading) individual processors in larger symmetric multiprocessor (SMP) systems. Though initial vertical system costs may be higher, subsequent processor and memory additions augment the same chassis, providing greater value over time. With horizontally-scaled architectures, entire systems are usually added to the aggregate mix of available computational power, often with additional physical requirements (footprint, power, cooling, etc.) License costs and support costs must also be factored into any evaluation. For example, it hardly makes sense to pursue a small savings on system acquisition cost, only to pay more in management, licensing, and recurring support costs.

Once a general understanding of required numbers of processors is available for the desired task, it is useful to evaluate how well those processors will be utilized in the context of the proposed architecture. A lower utilization rate, for example, might indicate that many more processors (and possibly systems) are required than initially thought. It is also helpful to evaluate how well the system can adapt or scale (either up or down) to changing demands. How gracefully can the system take on additional projects or tasks? How well can excess capacity be utilized by other applications? In terms of investment protection, how well can the chosen architecture adapt to the availability of newer, faster processors and systems? Finally, and very importantly, how much disruption will be brought to bear on critical business services by future upgrades and conversions?

The N-Tier Data Center and System Architecture Trends

Most modern application services are deployed today using an "N-tier" data center architecture (illustrated in Figure 1). This model provides an intelligent way to organize distributed applications as well as a methodology for providing systems that are tuned to deliver appropriate levels of service for their role in the data center.

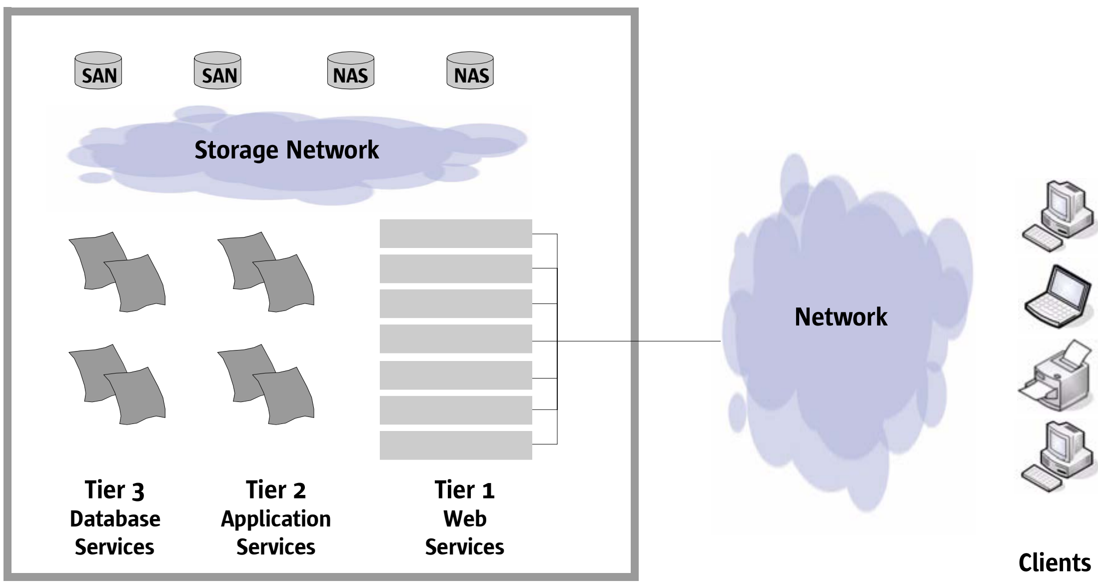


Figure 1. N-Tier architecture for the data center

Each tier has very different requirements that are best met by different computing architectures. Tier 1, often known as the presentation layer, is increasingly moving toward horizontal scaling with smaller servers providing most of the computational power. In contrast, the database layer (Tier 3) is typically dominated by larger vertically-scalable servers. The application layer (Tier 2) appears to be an area where both vertical and horizontal architectures are common. Diagonal architectures are also emerging at the application layer that leverage the advantages of both vertically- and horizontally-scalable designs.

It has been popular in the media to predict the obsolescence of vertical systems, asserting that medium-to-large servers are being rapidly replaced by racks and clusters of smaller servers. In reality, vertically-scalable systems offer advantages that are unmatched by other designs, and market data seems to indicate ongoing demand for vertically-scaled systems. Figure 2 shows the percentage of the total server market revenue comprised of servers priced above $500,000 according to Dataquest. These servers are all medium to large SMP servers that are capable of supporting large numbers of processors. As the graph shows, from 1998 to September 2003, the percentage of server revenue that can be attributed to servers in this price group has remained constant, despite economic and market fluctuations.

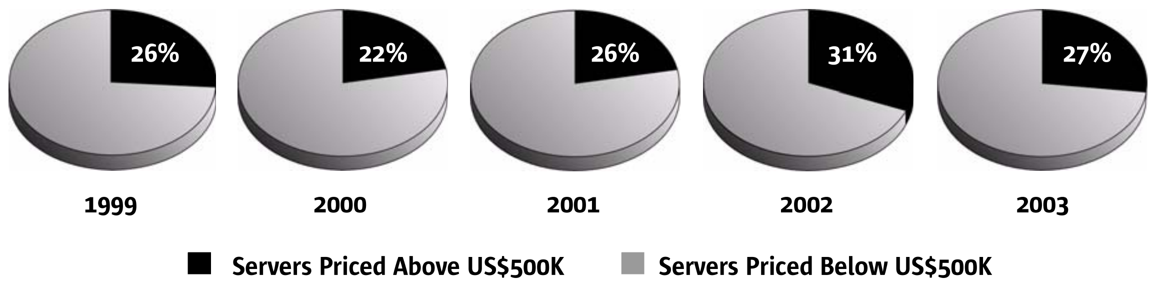


Figure 2. Percentage of servers priced above $500,000 (midrange to high-end servers), Source: Dataquest Q4 and Annual Worldwide Server Market Share Report, February 2003

It is also useful to try to understand how vertically- and horizontally-scaled architectures are being deployed in the data center. Data published in 2003 by IDC (Figure 3) provides an interesting view that breaks down data center tiers across a broad swath of systems ranging from one to 64 processors. Interestingly, this data shows that no one architecture dominates current deployments servers of all sizes are being deployed on all three tiers. In fact, organizations are deploying the application tier equally on systems that range from one to 64 processors. The percentage of each size of server dedicated to the application tier ranges from 35 percent for 1-processor servers up to 40 percent for 64-processor servers. IT departments seem to be deploying a variety of approaches, including horizontal, vertical, and diagonal architectures.

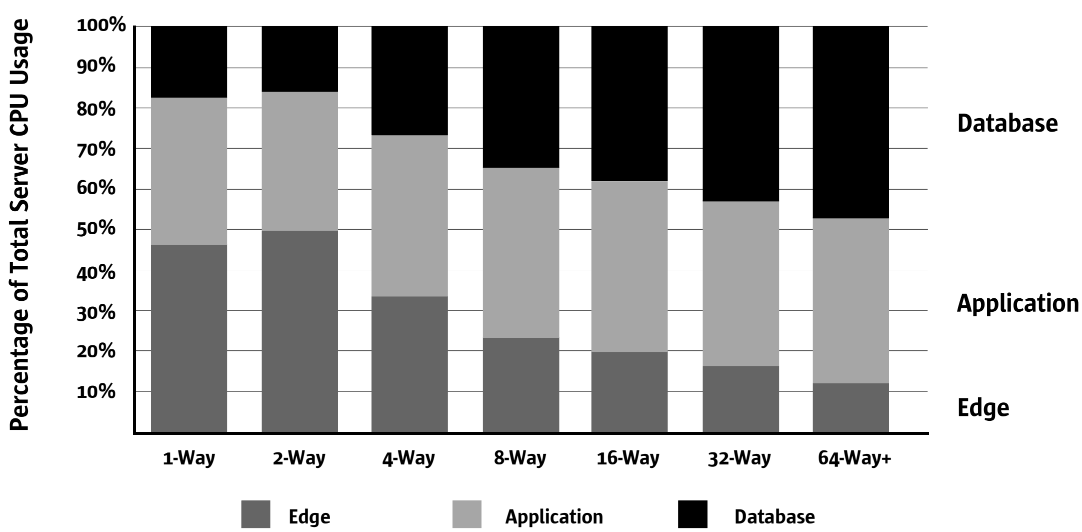


Figure 3. Servers of all sizes are being deployed in the data center (Source: IDC, 2003)

Sun Fire Servers and Solaris Operating System Technology

Rather than advocating a single architecture, Sun designs its technology to serve a broad range of deployment scenarios while providing considerable flexibility and investment protection. Beyond the merits of individual servers, Sun provides essential data center services, partnerships, and applications. In addition to these offerings, Sun Reference Architectures provide a flexible and informed approach that removes much of the risk from choosing deployment architectures and system components, giving organizations useful building blocks that help ensure rapid and successful deployments.

Most effective data centers feature a mix of horizontal, vertical, and diagonal architectures, depending on when applications were deployed and how they have evolved. Sun's binary-compatible SPARC® server product line spans the data center — letting organizations scale up, scale out, or both. Figure 4 illustrates the upper portion of Sun's server product line, including both UltraSPARC® III and UltraSPARC® IV based servers. For brevity, Sun's smaller horizontally-scalable SPARC and x86 servers are not pictured.

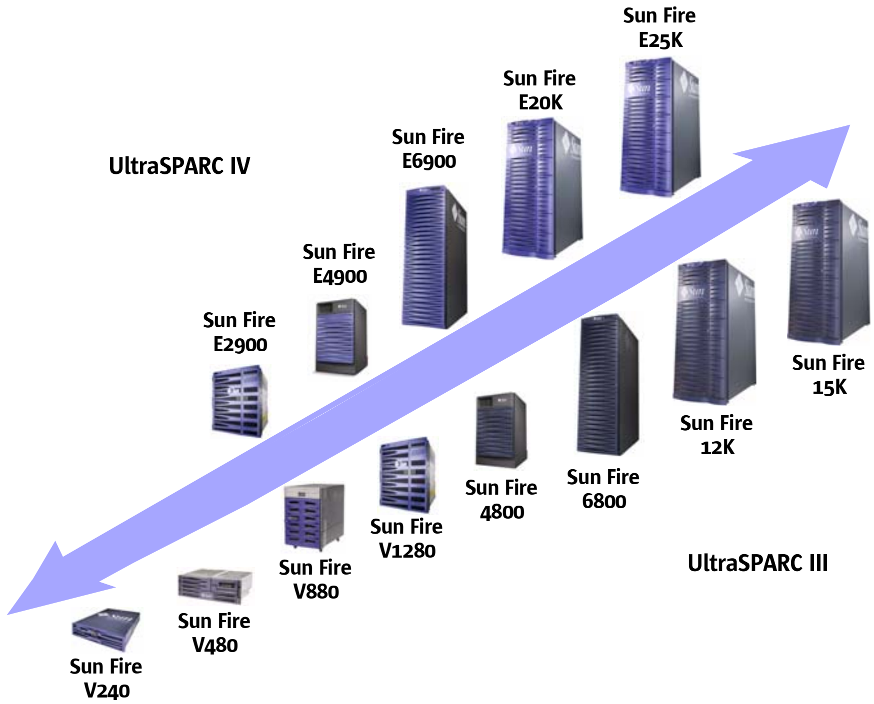


Figure 4. Sun's broad product line of binary-compatible systems for application- and database-tier deployment

The combination of powerful Sun Fire servers with the robust Solaris™ Operating System (OS) lets organizations deploy capacity where it will do the most good for the business. Table 1 lists the capabilities of servers that are most appropriate for application- and database-tier deployment. Chip multithreaded UltraSPARC IV processors available in Sun Fire Enterprise servers can process two threads simultaneously. Also, several Sun Fire Enterprise servers support multiple Dynamic System Domains (Solaris OS instances) per server as is discussed later.

Sun Fire Server	Maximum Number and Type of Processors	Maximum Concurrent processor threads	Memory Capacity	Dynamic System Domains (supported Solaris OS instances)
Sun Fire™ V480 server	4 UltraSPARC III	4	Up to 32 GB	1
Sun Fire™ V880 server	8 UltraSPARC III	8	Up to 64 GB	1
Sun Fire™ V1280 server	Up to 12 UltraSPARC III	12	Up to 96 GB	1
Sun Fire™ E2900 server	Up to 12 UltraSPARC IV	24	Up to 96 GB	1
Sun Fire™ 4800 server	Up to 12 UltraSPARC III	12	Up to 96 GB	1-2
Sun Fire™ E4900 server	Up to 12 UltraSPARC IV	24	Up to 96 GB	1-2
Sun Fire™ 6800 server	Up to 24 UltraSPARC III	24	Up to 192 GB	1-4
Sun Fire™ E6900 server	Up to 24 UltraSPARC IV	48	Up to 192 GB	1-4
Sun Fire™ 12K/15K servers	Up to 52/106 UltraSPARC III	52/106	Up to 288/576 GB	1-9/1-18
Sun Fire™ E20K/E25K servers	Up to 36/72 UltraSPARC IV	72/144	Up to 288/576 GB	1-8/1-18

Table 1. Capacities for several Sun Fire servers, appropriate for application- and database-tier deployments

Chapter 3

Vertical and Horizontal Architectures

Though all computing systems are comprised of the same essential components, vertical and horizontal architectures combine, connect, and utilize those components in different ways. These divergent design approaches result in distinct environments that accommodate applications differently. Understanding these key differences can be useful in selecting an effective approach for a particular deployment.

Architectural Characteristics of Vertical and Horizontal Architectures

Figure 5 contrasts the different processor, memory, and I/O designs of vertical and horizontal architectures.

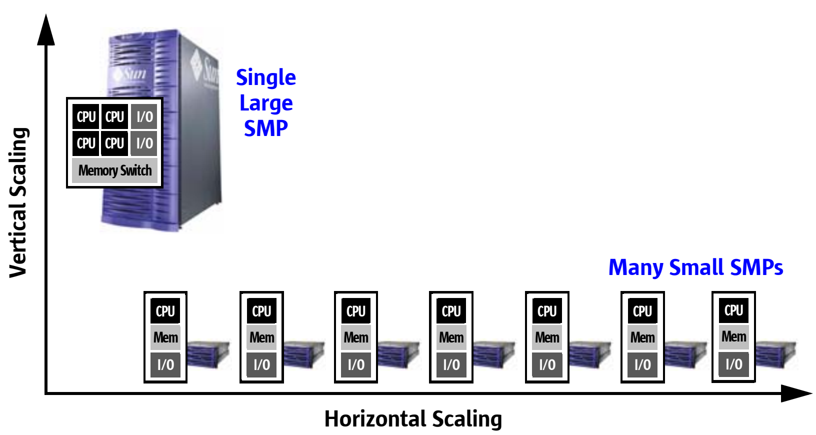


Figure 5. Contrasting designs approaches of vertically- and horizontally-scalable servers

Vertically-Scalable Architecture

Vertically-scalable servers are usually represented by larger SMP systems (typically hosting eight or more processors). These systems generally use a single instance of the operating system to manage multiple processors, memory subsystems, and I/O components, usually contained within a single chassis, rack, or box. Servers such as Sun Fire Enterprise servers can also be divided into separate Dynamic System Domains (domains) where multiple instances of the operating system manage different subsets of the resources in a particular chassis.

In a vertical design, the system interconnect is commonly implemented as a tightly-coupled centerplane or backplane that provides both low-latency and high-bandwidth (such as the Sun™ Fireplane interconnect in

Sun Fire™ Enterprise servers). In vertical or SMP systems memory is shared and appears to the user as a singular entity. All processors and all I/O connections have equal access to all memory. The cache coherent interconnect maintains information on the location of all data regardless of its cache or memory location. Resources are added to the chassis by inserting system boards with additional processors, memory, and I/O subassemblies. In addition to a single large SMP server, vertical architectures also include clusters of large SMP servers that can be used for a single large application.

Sun sells a wide range of vertically-scalable systems, including the Sun Fire Enterprise servers. For the purposes of this discussion, vertically-scalable servers are assumed to support eight or more processors in a single chassis. All of Sun's vertically-scalable servers support 64-bit operation.

Horizontally-Scalable Architecture

In contrast to vertically-scalable architecture, horizontal architecture provides scalability through network or cluster connectivity between multiple small systems or nodes (usually one to four processors). Horizontally-scalable systems are typically somewhat loosely-coupled, and are connected by standard network interconnects such as Fast Ethernet, Gigabit Ethernet (GBE), and Scalable Coherent Interconnect (SCI). These interconnect technologies provide much lower bandwidth and higher latency than the integrated vertical system interconnects.

In horizontally-scalable architecture, each node has its own processor(s) and memory. Depending on the system design, I/O subsystems may be internal to the node or shared with another system or chassis. Each horizontally-scalable node has a single instance of the operating system and resources are typically increased by adding more nodes rather than by adding more resources within a node. The memory in horizontal systems is distributed, and each node has its own memory that is directly accessed only by the node's processor and I/O connections. Any access to these resources by other nodes takes place over the slower, more loosely-coupled network interconnects and would be much slower than access by the node itself. There is also usually no cache coherency between nodes in a horizontal architecture since appropriate applications are generally small enough to fit within a single node. If more than one node is needed for the application, cache coherency must be handled by the application.

Sun sells a range of horizontal systems including blade servers and rack-mountable servers using both x86 and SPARC processor architectures. While many horizontally-scalable systems are limited to only 32-bit operation, Sun's product offerings support both 32- and 64-bit operation. Horizontally-scalable systems are usually deployed in clusters of small nodes, with each node supporting one to four processors.

Attributes of Different Architectures

Vertical and horizontal architectures coexist in the data center precisely because they complement each other and facilitate different kinds of applications. While horizontally-scalable systems may have a lower per-processor acquisition cost, they are not ideal for all situations. As will be discussed in later chapters, vertically-scalable systems can be competitive and even less expensive to acquire when issues such as utilization and licensing costs are taken into account.

Table 2 compares and contrasts the attributes of vertically- and horizontally-scaled systems along with ideal types of applications for each. It is important to note that in many cases it is not the type, but the size of the application that should be considered when selecting an architecture.

Attribute	Vertical Systems	Horizontal Systems
Number of processors	Eight or more	One to four processors
Processor architecture	64-bit	32-bit and 64-bit
Memory design	Shared and cache-coherent	Separate, per-node memory resources
Availability strategy	Redundant components, sophisticated reliability, availability, and serviceability (RAS) features	Replication of nodes
Components and Interconnects	Commodity components with high-speed proprietary system interconnects	Commodity components with slower standard system and network interconnects
Expansion/Upgrade opportunities	Add resources in existing chassis (CPU/memory boards, more and faster processors, more RAM, more and faster I/O connections)	Add more nodes or replace older nodes with newer, faster nodes
Ideal threads	Large execution threads, heavy inter-thread communication, threads with shared memory and/or data	Small, autonomous execution threads
Ideal applications	Large databases, transactional databases, data warehouses, data mining, application servers, non-partitionable HPTC applications	Web servers, firewalls, proxy servers, media streaming, directories, XML processing, JSP applications, SSL encryption/decryption, VPN, application servers, partitionable HPTC applications

Table 2. Attributes of vertically- and horizontally-scaled architectures.

Of course, there are no hard and fast lines between vertically- and horizontally-scalable systems and architecture, and many intermediate solutions can be built that combine the best aspects of both approaches. Diagonally-scalable architectures leverage the size, availability, and manageability features of medium-sized SMP systems while providing a degree of replication (Figure 6).

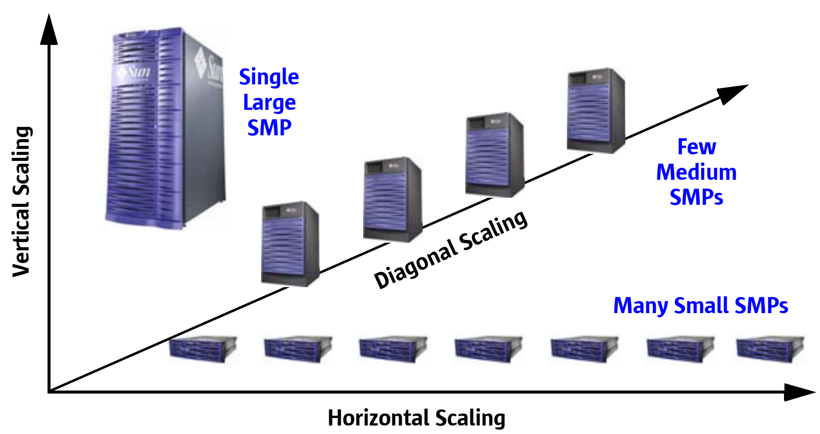


Figure 6. Binary-compatible Sun Fire servers can function in vertically-, horizontally-, and diagonally-scaled architectures

Applications that are stateless, small, and easily replicated are ideal for small horizontally-scaled servers such as blade servers. Applications that are stateful or that require significant amounts of data, users, and large-scale internal data communication are ideal for vertically- or diagonally-scaled deployments. For example, many HPTC (High Performance Technical Computing) applications require large amounts of shared memory and utilize large numbers of interdependent threads that communicate frequently with each other. These applications are ideal for large SMP servers, and even clusters of SMP servers. Other HPTC applications feature execution threads that are not dependent on each other and do not need large shared memory. These applications are partitionable and are ideal for clusters of small servers.

The database tier commonly has dependent execution threads that require large amounts of processor, memory and I/O resources, making vertical servers well-suited. The application tier typically features many small- to medium-sized application instances, allowing both vertical and horizontally-scaled architectures. Some commercial applications are partitionable and can be deployed on horizontally-scaled servers, while others are not easily partitionable and are better suited for vertically-scaled deployments. Diagonally-scaled architecture may be applicable to larger applications with more demanding memory and processor requirements. Figure 7 provides a graphical representation of these different types of workloads. Many applications can be deployed on all of these architectures, depending on the size of the applications, the overall workload, and existing infrastructure.

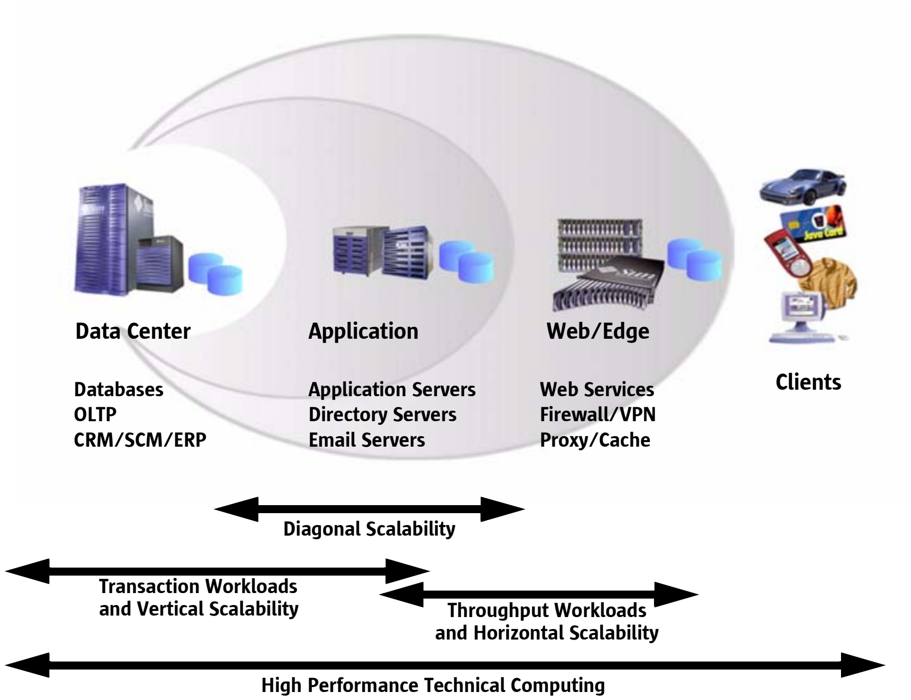


Figure 7. Appropriate applications for vertically- and horizontally-scalable architectures

Performance Implications of Architecture

Vertically-scalable servers and clusters of small computation nodes ultimately represent different approaches to parallel computing. To achieve good application performance, a balanced system is needed with fast processors, a fast interconnect, fast I/O, a scalable operating system, optimized applications, and a high degree of reliability, availability, and scalability.

- *Processors and System Interconnect*
  It is common — though somewhat misleading — to rate systems by their processor frequencies. Processor speeds are important, but they are only part of the equation. For application performance, it is generally more important that processors run at their maximum capacity (with high rates of utilization) than it is to have the

fastest available processor. A faster processor running at 50 percent of its capacity may actually be slower in terms of delivered performance than a slower processor running at 80 percent capacity.

Additionally, as the number of processors in a system increases, the bandwidth and latency of the system interconnect becomes more significant than the speeds of individual processors. The system interconnect moves data from disk, memory, and network interfaces to the processors. The system interconnect is also used to move data addresses, a task that is critical in maintaining cache coherency. If the system interconnect has slow address bandwidth then the processor will often become idle awaiting data.

The key technical difference between horizontal and vertical systems is the bandwidth and latency of the respective interconnects. In a clustered environment, network connections such as Fast Ethernet or Gigabit Ethernet move data between nodes while system interconnects move data within individual systems. If any of these interconnects are too slow, the processors will sit idle awaiting data. Cluster interconnects range in bandwidth from 125MB/second for Gigabit Ethernet (GBE) up to 200MB/second for scalable coherent interconnect (SCI). Latencies range from 100,000ns for GBE to 10,000ns for SCI. Infiniband promises a faster interconnect with peaks ranging from about 250MB/sec in early implementations up to a possible peak of 3GB/sec with end-to-end latencies depending on packet size. In contrast, SMP interconnect speeds are much faster, ranging from 9.6GB/sec for smaller Sun Fire servers up to an aggregate peak bandwidth of 172GB/sec for the Sun Fireplane interconnect featured in Sun Fire E20K/E25K servers. Latency for Sun Fire servers ranges from about 200ns to 450ns.

- *Input and Output*
  For many applications, fast I/O is essential in order to get data from the disks and the network to the system interconnect and on to the processors. An I/O bottleneck can adversely affect even the fastest interconnect and processors. Vertical systems such as Sun Fire Enterprise servers have high-speed, high-bandwidth I/O subsystems that interface directly with fast system interconnects. Smaller horizontal systems typically use slower standard interconnects to communicate with I/O modules.

- *Operating System*
  The best hardware can be ineffective if the operating system does not scale to accommodate more capacity. The Solaris Operating System has been used to operate large SMP systems since 1993 and has been optimized for increased performance and scalability ever since. Operating-system scalability is less of a requirement for horizontally-scalable systems because individual nodes or individual instances of the operating system are usually comprised of four processors or less.

- *System Availability*
  The chosen architecture also has important implications for system availability. In large vertically-scalable SMP systems, reliability and high-availability features are built into the system and can be augmented with failover solutions. In horizontal systems the nodes themselves may not be as reliable but high-availability features can be achieved through replication of nodes.

- *Optimized Applications*
  To get the best performance, applications must be optimized for the target computing architecture. Most commercial applications are optimized for single vertically-scalable systems because software developers have developed and optimized on dominant SMP systems for many years. Horizontally-scalable architectures typically require the creation or adoption of a distributed computing framework in order to realize an optimized result.

Application Size

As mentioned, large SMP systems typically have very fast interconnects that yield balanced system performance. In typical horizontal systems, each node conducts the small task assigned to it independent of what the other nodes are doing. Due to their low bandwidth and high latency interconnects, horizontal systems may have performance issues if data must move frequently from node to node. At the same time, certain applications do not require high-speed interconnects to deliver good performance. These applications are typically small and can be easily replicated, including Web servers, proxy servers, firewalls, and small application servers.

Figure 8 illustrates a horizontal (or distributed memory) architecture. In this example, four-processor nodes (each with their own RAM and direct-attached I/O) are connected by a network interconnect such as Gigabit Ethernet. In the illustration, three sizes of workloads are mapped onto this computing environment. The smallest workload can fit easily onto individual nodes. As the workload gets larger it becomes too large to be handled by a single node, so multiple nodes are applied to each workload. When more than one node is used, inter-node communications over the relatively slow network interconnect can greatly hinder performance. Small workloads that do not require communication with each other are well suited for horizontal architectures, but large workloads may become problematic.

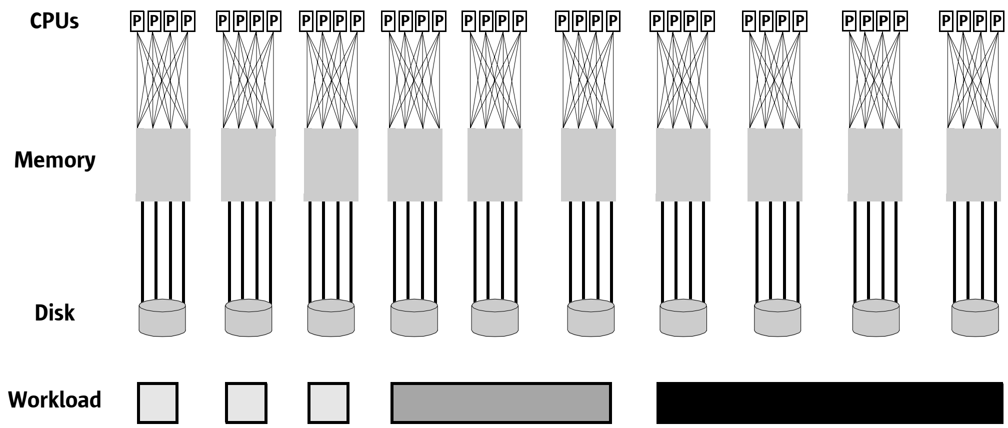


Figure 8. Larger workloads may be adversely affected by slower inter-node communication present in horizontally-scalable architectures

Figure 9 shows a large SMP system a large number processors, shared memory, and a high-speed interconnect (such as the Sun Fire E25K server). This server can handle all three sizes of workloads illustrated in Figure 8, since inter-node or inter-process communication is no longer an issue. In addition, all of the processors in the system can access all of the disks, memory, and network connections, a key characteristic of SMP systems.

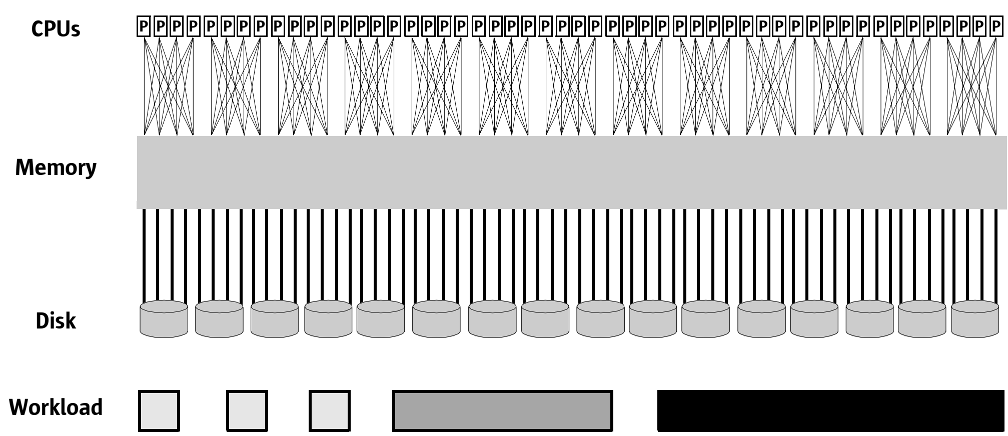


Figure 9. Vertical architecture allows multiple application sizes to coexist in a high-speed environment

Even though it is possible to deploy smaller applications and workloads on vertically-scalable systems, it may not make good financial sense. Per-processor acquisition cost for a large SMP server is typically greater than that of smaller horizontally-scalable systems. If an application can run with good performance on a small node or a few small nodes without undue increase in management complexity, then horizontal scaling is a more economical way to deploy. However, if the application is too large to run on a small node (or a few small nodes), a large SMP server is a more attractive option both from a systems management and a performance perspective.

Chapter 4

Application Layer Considerations

In modern data centers, many systems are alternately over-taxed, or vastly under-utilized. Resource utilization, and particularly processor and system utilization can have a significant effect on the economics of data center architecture and operations. Rather than looking only at the application tier, this section examines the overall application architecture, evaluating ways that vertically- and horizontally-scalable systems can be combined in order to improve overall utilization and performance.

Application Tier Resource Utilization

Among other factors, acquisition cost can be greatly reduced by improving system utilization rates. By improving the utilization of individual processors and systems, fewer of both may be needed to accommodate a given level of performance. Application-tier topology and resource management tools can each bring about substantial improvements in utilization that can increase performance while reducing costs.

Application Tier Topology

As recently as a few years ago, it was common to dedicate an individual server for a single application or database instance. Though this practice kept design and operational issues separate, it lead to large inefficiencies since each server was configured for peak load. The result was data centers that were littered with servers — most of them idle or running at very low rates of utilization during off-peak operation.

Making matters worse, many modern applications such as PeopleSoft, SAP and Oracle Applications contain components that run on different tiers of the data center. Each module, such as PeopleSoft's Human Resources module, will have components that run on the presentation layer (tier 1), the application layer (tier 2), and the database layer (tier 3). One way to deploy these modules would be to set up separate database servers, application servers and Web servers for each application module with no sharing between modules (Figure 10).

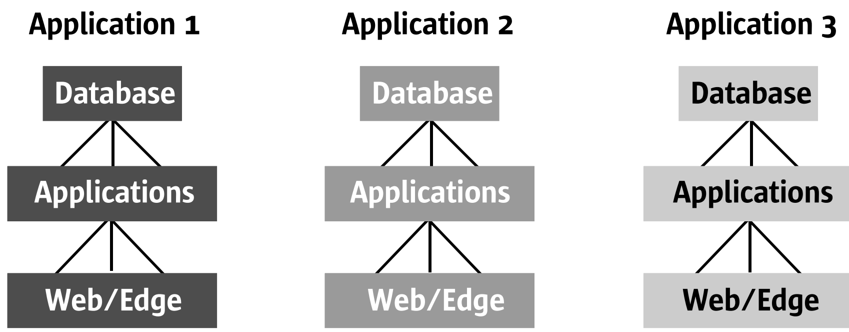


Figure 10. Application components deployed on completely separate servers

It is easy to see how this approach can quickly lead to a proliferation of servers, and rapidly-accelerating complexity and management costs. Today most organizations are looking for ways to consolidate the myriad servers in their data centers, taking better advantage of the computational resources they already own while making sure that new purchases are utilized to the fullest extent possible. This consolidation trend is focused on finding ways to safely and reliably configure more than one application instance on each server.

One efficient way to better utilize resources is to mix and match application tier instances, running more than one application tier instance on a given server. If the application instances are properly matched to server capacities, individual server resource utilization can be greatly increased (while the number of servers is greatly reduced). Figure 11 shows consolidation of the applications by tier. With larger servers, this approach increases the likelihood of being able to run multiple instances of a given tier on a particular server.

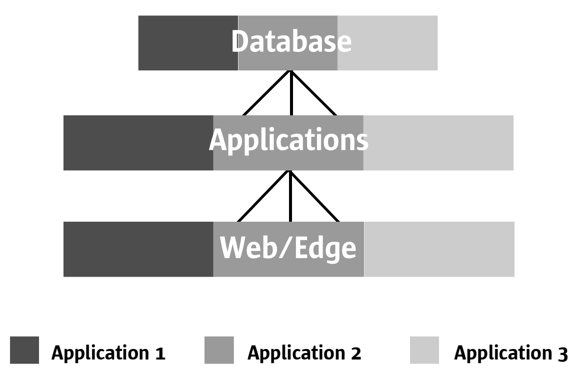


Figure 11. Consolidated deployment with application components deployed on similarly-purposed servers for better utilization and simplified management

Resource Management Tools

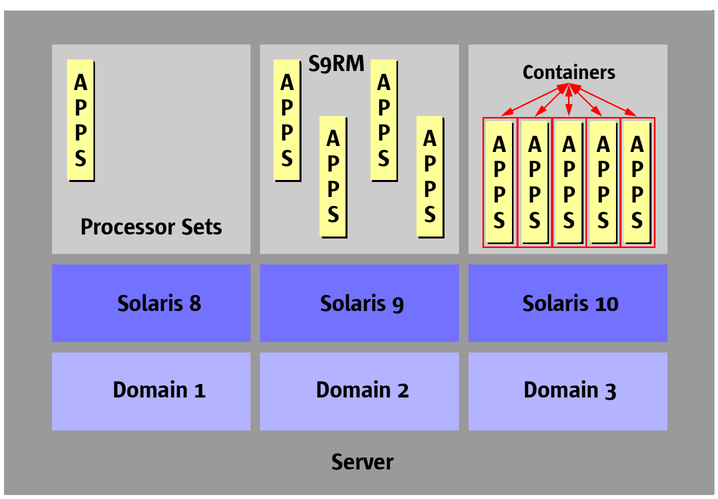
The ability to effectively run multiple application tier instances (and multiple database tier instances as well) on a single server requires that sophisticated resource management capabilities are built into the operating system environment. It clearly makes little sense to consolidate application components only to have them compromised when other application components demand additional resources.

In the Solaris Operating System, integral tools such as processor sets, Solaris™ 9 Resource Manager (S9RM) and, N1™ Grid Containers (Solaris 10 Operating System and Software Express) can improve resource utilization while helping to ensure that application components get the resources they need. Processor sets and S9RM act to partition processor resources within a single instance (or domain) of the Solaris Operating System. These tools help ensure that a given application is allocated at least the minimum resources necessary to meet performance requirements, even if it is sharing a server or domain with other application instances.

N1 Grid Containers improve greatly on S9RM and processor sets, adding security isolation and fault isolation between containers. Each container has its own IP address, its own file system, its own access privileges and it can reboot independently from other containers. Multiple containers can utilize the resources of a single processor, allowing very fine granularity and use on systems with a small number of processors. All of this flexibility and isolation requires only one copy of the Solaris Operating System so that management is simplified as well. With this approach, updating the operating system for a server or domain automatically updates all of the containers in that domain.

In addition to N1 Grid Containers, Sun Fire 4800/E4900, 6800/E6900, 12K/E20K, and 15K/E25K servers all support Sun's fifth-generation Dynamic System Domains and fifth-generation Dynamic Reconfiguration. Together, these powerful tools allow administrators to divide larger vertically-scaled servers into multiple fully fault-isolated systems, each running a different instance (and potentially a different version) of the Solaris Operating system. Dynamic Reconfiguration allows resources to be added to domains, or moved between domains without requiring downtime.

By combining Dynamic System Domains with processor sets, Solaris 9 Resource Manager, and new N1 Grid Containers, these systems can deliver considerable autonomy for applications while greatly enhancing the utilization of both individual processors and the system as a whole. Figure 12 illustrates the ability of a single Sun Fire Enterprise server to host many diverse domains and support many highly-specialized applications.



S9RM = Solaris 9 Resource Manager

Figure 12. A single Sun Fire Enterprise server can host multiple domains with different versions of the Solaris Operating System, and offers a variety of fine-granularity resource management technologies

Login Persistence and Load Balancing

In many typical application-tier deployments, application components are deployed across more than one server. In these situations, the ability to load balance many user sessions across the multiple servers that provide the application can be critical to good performance and high overall levels of utilization. While many applications now feature Java™ technology based middleware that enables effective dynamic load balancing, other applications provide no load balancing or only a rudimentary static load balancing system.

In the case of SAP, PeopleSoft, and the Oracle forms-based modules of the Oracle Applications suite, users are automatically sent to the least-busy server when they first login. For these and similar applications, if there are 10 servers in the application tier, the first 10 users will probably be assigned their own server (Figure 13).

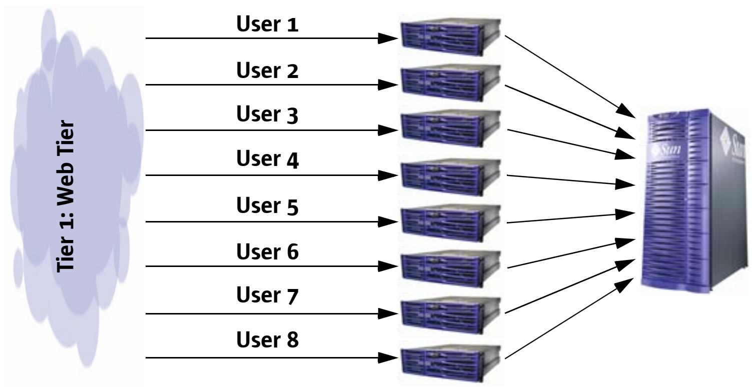


Figure 13. With statically load-balanced applications, users connect first to the least-busy server

In this scenario, the next 10 users will then be assigned to the least busy servers. Ultimately, a different number of users will end up logged in to each server with some servers busier than others (Figure 14). Once a user is logged in to a physical server they generally stay connected to that server and cannot dynamically move from a busy server to a less-busy server without logging out and logging back in again.

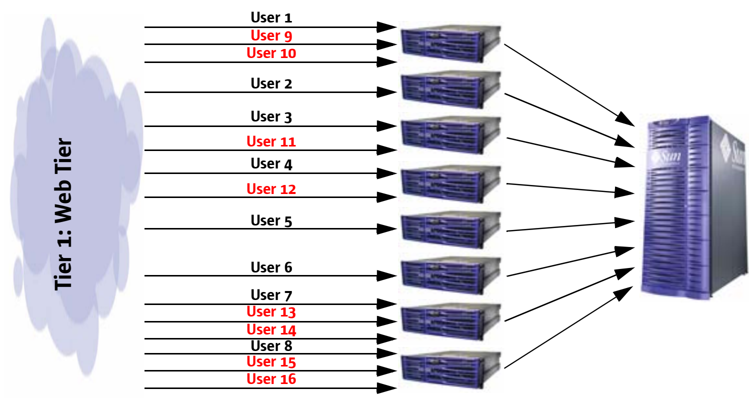


Figure 14. Users remain connected until logout leading to unbalanced load across the application servers

As a result of this static load balancing, application servers can easily become overloaded, creating so-called “hotspots” that exhibit very poor application performance. At the same time, other servers will likely be idle or under-utilized (Figure 15). One approach to avoiding these server hotspots is to configure each of the many servers with sufficient resources to handle any possible peak load and the highest number of application users. Alternately enough servers must be provided so that no one server ever has too many users logged in. Both of these approaches can lead to inefficiencies as the need to deploy larger numbers of servers can produce significantly under-utilized resources.

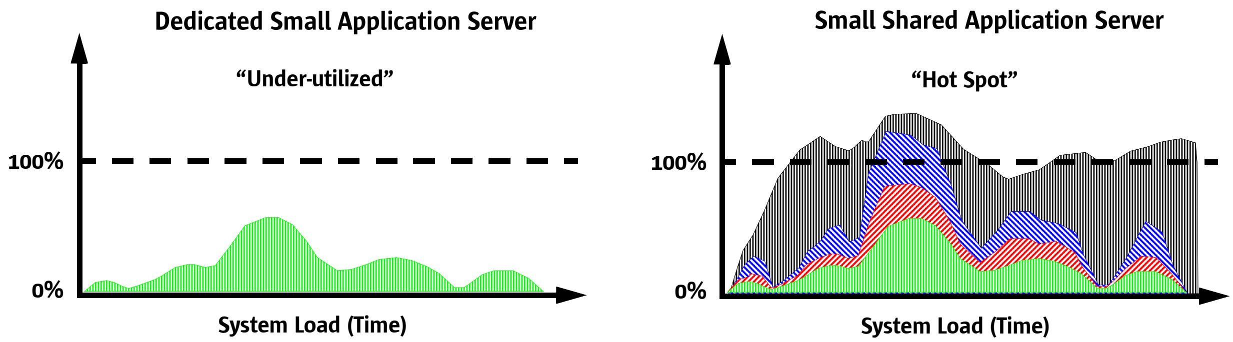


Figure 15. In many statically load-balanced application deployments, smaller servers are alternately either overloaded, creating "hotspots" or under-utilized, wasting valuable resources

Deploying a large number of small servers exacerbates this problem. The unused capacity (or "headroom") on smaller servers is often either insufficient, or wasted and unavailable for use by other applications. Unfortunately, headroom generally cannot be shared between servers in the application tier so excess headroom on one server cannot be utilized by servers that have exhausted their resources. The larger the number of servers that are deployed, the larger the amount of resources that cannot be shared and may be wasted. The smaller the server, the larger the quantity required and the fewer users or the fewer application instances that can be supported on each server.

One way to substantially improve load balancing in these situations and overcome some of the issues of login persistence is to use fewer large servers for the application tier and let the Solaris Operating System perform load balancing. This technique assigns multiple application instances to a single instance of the Solaris Operating System running on a multiprocessor server. Consolidating many smaller servers into fewer larger servers combines headroom and allows it to be shared throughout the application tier, improving system utilization and performance, and requiring deployment of fewer total processors.

The Solaris Operating System is very efficient at scheduling large numbers of processes among all the processors that exist in a given server or domain, and processes dynamically migrate from one processor to the next based on workload. With this approach, a large vertically-scaled server essentially functions as a pool of compute resources that are assigned, as needed, to the many users and application instances that reside on that server (Figure 16). Using the Solaris Operating System to load balance can reduce the need for processing, resulting in fewer processors, smaller memory, and lower acquisition costs.

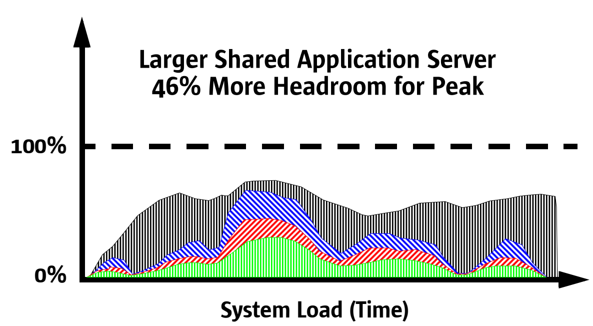


Figure 16. The Solaris Operating System can perform load balancing functions across multiple application instances on a large server to enhance performance, share resources, and improve utilization

To illustrate this point, a paper by Dr. Christian Tolkes of Sun's SAP Competency Center describes in more detail the possible reduction in hardware when using larger servers to consolidate resources. Dr. Tolkes' paper demonstrates how it was possible to deploy SAP with 50 percent fewer processors using a Sun Fire 6800 server rather than smaller two-way servers[1].

64-bit or 32-bit Servers and Number of Processors

When looking at options for lowering server acquisition costs, many organizations consider small 32-bit platforms because they seem inexpensive. However, many application-tier instances can have very high per-user processor and memory requirements that strain smaller 32-bit servers. Oracle Applications, for example, has a per-user memory requirement that ranges from 5MB up to 80MB with an average of 20MB. While memory requirements can vary from one application to the next, SAP, PeopleSoft, and Oracle Applications are three common examples with significant memory requirements.

As a result of these large memory demands, 64-bit systems are strongly recommended for the application tier, regardless of the resulting deployment architecture. 64-bit systems can support significantly more memory per processor and more memory per application process than 32-bit systems. In fact, the Sun SAP Competency Center and the Sun Oracle Advanced Technology Center (SOATC) advise customers to use 64-bit servers with at least 4 processors based on the nature of those applications. Servers with fewer processors can easily become a hotspot. For example, a single batch job for SAP can easily overwhelm a two-processor server. Many other applications behave similarly to SAP, PeopleSoft, and Oracle and that these recommendations are likely to be appropriate for other application-tier deployments.

Application Tier Architectures

In evaluating IT architectures, it is useful to compare different vertical and horizontal approaches to help understand their performance, availability, and cost implications. This section examines the merits of four different architectures that may be used in application-tier deployments against the technical criteria previously discussed. A subsequent section analyzes these solutions from a financial perspective. Table 3 summarizes the characteristics of the architectures to be discussed in this section.

Architecture	Performance	Availability	Complexity
Vertical 2-tier	Very Good	Marginal	Low
Vertical 3-tier	Excellent	Good	Low
Horizontal 3-tier	Good	Excellent	High
Diagonal 3-tier	Good	Very Good	Medium

Table 3. Multiple N-tier application architectures are possible, with different advantages and disadvantages

Vertical 2-Tier Architecture

Constructing a vertical 2-tier architecture involves deploying both the application instances and the database instances on a single system or on a single domain of a large, vertically-scalable server (Figure 17). This solution promises to have the fewest operating system instances and hence the most centralized management, leading

1. For more info please see: Consolidated SAP Solutions, A Review of Costs (reference 5), http://www.sun.com/datacenter/pdf/SAP_Consolidation.pdf

ideally to lower management costs. In general, the more operating system instances that need to be managed the higher the management costs are likely to be.

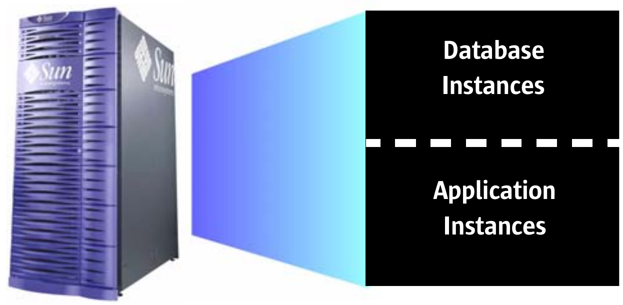


Figure 17. Vertical 2-tier architecture: database and application instances in one vertical server or domain

This vertical 2-tier architecture generally exhibits very good performance since the application instances and the database instances are able to communicate with each other internally to the server instead of over an external network interconnect. Some industry-standard benchmarks (such as the SAP 2-Tier benchmark) are run using this architecture. Resource utilization for this architecture is also very good since the Solaris Operating System can balance resources efficiently between the various application and database instances. Resource management tools such as processor sets, Solaris 9 Resource Manager, and N1 Grid Containers can be used to run all of the application and database instances effectively in a single domain.

Unfortunately, the availability characteristics of the vertical 2-tier architecture are not ideal. If the single large server or large domain goes down, then all application instances and all database instances go down as well. To improve availability, clustering can be used to provide failover capabilities, but that requires twice as much hardware and can increase complexity. As a result, while the vertical 2-tier architecture can provide some advantages in terms of management, resource utilization, and performance, its lack of high availability makes this architecture undesirable in production environments.

Note that the software fault isolation capabilities of N1 Grid Containers may make vertical 2-tier architectures somewhat more viable as the availability characteristics will improve.

Vertical 3-Tier Architecture

Similar to the vertical 2-tier option, the vertical 3-tier architecture places all of the application and database instances on a single vertically-scalable server (Figure 18). However, in this option the application tier instances are separated from the database instances by placing them in different Dynamic System Domains. Because domains provide resource, fault, service, and security isolation, they help enable the deployment of different application environments on a single physical system with much better availability characteristics.

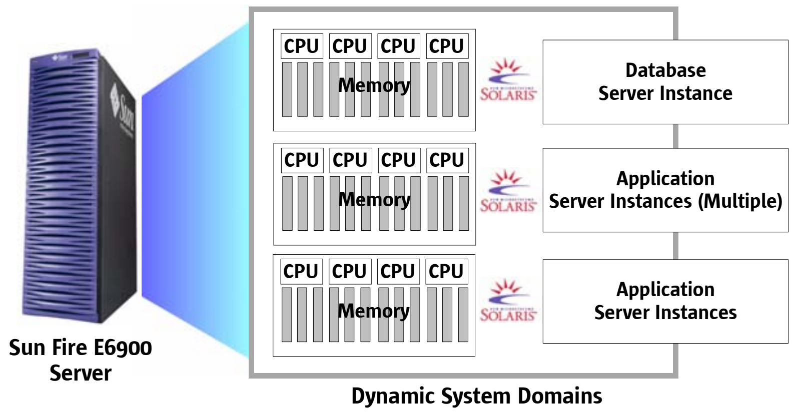


Figure 18. Vertical 3-tier architecture: application and database instances are placed in separate fault-isolated domains on a single server

By deploying the application instances in different domains, and separate from the database instances, the vertical 3-tier architecture overcomes the availability drawbacks inherent in the vertical 2-tier architecture. At the same time this configuration realizes the performance and resource management advantages of deploying many application instances on a large Solaris Operating System instance. From a technical perspective, the vertical 3-tier architecture is a very strong solution.

Horizontal 3-tier Architecture

The horizontal 3-tier architecture involves deployment of a large number of application instances on many relatively small servers (Figure 19). These small servers are normally from one to four processors in size and (as mentioned previously) and it is best if 64-bit servers are employed to meet the stringent memory demands of modern applications.

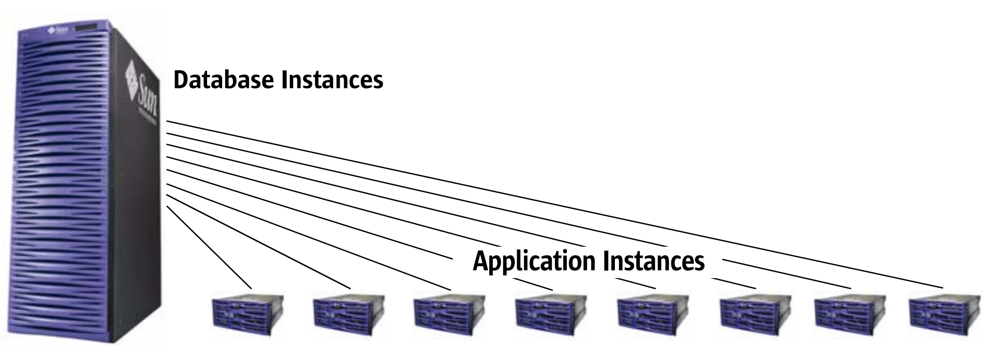


Figure 19. Horizontal 3-tier architecture: database instances on a large vertical server, application instances on many small (one- to four-processor) servers

With a large number of deployed servers, the horizontal 3-tier architecture can provide very high availability. While each node may not have high-availability features, the loss of one node is not usually detrimental to the overall production environment because of the inherent redundancy. Because the nodes are relatively small, there may be some resource usage issues due to static load balancing and persistent login.

Performance is generally good for this architecture, as production environments tend to deploy as many nodes as are necessary to meet workload requirements. Management cost can be an issue, however, as the use of small nodes means that there are many instances of the operating system to administer and maintain. These management challenges may be lessened in the future when Sun's N1 Grid Provisioning Server becomes available for data center servers (as of this writing, the N1 Grid Provisioning software is available for Sun's blade servers).

Diagonal 3-Tier Architecture

The diagonal 3-tier architecture is very similar to the horizontal 3-tier architecture except that larger servers are used as nodes. Instead of many one- to four-processor servers, the diagonal 3-tier architecture uses fewer eight- to 24-processor servers (Figure 20).

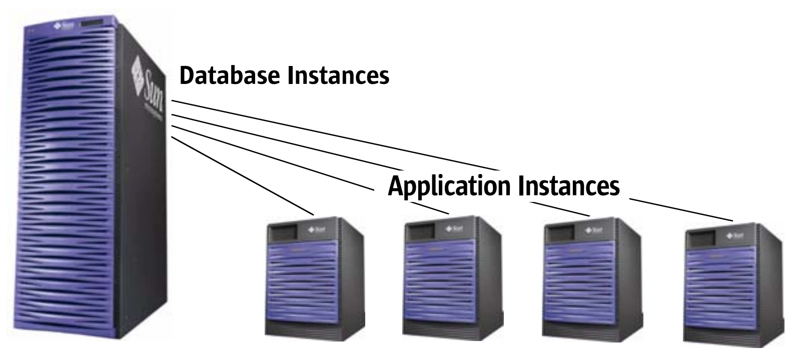


Figure 20. Diagonal 3-tier architecture: Database on a vertical server, applications on medium-sized multiprocessor servers

At least two application servers are typically used for availability so the size of the node is dependent on the total number of processors required to meet service level requirements. Since the nodes are larger (and fewer) than the horizontal architecture, resource utilization is generally better than the pure horizontally-scalable option. As the application deployment gets larger (more total processors) the diagonal option gets stronger, since the node count increases and the nodes themselves can host larger numbers of processors.

Since there are fewer nodes than the horizontal 3-tier architecture, complexity is reduced and management costs are likely to be lower in the long term. With the diagonal architecture, capacity can be increased within the server nodes themselves (more or faster processors) or by adding more nodes. A *mixed diagonal architecture* can also be created where the application servers are different sized systems. This approach is also highly flexible. Deployments might start with 12-processor nodes but add four-processor nodes if small increments of capacity are needed. Likewise, larger-capacity nodes could be added to respond to more significant capacity demands.

Application-Specific Issues

There are some application-specific issues that should be considered in concert with the architecture choices. For example, Oracle does not consolidate connections between the application tier and the database tier and there is a limit of approximately 8000 database connections per server. These issues may ultimately limit server size in the application tier.

Financial Analysis

Ultimately, no application architecture can be deployed independent of cost considerations. This section examines the various application deployment architectures from a financial point of view. For the application tier, the financial analysis is limited to the cost of hardware acquisition because this is easy to quantify and can be generalized. For the database tier (chapter 5) the analysis assumes the use of the Oracle database and includes both the server hardware acquisition costs as well as Oracle licensing costs.

Total cost of ownership (TCO) analysis is outside the scope of this paper but is recommended for customers seriously investigating application-tier architecture. TCO studies are best done on specific customer environments and are hard to generalize. Sun and its partners can provide detailed TCO analysis for IT departments that need them.

In order to compare the cost of hardware acquisition for various application-tier deployment options, a tool called the *App Calculator* was developed by Sun[1]. This internal tool is a simple spreadsheet that calculates the acquisition costs of two configurations under consideration. The calculator compares configurations that use small servers (called the horizontal option) and servers that are medium to large in size (the vertical option). It should be noted that all options are horizontal in some way in that they include more than one server or domain (except for the vertical 2-tier architecture). The key difference between the horizontal, diagonal, and vertical scenarios is the relative size, type, and number of servers. The inputs to the App Calculator as used in this analysis include:

- Number of 1.2 GHz UltraSPARC III processors required to meet the workload
- Baseline server utilization assumed in determining the number of 1.2GHz UltraSPARC III processors
- Horizontal node list price
- Assumed horizontal node hardware discount
- Number of processors in the horizontal nodes
- Vertical node list price
- Assumed vertical node hardware discount
- Number of processors in the vertical nodes
- Relative processor performance vs. 1.2GHz UltraSPARC III processors
- System utilization for horizontal nodes
- System utilization for vertical nodes

The analysis is normalized to the processor performance of 1.2 GHz UltraSPARC III processors for comparative purposes and the App Calculator adjusts the number of processors and servers needed using the relative processor performance input parameter. For example, a server with processors that are faster than a 1.2GHz UltraSPARC III processor will require fewer processors to meet the application requirements. The cost of acquisition for both the horizontal and vertical options are calculated using the number of adjusted servers multiplied by the per-server cost and less the assumed hardware discount. For this analysis, a 20-percent discount is assumed for Sun Fire V480, V1280, and E2900 servers while a 40-percent discount is assumed for Sun Fire 4800/E4900, 6800/E6900, and 15K/E25K servers[2].

The App Calculator makes further adjustments based on assumed levels of server utilization. The tool compares the server utilization of each server type with the baseline or assumed server utilization. If the server has a utilization that is different from the baseline utilization, the number of processors needed to meet the requirements is adjusted as necessary. A server utilization figure better than the baseline results in fewer processors (and possibly fewer systems) needed to meet the application performance requirements. Only Sun servers were used to compare the relative performance of the systems in order to reduce the number of variables in the analysis.

In most cases, application benchmarks run by Sun as well as reports from customers indicate that 1.2GHz UltraSPARC IV processors provide 1.8 to 2.0 times faster performance than 1.2GHz UltraSPARC III processors. For

1. The App Calculator is available through Sun Sales Representatives and partners
2. All of the system prices are based on U.S. list prices for anounced and shipping servers posted on www.sun.com as of March 25, 2004. All assumed discounts are for example purposes only and may not apply to all customer situations or geographies.

simplicity in this analysis, the 1.2GHz UltraSPARC IV processor is assumed to offer twice the relative processor performance of the 1.2GHz UltraSPARC III processor. Actual real-world speedup is highly application dependent and will vary with different circumstances. So that the cost of processor power can be compared equitably across all systems, the cost-per-equivalent-processor for the 1.2GHz UltraSPARC IV based Sun Fire Enterprise servers (Sun Fire E2900, E4900, E6900, E20K, and E25K servers) has been adjusted through division by 2 (or 1.75 for the 1.05GHz UltraSPARC IV processor). Clearly the servers that use UltraSPARC IV processors compare favorably due to their ability to process up to twice the number of concurrent threads.

Based on the earlier discussion, larger servers are assumed to be approximately 50 to 100 percent more efficient than the smaller servers. For this analysis, the system utilization for horizontally-scalable Sun Fire V480 servers is set at 30 percent. 30-percent utilization is generally optimistic for smaller application servers and many customers report utilization levels as low as 15 percent. Larger vertically- or diagonally-scalable servers are evaluated at both 30-percent utilization and also at a higher level of utilization (45-, 50-, or 60-percent) depending on the size of the server. These numbers are also conservative as Sun engineers have demonstrated 80- to 90-percent server utilization by using many of the technologies available to larger servers (processor sets, S9RM, Dynamic System Domains, and N1 Grid Containers). While the utilization numbers are somewhat arbitrary, the relative value of server utilization numbers is more important than the utilization numbers themselves. Relatively small differences in utilization can bring about dramatic differences in TCA.

The results of the App Calculator are graphed and shown in Figure 21 through Figure 26. The vertical axis in each graph represents the acquisition cost in dollars for the various systems based on the analysis and the bars show the acquisition costs of three different server options. The left-most bar is the cost of the horizontal 3-tier architecture using various numbers of fully-populated four-processor Sun Fire V480 servers. The middle bar is the acquisition cost of the corresponding vertical or diagonal architecture with the same assumed system utilization rate (30 percent in all cases). The right-most bar represents the acquisition cost for a different vertical or diagonal system option, made possible with the higher assumed system utilization. The right-most bar shows the affect on acquisition costs from higher projected levels of system utilization that is possible with larger servers or domains.

Vertical 3-Tier Versus Horizontal 3-tier Architecture

The App Calculator tool is designed to compare two different architectures and several scenarios were run to determine the relative acquisition costs of each architecture. To simplify the analysis, the horizontal 3-tier architecture is always one of the architectures in each comparison. The vertical 2-tier architecture was not analyzed because of the potential availability issues. The financial results comparing vertical 3-tier and horizontal 3-tier architectures are shown in Figure 21, Figure 22, and Figure 23.

UltraSPARC III Domains Versus a Cluster of 12 Sun Fire V480 servers

The first analysis assumes that the application servers in the vertical architecture are provided by UltraSPARC III domains added to existing servers. Multiple domains were used in this cost analysis, each representing a number of four-processor Uniboards (processor and memory boards) that can be deployed in any server from the Sun Fire 4800 to the Sun Fire E25K server. The cost of horizontally-scalable Sun Fire V480 servers was compared to the cost of adding the requisite number of UltraSPARC III Uniboards to an existing Sun Fire server (Figure 21).

This application deployment scenario should be of interest to IT managers that already have larger UltraSPARC III based servers with some available slots remaining. Even if a larger server was primarily purchased as a database server, the ability to add the application tier in one or more separate domains in the same server provides an attractive alternative to adding several smaller servers.

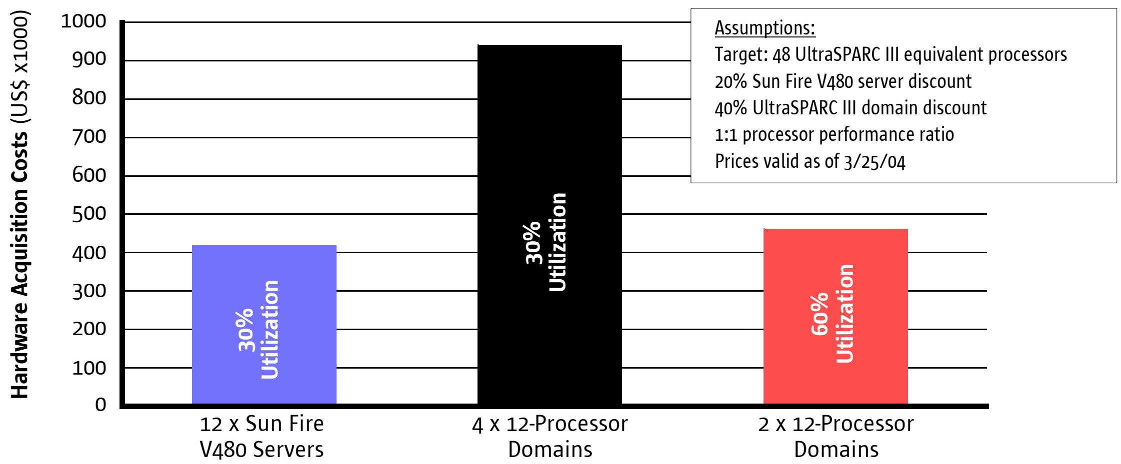


Figure 21. Assuming better utilization (60 percent vs. 30 percent), adding two 12-processor UltraSPARC-III domains to an existing Sun Fire 15K server compares well with the cost of twelve Sun Fire V480 servers

The target number of processors for this analysis was 48, requiring 12 Sun Fire V480 servers or four 12-processor UltraSPARC III domains in a larger Sun Fire 15K server. In this case the utilization was then doubled to 60 percent (versus the 30-percent baseline) to explore the effect on cost. Again, this increase in system utilization is certainly reasonable and even conservative in Sun's experience. With this better utilization, the processing requirement could be met with only two 12-processor UltraSPARC III domains, resulting in comparable costs to the horizontal

2-tier configuration. This cost difference is small enough to merit consideration by IT managers, particularly if they are concerned about floor space, are charged by the server in outsourcing contracts, or prefer the extra management, reliability, availability, and serviceability features of larger systems.

UltraSPARC IV Domains Versus a Cluster of 16 Sun Fire V480 servers

A similar cost analysis was done comparing multiple Sun Fire V480 servers to domains populated with chip-multithreaded UltraSPARC IV processors (Figure 22). In this case, the target was the equivalent of 64 UltraSPARC III processors, requiring either 16 four-processor Sun Fire V480 servers or three 12-processor UltraSPARC IV domains added to an existing Sun Fire Enterprise server (such as a Sun Fire E25K server).

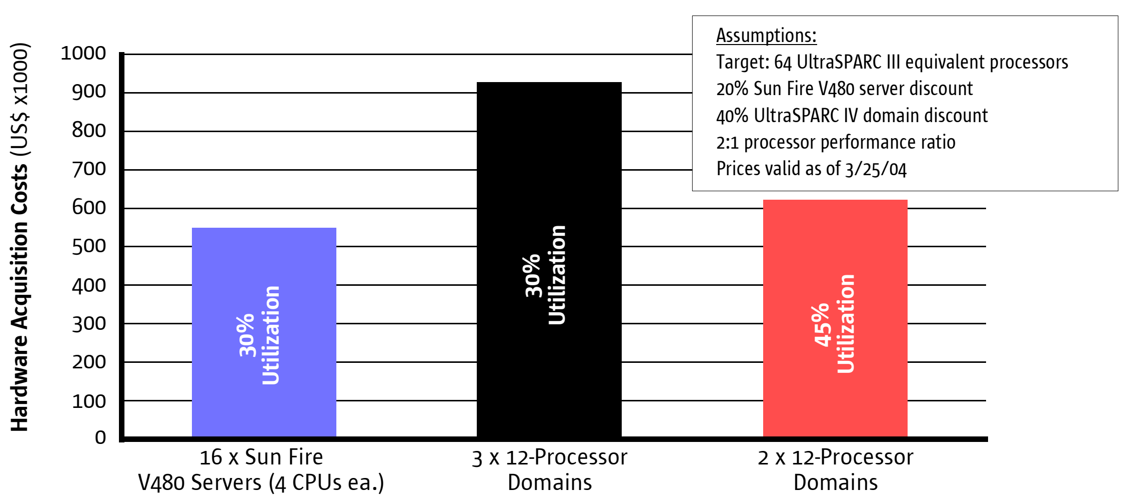


Figure 22. Assuming better utilization (45 percent vs. 30 percent), two 12-processor UltraSPARC IV domains are only slightly more expensive to acquire than 16 UltraSPARC III based Sun Fire V480 servers

When applying an assumed 45-percent utilization rate to the vertically-scalable option, only two 12-processor UltraSPARC IV domains were required to meet the capacity requirement. The two UltraSPARC IV domains cost only slightly more acquire than the 16 Sun Fire V480 servers, and are typically less complex to manage. These results clearly demonstrate that the vertical 3-tier architecture is a viable option assuming that the customer either has open slots in a larger server or that a large vertical server has already been chosen as the database server.

Sun Fire E6900 Servers Versus a Cluster of 16 Sun Fire V480 Servers

To further evaluate alternatives to a 3-tier horizontal architecture, a Sun Fire E6900 server was compared against a cluster of four-processor Sun Fire V480 servers. For this analysis, a total of 64 UltraSPARC III processors were required (configured into the 16 four-processor Sun Fire V480 servers). At the baseline 30-percent utilization rate, the App Calculator results showed that this computational challenge could be handled with two Sun Fire E6900 servers, each configured with 16 UltraSPARC IV processors.

At the higher assumed 45-percent utilization rate, a single Sun Fire E6900 server with 24 UltraSPARC IV processors could be deployed to meet the computational requirements. For redundancy and availability, this single Sun Fire E6900 server should be configured with at least two domains. With the Sun Fire E6900 server, the centerplane and the internal power grid can also be configured so that no single hardware failure can bring down all of the domains. In this case the adjusted costs for the 24-processor Sun Fire E6900 server is less than 20 percent higher than the 16 four-processor Sun Fire V480 servers, making this a viable option as shown in Figure 23.

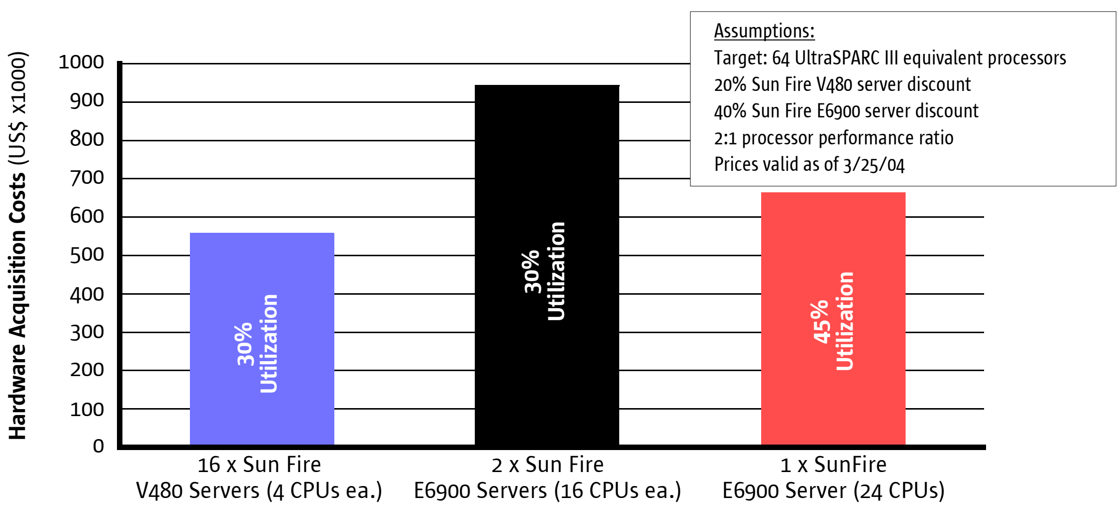


Figure 23. Assuming 45-percent utilization, a single 24-processor Sun Fire E6900 server compares well with the TCA of 16 four-processor Sun Fire V480 servers

Given the roughly equivalent cost, the greater simplicity (one chassis and two Solaris domains vs. 16 chassis and 16 Solaris instances) could be a factor for consideration. In addition, the better reliability and availability features available on a Sun Fire E6900 server domain (such as dynamic reconfiguration and online servicing) could add to the attractiveness of a vertical architecture solution in this scenario. While the Sun Fire E6900 server can be deployed in two domains with power and centerplane fault-isolation, some users may be uncomfortable of deploying their entire app tier in one server. In addition, a 24-processor Sun Fire E6900 server is at its maximum processor capacity, though a diagonal approach could be used for future expansion.

Diagonal 3-Tier Versus Horizontal 3-Tier Architecture

Three different analyses were done comparing a horizontal 3-tier architecture to a diagonal 3-tier architecture. The difference in each scenario is the size and type of the server used in the diagonal configuration. Again, the horizontal 3-tier architecture was represented by clusters of four-processor Sun Fire V480 servers in all cases. For the diagonal 3-tier architecture, eight-processor Sun Fire E2900 servers, 12-processor Sun Fire E4900 servers, and 16-processor Sun Fire E6900 servers were used in different scenarios.

Sun Fire E2900 Servers Versus a Cluster of 16 Sun Fire V480 Servers

Figure 24 shows the App Calculator results for two Sun Fire E2900 servers equipped with UltraSPARC IV processors. The target number of UltraSPARC III processors was 64, requiring a cluster of 16 Sun Fire V480 servers compared to just four eight-processor Sun Fire E2900 servers. Even with the same discounts, and the same 30-percent baseline utilization, the acquisition cost for these two options is very close. When the utilization is adjusted to 45 percent, the processing demands can be met by just two 12-processor Sun Fire E2900 servers (24 total UltraSPARC IV processors), yielding a lower projected TCA for the diagonally-scalable option.

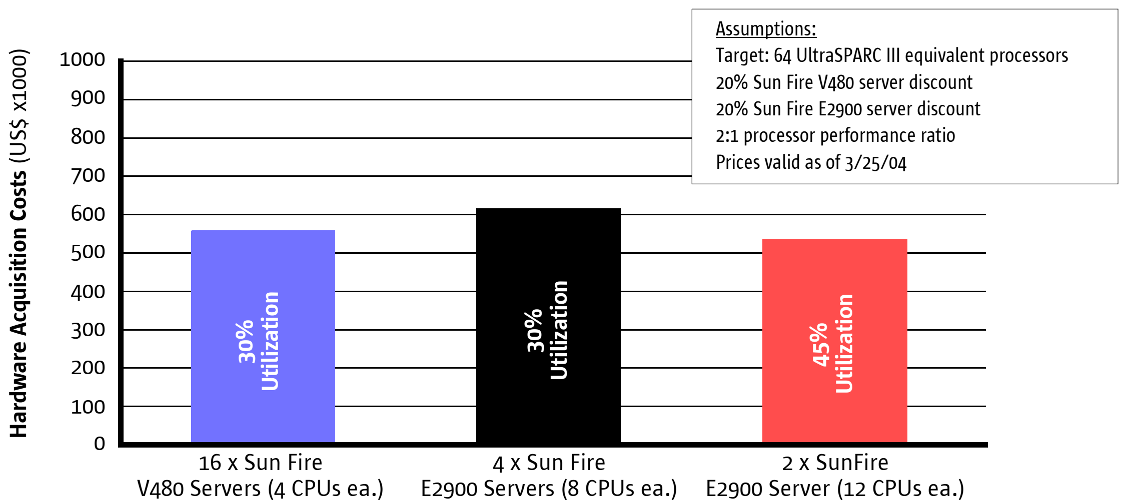


Figure 24. Assuming a 45-percent utilization rate, two 12-processor Sun Fire E2900 servers provide a lower TCA than 16 Sun Fire V480 servers

In this scenario, the utilization-adjusted cost for the diagonal solution doesn't change as much as other examples due to a higher per-processor price for the two 12-processor Sun Fire E2900 servers. Though not shown, three eight-processor Sun Fire E2900 servers would also fit the processing requirement, reducing the cost of the diagonal solution by 25 percent (one fewer eight-processor Sun Fire E2900 server). Three eight-processor Sun Fire E2900 servers would also provide for greater internal expandability than fully-populated (12-processor) systems.

Sun Fire E4900 Servers Versus a Cluster of 12 Sun Fire V480 Servers

In this scenario, 48 UltraSPARC III processors were the target and 12 clustered Sun Fire V480 servers were required to meet the performance requirement. Two 12-processor Sun Fire E4900 servers were used for the corresponding diagonal architecture (Figure 25).

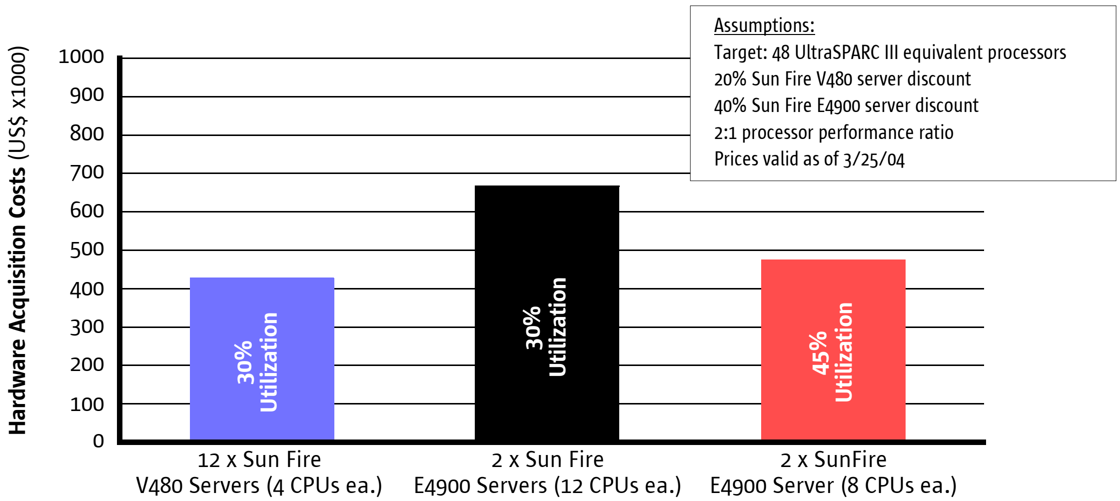


Figure 25. Assuming 45-percent utilization, two eight-processor Sun Fire E4900 servers compare well against 12 Sun Fire V480 servers

As shown in Figure 25, the costs for the baseline (30 percent) utilization Sun Fire E4900 server configuration is substantially higher. However, when the assumed server utilization is set to 45 percent, the diagonal option requires only two eight-processor Sun Fire E4900 servers, resulting in similar costs for the two options. In this case, both horizontal and diagonal options are technically viable and both have similar hardware acquisition costs.

Though the horizontal option may appear to have somewhat higher availability given the greater number of deployed servers, it also reflects greater complexity and possibly greater management costs. Each Sun Fire E4900 server can be partitioned into two Dynamic System Domains, increasing redundancy and promoting higher availability. High-availability technologies found on the Sun Fire E4900 and larger servers include complete hardware redundancy, hot-swap of critical components for online servicing, memory and processor offlining, and proactive diagnosis. These features serve to close any availability gap due to having fewer servers deployed in the diagonal architecture.

Sun Fire E6900 Servers Versus a Cluster of 16 Sun Fire V480 Servers

The results using a single Sun Fire E6900 server against 16 Sun Fire V480 servers are shown in Figure 23 and have been previously discussed in the section on vertical 3-tier architecture. Those results show that using a Sun Fire E6900 server in a vertical architecture for larger deployments is financially reasonable. This example provides a larger number of target processors (96 UltraSPARC III processors) to explore the cost implications of the Sun Fire

E6900 server in a diagonal configuration. In this case, 24 four-processor Sun Fire V480 servers would be needed to meet the 96-processor requirement (Figure 26).

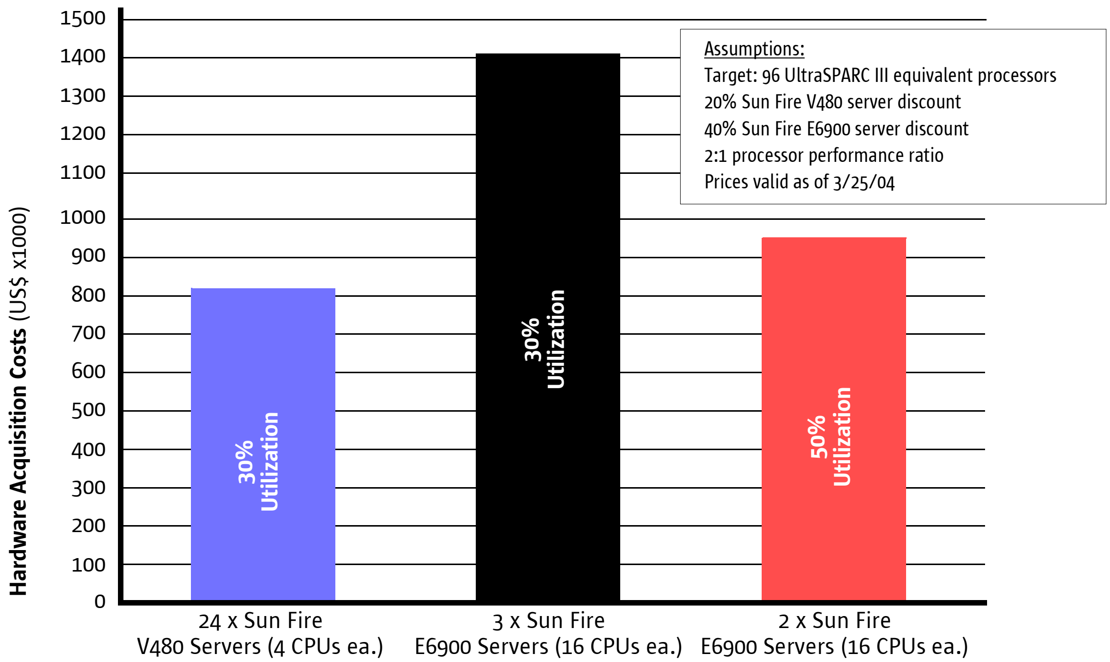


Figure 26. Assuming 50-percent utilization, two 16-processor Sun Fire E6900 servers compare well against 24 Sun Fire V480 servers

The same processing requirement can be met with only three Sun Fire E6900 servers, each with 16 UltraSPARC IV processors. By applying an assumed higher 50 percent utilization rate for the larger systems, the target load can be met by just two 16-processor Sun Fire E6900 servers. Again, the cost of the configuration is only slightly higher than the horizontally-scalable option with four-processor Sun Fire V480 servers. Like the Sun Fire E4900 server, the Sun Fire E6900 server can be partitioned using Dynamic System Domains for greater availability. In addition, the two Sun Fire E6900 servers should be less complex and costly to manage than 16 Sun Fire V480 servers and there is room for expansion in both Sun Fire E6900 chassis.

Chapter 5

Database Layer Considerations

The database layer has traditionally been deployed by vertically-scalable systems. With the advent of distributed databases such as Oracle 9i and 10g RAC, database architects are no longer strictly limited to vertical architectures. Databases can now be distributed across a number of systems, both to improve availability and to provide increased scalability. To evaluate the scalability and cost implications of this new architecture, Sun compared both single medium and large SMP servers against a cluster of smaller horizontal servers (four processors or less) from both Sun and Dell.

Performance Implications

In order to prevent confusion, it is helpful to define the terms used in this paper when discussing scalability:

- *Speedup* for an SMP server is defined as the degree of performance improvement on multiple processors as compared to single processor performance. Speedup does not vary based on the number of processors, meaning that the speedup value at 24 processors will be the same if it is measured at 48 processors.
- *Linear speedup* is ideal and is realized when an application runs, for example, 40 times faster on 40 processors than it runs on a single processor.
- *Cluster speedup* represents a similar concept but the base measurement is the number of nodes instead of the number of processors. Like SMP speedup, cluster speed up will be the same regardless of the number of nodes used in the measurement.
- *Scaling efficiency* is commonly used to measure the ability of applications (especially cluster applications) to scale to large numbers of nodes. However, scaling efficiency as defined here varies depending on how many nodes are used in the measurement.
- *SMP scaling efficiency* is the speedup divided by the number of processors.
- *Cluster efficiency* is the cluster speedup divided by the number of nodes. This value may be misleading since a 90-percent scaling efficiency on two nodes is not equal to 90-percent efficiency on four nodes[1].

Published benchmarks show that Oracle 9i RAC shows a cluster speedup of 1.8 or a scaling efficiency of 90 percent[2]. In contrast, internal Sun tests on Sun Fire SMP servers show that many Oracle database workloads deliver very good speedup including a 23-times speedup at 24 processors, a 57- to 60-times speedup at 64 processors and a 62.5-times speedup at 72 processors. The 95-percent scalability curve used for reference here is conservatively derived from a 23-times speedup at 24 processors. Interpolation of these SMP results to eight processors shows an

1. For a detailed discussion on this topic, please see “Deploying Oracle in the SunPlex™ Environment” (reference 3)
2. Published benchmark results as of April 2003 from www.oracle.com and www.SAP.com

efficiency of 99.5 percent. These results are graphed and extrapolated in Figure 27. The graph plots the following lines and curves:

1. Perfect linear scalability
2. 95% scalability on a 24-processor SMP server
3. 90% scalability on a cluster of two four-processor servers (extrapolated 2-processor data)

In the graph, both of the scalability results have been extrapolated beyond their measured values by using Amdahl's Law. Amdahl's Law can be used to predict speedup and to extrapolate to higher numbers of processors. This extrapolation can also be done using the geometric formula:

$$\text{speedup}(n)=1+(se)+(se)^2+(se)^3+\ldots.(se)^n$$

where *se* is the scaling efficiency and *n* is the number of nodes[1].

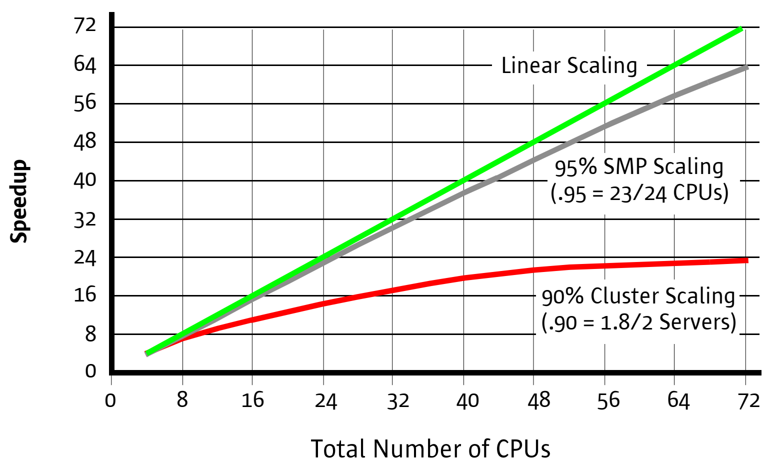


Figure 27. Extrapolated scaling curves for typical 90-percent cluster scaling and 95-percent SMP scaling

While 90-percent cluster scaling efficiency may sound good, it is clear from the graph that it is not very efficient when extrapolated to four or more nodes, or when compared to the superior scalability of large SMP servers. In fact, this graph shows that there are definite limits to the scalability of databases in clustered (or horizontally-scaled) environments. Connecting many small servers together does not actually deliver the scalability needed for medium and large sized applications. The reasons may be traced to the limitations of cluster interconnects, overhead in database software to manage cluster environments, and the difficulty of writing applications for distributed-memory clustered environments.

It is also apparent that it would be very difficult for clusters of horizontal systems to match the workloads that can be handled by large SMP servers of 24 processors and above. The SMP results with standard Oracle 9i (non-RAC) show much greater efficiency. Consequently, matching the results of a single 24-processor SMP would require 13 4-processor servers assuming the same processors, operating system, etc. The TCA implications of these results will be examined later in this chapter.

It should be noted that the cluster scaling used for this graph is based on OLTP applications. Some read-only databases that have well-behaved or pre-defined queries may show better horizontal scaling. In some cases it is possible to partition data to significantly reduce the amount of cross-node communication. In those cases cluster scaling may be better, but those cases are not common in commercial computing. Such partitionable workloads

1. For more information on geometric extrapolations, please refer to "Deploying Oracle in the SunPlex Environment" (reference 3)

are much more common in HPTC, including seismic processing, Monte Carlo simulation, and risk management. Cluster (or horizontal) environments work very well on these highly partitionable workloads.

Total Cost of Acquisition (TCA) Implications

Different levels of scalability for distributed versus single-system databases clearly have implications for the number of systems or processors that must be deployed to achieve an equivalent level of performance. Using the scalability issues for Oracle 9i/10g RAC versus non-clustered versions of the database, this section analyzes the effect of these scalability limitations on total cost of ownership (TCO). To make things easier to quantify, the analysis is restricted to total cost of acquisition (TCA). TCA as discussed here includes the cost of software licenses and servers. Storage and cluster interconnect costs are not included.

These issues are an important driver of cost because Oracle 9i RAC requires a different licensing scheme than the standard (non-clustered) version of Oracle 9i. For any system regardless of the size, Oracle Enterprise Edition licensing costs $40,000 per processor[1] for the database. This license fee is the same for both Oracle 9i RAC and standard Oracle 9i (non-RAC). For the RAC option, there is an additional $20,000 per-processor fee and an option of paying another $10,000 per processor for the capability to partition data. While the $10,000-per-processor fee is optional, tests show that data partitioning can provide better scalability in distributed databases. These amounts result in RAC license fees of $70,000 per processor as compared to $40,000 per-processor for standard Oracle 9i[1].

With these higher license fees for Oracle 9i RAC, along with lower scalability, it is easy to see how software licensing costs can quickly dwarf any savings in initial system cost from purchasing smaller, less-expensive servers.

Availability Implications

Availability is of great importance in today's data centers. Application services must be available around the clock without fail. Various schemes can provide high availability, depending on the requirements of individual applications. Vertical servers provide high availability with numerous reliability and availability features that minimize planned and unplanned downtime. Horizontal servers usually do not provide these features but achieve high availability by deployment of many redundant servers. The lack of reliability and availability technology, as well as slower system interconnects are the primary reasons why horizontally-scalable servers generally offer lower cost per unit of processing power than vertically-scalable servers.

Web server deployments provide a good example of horizontal high availability, since it is possible to deploy many small servers, each with a copy of the Web server software. If one Web server fails, the transactions are routed to the surviving servers. In the case of application servers, though both horizontal and vertical architectures may be deployed, replication remains the main technique for high availability.

The database layer, however, presents unique data sharing and availability requirements. Databases are stateful and by nature usually require that data is shared and accessible among all processors or nodes. To provide the replication necessary for high-availability databases, cluster software is required, such as Sun™ Cluster software or Oracle 9i RAC (for very high availability). Selecting a high-availability solution for the database layer requires an understanding of the tolerance for downtime (whether planned or unplanned). Figure 28 shows various availability percentages and the resulting hours of annual downtime.

1. All costs are U.S. list prices effective as of March 2004, and posted at: www.oracle.com

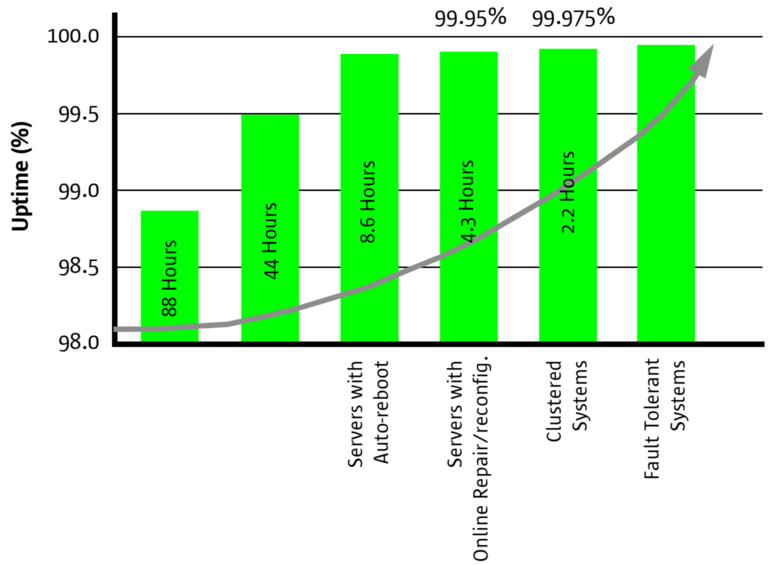


Figure 28. Uptime percentages and corresponding annual hours of downtime

As availability requirements increase, the cost and complexity of achieving those levels increases as well. Data center managers need to determine the best mix of cost, complexity and availability that meets their service level requirements. Those data centers that require up to 99.95-percent availability may be able to deploy single SMP servers (e.g., Sun Fire E2900 to E25K servers) with availability features such as full hardware redundancy and online serviceability through Dynamic Reconfiguration.

If greater than 99.95-percent availability is required, a cluster deployment provides a viable solution. Solutions such as Sun Cluster software with high availability (HA) failover can provide up to 99.975% availability through an *active/passive* configuration. HA failover works by using a primary server and a hot standby server. If the primary server fails, the standby server will take over processing. Time to restart the service for HA failover is application dependent and may take several minutes, especially for database applications that require extensive log roll-back to recover transactions.

If the data center cannot tolerate several minutes of downtime then the next option is an *active/active* solution such as Oracle 9i RAC where two or more nodes are deployed for the application to provide greater availability. Availability is perhaps the best usage of RAC technology and it is important to note that even two or more vertically-scalable nodes can be configured with Oracle 9i RAC. In this scenario, if one node fails the remaining node(s) continue to handle the workload. In case of a node failure with this configuration, users may not notice an outage or it may be very short (some customers report outages of less than one minute).

Architectural Case Studies

To evaluate the scalability and cost ramifications of vertical and horizontal architectural approaches, Sun has created an internal facts-based tool to calculate and compare the TCA of a single SMP system running Oracle 9i/10g versus a cluster environment running Oracle 9i/10g RAC that delivers the same performance. Called the RAC Calculator, this tool is not a model but simply an aid in calculating TCA of the server hardware and the Oracle database licenses[1]. All of the key input parameters are user-provided so that various scenarios and assumptions can be tested.

1. The RAC Calculator is available through Sun Sales Representatives and partners

The user enters the following information into the RAC Calculator:

- Name/model of target SMP server
- Number of processors of the target SMP server
- List price of the target SMP server
- Name/model of the cluster node
- Number of the cluster nodes
- Performance ratio of the processors used in the cluster node vs. the vertically-scalable system
- Per processor price of Oracle 9i/10g
- Per processor price of RAC option
- Per processor price of partitioning option
- Assumed SMP discount percentage
- Assumed cluster node discount percentage
- Assumed Oracle discount percentage
- Scalability percentage
- Percentage decay from one node to the next

The RAC Calculator first establishes the number of cluster nodes required for the horizontally-scaled cluster configuration and then identifies the number of processors required to match the performance of the single SMP system. The RAC Calculator uses a slightly different method to extrapolate performance than the method explained earlier (given two-node speedup). Instead, the calculator uses a straight line decay. The input scalability percentage (*sp*) is the percentage of extra performance benefit that each additional node provides. The decay (*de*) is applied to each successive node to reduce that node's contribution to the total result, as follows:

speedup(n)=1+(sp)+(sp-de)+(sp-2de)+...(sp-(n-2)de)[1]

Since the amount of decay is the same for each node, at some point the contribution of additional nodes approaches zero. The number of nodes(*n*) that contribute to increased performance in a cluster is *n=sp/de* so the smaller the size of *de* or the larger the size of *sp,* the more nodes will contribute. After the number of cluster nodes and processors are calculated the tool provides the TCA for both configurations using the user-supplied pricing data and assumed discounts. Lastly, the tool plots the TCA for both solutions and breaks out the amount of the TCA in terms of hardware and software costs as applicable. There is an option to add in the cost of SAN connections and cluster interconnects, but for the purposes of this paper that cost was assumed to be zero in all cases.

The following sections detail several case studies that analyze the TCA of both horizontal and vertical architectures in the database tier. The first three case studies use Sun Fire servers with UltraSPARC III processors as the target vertical servers. As previously mentioned, the Oracle licensing cost of UltraSPARC III processors is $40,000 per processor and the Oracle RAC license cost is $70,000 per processor[2]. Two case studies are also provided that use Sun Fire Enterprise servers with the faster UltraSPARC IV processors. Each chip-multithreaded UltraSPARC IV processor can run up to two concurrent execution threads and Oracle has chosen to treat each thread as a processor for licensing purposes. For this reason, the standard Oracle licensing cost for an UltraSPARC IV processor is twice that of an UltraSPARC III processor or $80,000. The RAC pricing is also double for UltraSPARC IV processors or $140,000 per processor with the data partitioning option. These costs are offset by the fact that UltraSPARC IV processors typically deliver up to twice the per-processor performance of UltraSPARC III processors.

1. This formula is very similar to the geometric formula described in "Deploying Oracle in the SunPlex Environment" (reference 3) for six-eight nodes, depending on the decay factor
2. All costs are U.S. list prices effective as of March 2004, and posted at: www.oracle.com

Case 1: Sun Fire 4800 and 6800 Servers Compared to a Sun Fire V480 Server Cluster

The first scenario compares Sun Fire 4800 and 6800 servers to a cluster of Sun Fire V480 servers (Table 4).

Server Configuration	Processors	Memory	Calculator Parameters	Assumed Discounts[a]	Configuration
Sun Fire V480 server	4 x 900 MHz UltraSPARC III	8 GB RAM	90% scaling 10% decay	50% Oracle discount 20% Sun discount	Oracle 9i RAC
Sun fire 4800 server	12 x 900 MHz UltraSPARC III	20 GB RAM	90% scaling 10% decay	50% Oracle discount 40% Sun discount	Oracle 9i
Sun Fire 6800 server	20 x 900 MHz UltraSPARC-III	32 GB RAM	90% scaling 10% decay	50% Oracle discount 40% Sun discount	Oracle 9i

a. Assumed discounts are for illustrative purposes only, and may not be applicable to all situations and geographies

Table 4. Server configurations and analysis parameters for Sun Fire V480, 4800, and 6800 servers

The results of the RAC Calculator analysis are shown in Table 5. Sun Fire V480 server clusters with different numbers of nodes were compared against equivalent Sun Fire 4800 and 6800 servers.

Servers Comparisons	Number of Nodes	Processors per Node/Total Processors	Hardware Cost[a]	Software Cost[a]	Total Cost
Sun Fire 4800 server versus	1	12/12	$229,680	$240,000	$469,680
a four-node Sun Fire V480 server cluster	4	4/16	$172,000	$560,000	$732,000
Sun Fire 6800 server versus	1	20/20	$346,360	$480,000	$826,360
an eight-node Sun Fire V480 server cluster	8	4/32	$345,600	$1,120,000	$1,465,600

a. U.S. list prices as of April 2003 from www.sun.com and www.oracle.com respectively with assumed discounts applied

Table 5. Single Sun Fire E4900 and E6900 servers compared to equivalent Sun Fire V480 server clusters.

Figure 29 graphically illustrates the distribution of costs for the above analysis, comparing a single Sun Fire 4800 server using standard Oracle 9i to a cluster of four Sun Fire V480 servers running Oracle 9i RAC. While the horizontally-scalable cluster solution has a lower hardware acquisition cost, the greater number of required processors and higher Oracle 9i RAC licensing costs make the horizontal solution more expensive to acquire by $262,320.

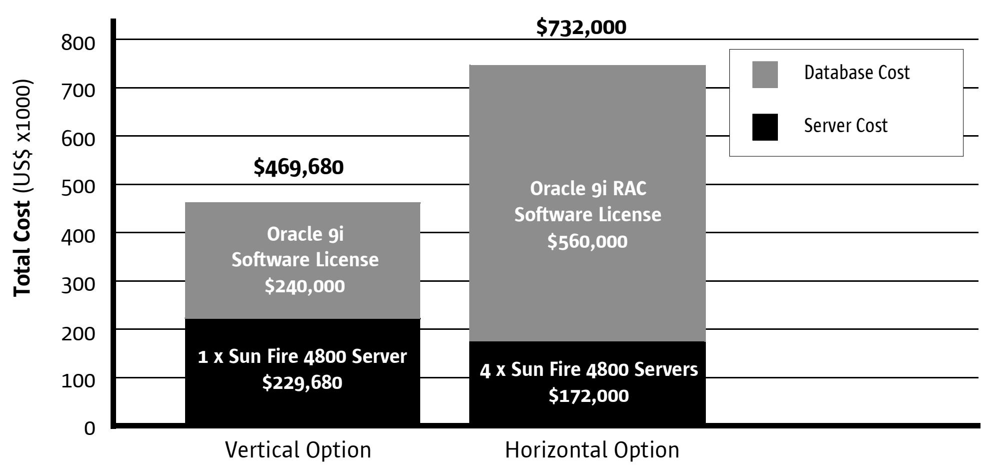


Figure 29. A single Sun Fire 4800 server running Oracle 9i has a lower TCA than a four-node, 16-processor Sun Fire V480 cluster licensed for Oracle 9i RAC

Case 2: Sun Fire SMP Servers Compared to x86 Clusters from Dell and Sun

The analysis was extended to compare single vertically-scalable UltraSPARC based Sun servers against clusters of x86 servers from both Dell and Sun. The Dell server used in the analysis was the Dell PowerEdge 6650 server with four 2.8 GHz Intel Xeon processors running Microsoft Windows 2000 Advanced Server. The Sun Fire V40z server with four 4.2 GHz AMD Opteron 850 processors running SUSE Linux 8.0 was also evaluated. Table 6 provides the configurations, prices, calculator parameters, and assumed discounts for these systems.

x86 Server Configurations	Processors	Memory/Disk	Calculator Parameters and Assumed Discounts	U.S. List Price
Dell PowerEdge 6650 server with Microsoft Windows Advanced Server 2000	4 x 2.8 GHz Intel Xeon	8 GB RAM/ 2 x 73 GB disks	90 percent scaling 10 percent decay 50 percent Oracle discount[a] 20 percent Dell discount[a]	$22,500[b]
Sun Fire V40z server with SUSE Linux	4 x 2.4 GHz AMD Opteron	8 GB RAM/ 2 x 73 GB disks	90 percent scaling 10 percent decay 50 percent Oracle discount[a] 20 percent Sun discount[a]	$25,000[b]

a. Assumed discounts are for illustrative purposes only, and may not be applicable to all situations and geographies
b. Source: U.S. list price as of September 23, 2004 from www.dell.com and www.sun.com

Table 6. x86 server configuration and analysis parameters

For the previous case study, the comparison between Sun SPARC systems was straightforward, given that the systems used the same processors, memory, operating systems, and storage. This analysis, however, required comparing both Intel Xeon and AMD Opteron based servers with servers based on both UltraSPARC III and UltraSPARC IV processors. In order to use the RAC calculator tool to compare servers with different processors, it was necessary to estimate relative per-processor performance. Unfortunately, no available database benchmarks directly compare similarly configured servers so the SPEC® CINT2000® benchmark was used to derive an approximation of relative performance. These results will probably differ for actual database applications, and actual application and database metrics should be utilized when planning actual application deployments.

Table 7 lists the CPU performance ratios used for the RAC calculator, derived solely from published SPEC CINT2000 results for comparable four-processor servers with different types of processors.

Performance Comparison	Systems	SPEC CINT2000[a] Values	CPU performance ratio
Four-processor UltraSPARC III versus Four-processor Intel Xeon	Sun Fire V480 server vs. Dell PowerEdge 6650 server	31.7 47.4	1.5
Four-processor UltraSPARC III versus Four-processor AMD Opteron 850	Sun Fire V480 server vs. Sun Fire V40z server	31.7 63.5	2.0
Four-processor UltraSPARC IV versus Four-processor Intel Xeon	Sun Fire V890 server vs. Dell PowerEdge 6650 server	59.8 47.4	0.8
Four-processor UltraSPARC IV versus Four-processor AMD Opteron 850	Sun Fire V890 server vs. Sun Fire V40z server	59.8 63.5	1.0

a. SPEC and the benchmark name SPEC CINT2000 are registered trademarks of the Standard Performance Evaluation Corporation. Competitive benchmark results stated above reflect results published on www.spec.org as of September 23, 2004. The following analysis was done using SPEC CINT2000 numbers that represent the best published results for the currently shipping versions of the Dell 6650 with four Intel 2.8 GHz Xeon processors, the Sun Fire V40z server with four 2.4 GHz AMD Opteron processors, the Sun Fire V480 server with four 1.2 GHz UltraSPARC III processors, and the Sun Fire V890 server with four 1.2 GHz UltraSPARC IV processors. The SPEC CINT2000 results for the Sun Fire V480 and V890 servers were generated using the Solaris 9 operating system. SPEC CINT2000 results for the Dell PowerEdge 6650 server were generated with Windows Advanced Server 2000 and the CINT2000 results for the Sun Fire v40z server were generated with Suse Linux 8.0. For the latest SPEC CINT2000 results, visit www.spec.org/cpu2000/results.

Table 7. Relative performance of several four-processor servers by SPEC CINT2000 ratios

Cluster deployments using both the Dell PowerEdge 6650 server and the Sun Fire V40z server were compared to single Sun Fire V1280 and E2900 servers, each with 12 UltraSPARC III processors and 12 UltraSPARC IV processors respectively. The results show that the faster x86 nodes narrow the hardware and software acquisition costs between single vertically-scalable servers and clustered servers. However, the first two results in Table 8 demonstrate that the hardware and Oracle software acquisition costs of single Sun Fire v1280 or E2900 servers are lower than the combined cost of the Dell PowerEdge 6650 servers equipped with Oracle 9i RAC.

Server Comparisons	Number of Nodes	Processors per Node/Total Processors	Hardware Cost[a]	Software Cost[a]	Total Cost of Acquisition
Example 1: Sun Fire V1280 server versus	1	12/12	$83,997	$240,000	$323,997
Three-node Dell 6650 server cluster	3	4/12	$54,000	$420,000	$474,000
Example 2: Sun Fire E2900 server versus	1	12/12	$199,797	$480,000	$679,797
Five-node Dell 6650 server cluster	5	4/20	$90,000	$700,000	$790,000
Example 3: Sun Fire E2900 server versus	1	12/12	$199,797	$480,000	$679,797
Four-node Sun Fire V40z server cluster	4	4/16	$80,000	$560,000	$640,000
Example 4: Sun Fire E2900 server versus	1	12/12	$199,797	$480,000	$679,797
Six-node Sun Fire V40z server cluster (60-percent scaling)	6	6/24	$120,000	$840,000	$960,000
Example 5: Sun Fire E6900 server versus	1	24/24	$672,777	$960,000	$1,632,777
11-node Sun Fire V40z server cluster (95-percent scaling)	11	11/44	$220,000	$1,540,000	$1,760,000

a. U.S. list prices as of September 23, 2004 from www.sun.com, www.dell.com, and www.oracle.com with assumed discounts applied

Table 8. Cost analysis comparing single Sun Fire V1280, E2900, and E6900 servers with Oracle 9i to equivalent Dell PowerEdge 6650 and Sun Fire V40z server clusters licensed for Oracle 9i RAC

In the third example, a single Sun Fire E2900 server was compared to a cluster of Sun Fire V40z servers, yielding only a small difference in combined hardware and software acquisition costs. It is important to note, however, that costs for the cluster interconnect, storage area networks (SANs) costs, as well as any cluster implementation and management costs are not included in this analysis. These additional costs will likely make clustered environments more expensive than the single server. In any case, this analysis shows that the speed of the individual clustered node does have a perceptible effect on deployment costs. Though all performance comparisons represent current “snapshots”, the smaller x86 servers will likely not increase in performance once they are deployed. While larger vertically-scaled systems such as the Sun Fire V1280, E2900, and E6900 servers are frequently upgraded with more or faster processors, horizontally-scalable servers are typically replaced with faster servers.

The previous examples in this case study use a scaling factor of 90 percent with a 10-percent decay factor. As has been mentioned before, these numbers may be very generous and some applications may not scale as well in clustered configurations. To understand the impact of lower scaling, the fourth example compares clustered Sun Fire V40z servers against the Sun Fire E2900 server as before, but with a reduced scaling factor of 60 percent. As shown in Table 8, much higher combined hardware and database acquisition costs result for the clustered solution.

Since many large databases can easily require more than 12 UltraSPARC III or UltraSPARC IV processors, the final comparison in Table 8 illustrates a single Sun Fire E6900 server with 24 UltraSPARC IV processors against a comparable cluster of AMD Opteron based Sun Fire V40z servers. In this example, eleven Sun Fire V40z servers were required to match the performance of a single 24-processor Sun Fire E6900 server. Because the effects of scaling and decay factors become more pronounced as more nodes are added, this comparison also required an optimistic 95 percent scaling factor for the clustered servers to match the single vertically-scaled server.

In most of the examples shown, the single vertically-scalable Sun Fire server had a lower TCA than the clustered solution, even though the hardware costs of clustered nodes were generally lower. In fact, the cost of Oracle 9i RAC software licensing (even at a 50-percent discount) so dominates the equation that even if the horizontally-scalable hardware were completely free of charge, that option still costs more than the vertical solution in most cases. Figure 30 graphically illustrates the second result from Table 8, showing that a 12-processor Sun Fire E2900 server licensed for Oracle 9i costs $110,203 less than an equivalently configured 20-processor Dell PowerEdge 6650 cluster licensed for Oracle 9i RAC. In this case, as with several other examples, the cost of the Oracle 9i RAC software license alone is more than the combined hardware and software acquisition costs for the vertically-scalable option. The high price for RAC licenses is made even more expensive by the lower scalability of the horizontal solution as compared to the vertically-scalable server (Figure 27).

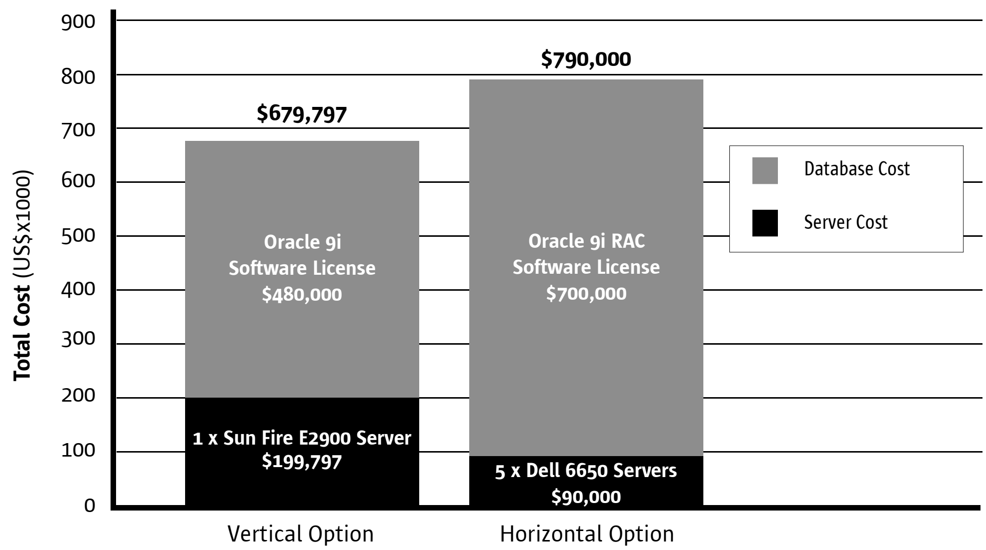


Figure 30. A single 12-processor Sun Fire E2900 server with Oracle 9i cost less than a five-node, 20-processor Dell 6650 server cluster licensed for Oracle 9i RAC

Case Study 3: Sun Fire Enterprise 4900 Server Compared to a Sun Fire V480 Server Cluster

This scenario compares a cluster of Sun Fire V480 servers to single Sun Fire E4900 and Sun Fire E6900 servers (Table 9). The Sun Fire E4900 server is equipped with the newer 1.2Ghz UltraSPARC IV processors. Several Sun internal and customer benchmarks indicate that a these new UltraSPARC IV based systems run applications at up to twice the speed of UltraSPARC III based systems with the same number of processors. As discussed, this increase in performance offsets the increase in Oracle licensing costs for UltraSPARC IV processors.

Server Configuration	Processors	Calculator Parameters	Assumed Discounts[a]	Configuration
Sun Fire V480 server	7 x 1.2 GHz UltraSPARC III	90% scaling 0% decay	50% Oracle discount 20% Sun discount	Oracle 9i RAC
Sun Fire E4900 server	12 x 1.2 GHz UltraSPARC IV	90% scaling 0% decay	50% Oracle discount 40% Sun discount	Oracle 9i
Sun Fire E6900 server	16 x 1.2 GHz UltraSPARC IV	90% scaling 0% decay	50% Oracle discount 40% Sun discount	Oracle 9i

a. Assumed discounts are for illustrative purposes only, and may not be applicable to all situations and geographies

Table 9. Server configurations and analysis parameters for Sun Fire V480, E4900, and E6900 servers

Note that for this analysis the value of the "Decay" is zero which assumes better scalability than the previous case studies. As can be seen in Table 10, and Figure 31, seven four-processor Sun Fire V480 server nodes would be needed to accomplish the same level of performance as a single 12-processor Sun Fire E4900 server. The same hardware and software discounts are applied in this case study as were used in the other examples. The results show that even with the doubling in per-processor Oracle licensing costs for UltraSPARC IV processors, the single vertical Sun Fire E4900 server is considerably less expensive than the horizontal solution.

Servers Under Comparison	Number of Nodes	Processors per Node/Total Processors	Hardware Cost[a]	Software Cost[a]	Total Cost
Sun Fire E4900 server vs.	1	12/12	$337,017	$480,000	$817,017
a seven-node Sun Fire V480 server cluster	7	4/28	$240,772	$980,000	$1,220,772
Sun Fire E6900 server vs.	1	16/16	$186,360	$640,000	$826,360
a nine-node Sun Fire V480 server cluster	9	4/36	$309,564	$1,260,000	$1,569,564

a. U.S. list prices as of March 2004 from www.sun.com and www.oracle.com respectively with assumed discounts applied

Table 10. Cost analysis comparing single Sun Fire E4900 and E6900 servers to comparable Sun Fire V480 server clusters

Figure 31 illustrates the cost advantage of a single vertically-scalable Sun Fire E4900 server licensed with Oracle 9i over the seven clustered Sun Fire V480 servers licensed for Oracle 9i RAC.

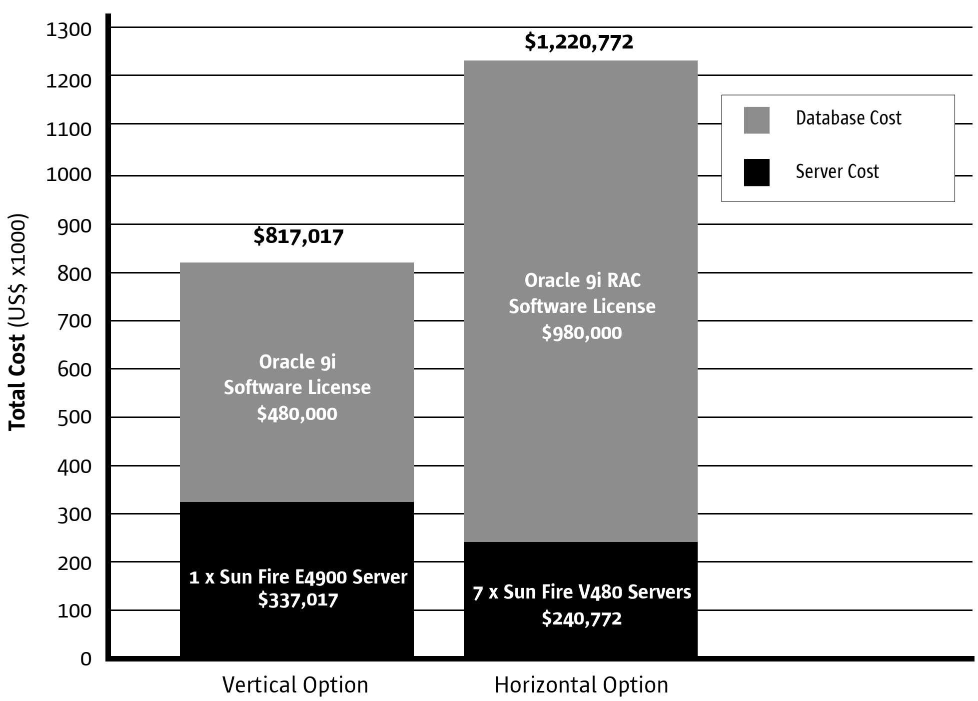


Figure 31. A12-processor Sun Fire E4900 server with Oracle 9i has a lower TCA than a cluster of seven 4-processor Sun Fire V480 servers with Oracle 9i RAC

Case Study 4: Sun Fire E4900 Server Compared to a Sun Fire v480 Server Cluster with Linear Cluster Scalability

Virtually all customer and public benchmarks indicate that Oracle RAC scalability is less than 90 percent. However, it is instructive to perform a TCA analysis assuming perfect scalability (100-percent scalability). For this assumption, matching a target Sun Fire E4900 server with 12 UltraSPARC IV processors would require six four-processor Sun Fire V480 server nodes to achieve the same level of performance (Table 11).

Server Configuration	Processors	Calculator Parameters	Assumed Discounts[a]	Configuration
Sun Fire V480 server	4 x 900 MHz UltraSPARC III	100% scaling 0% decay	50% Oracle discount 20% Sun discount	Oracle 9i RAC
Sun Fire E4900 server	12 x 1.2 GHz UltraSPARC IV	100% scaling 0% decay	50% Oracle discount 40% Sun discount	Oracle 9i

a. Assumed discounts are for illustrative purposes only, and may not be applicable to all situations and geographies

Table 11. Server configurations and analysis parameters assuming linear scaling

In this case, both the vertical and horizontal options require 24 Oracle licenses due to the different consideration for UltraSPARC IV vs. UltraSPARC III processors. However, the RAC Calculator analysis shows that even with 100-percent scalability, the more costly Oracle RAC licenses for the horizontal option more than offset the higher vertical server costs, making the vertical solution is less expensive (Table 12 and Figure 32).

Servers Under Comparison	Number of Nodes	Node/Total Processors	Hardware Cost[a]	Software Cost[a]	Total Cost
Sun Fire E4900 server vs.	1	12/12	$337,017	$480,000	$817,017
a six-node Sun Fire V480 server cluster	6	4/24	$206,376	$840,000	$1,046,376

a. U.S. list prices as of March 2004 from www.sun.com and www.oracle.com respectively with assumed discounts applied

Table 12. Cost analysis comparing single Sun Fire E4900 server to a six-node Sun Fire V480 server cluster

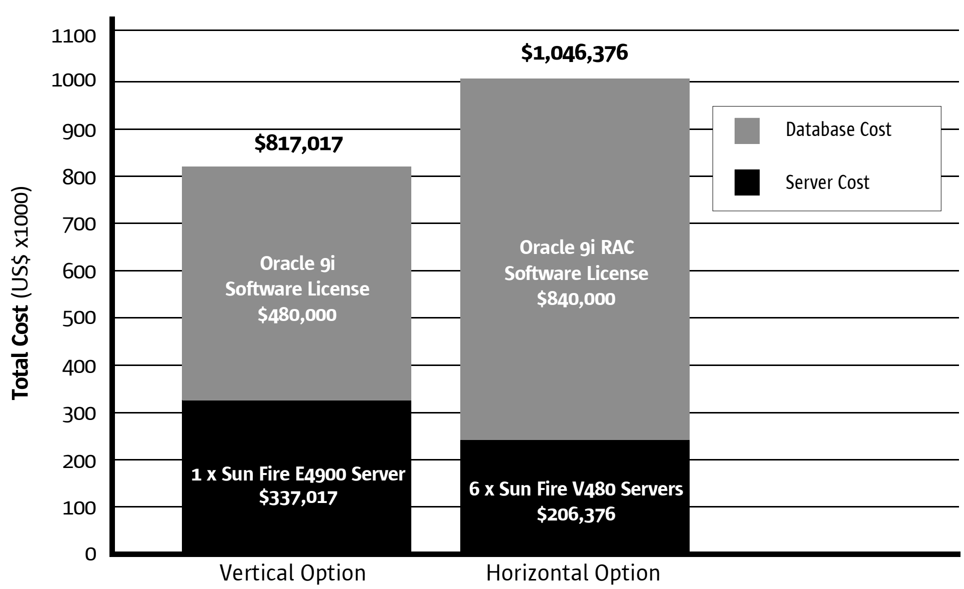


Figure 32. Even assuming linear scaling, a 12-processor Sun Fire E4900 server running Oracle 9i has a lower TCA than a six-node cluster of Sun Fire V480 servers licensed for Oracle 9i RAC

Case 5: Deploying Oracle 9i RAC for Availability: Many Small Servers or Fewer Large Servers

So far the analysis has considered the TCA and performance implications of vertical vs. horizontal (non-RAC vs. RAC) deployments. A third analysis was performed to evaluate the use of Oracle 9i RAC to achieve higher availability, independent of possible performance issues. This analysis was specifically designed to evaluate whether it is better to cluster a few large SMP servers or many smaller horizontal servers, assuming that the RAC option is necessary to meet the most demanding availability requirements.

For this analysis, The RAC Calculator tool was used to evaluate the TCA of a few large SMP servers as compared to many smaller servers for high availability. The target database was assumed to require the resources of a 20-processor Sun Fire 6800 server but with greater than 99.95% availability. The performance and availability requirements could be met with either two 12-processor Sun Fire 6800 servers or eight four-processor Sun Fire V480 servers both equipped with Oracle 9i RAC. The TCA analysis generated very similar results for both options and is shown in Table 13. The same pricing, discount, scaling and decay parameters were used in this analysis as provided earlier.

Server Configuration	Cost[a]	Configuration
1 Sun Fire 6800 server (20-way)	$826,360	20 processors with Oracle 9i licenses (non-RAC)
2 Sun Fire 6800 servers (12-way)	$1,461,360	24 total processors with Oracle 9i RAC licenses
8 Sun Fire V480 servers (4-way)	$1,465,600	32 total processors with Oracle 9i RAC licenses

a. U.S. list prices as of April 2003 from www.sun.com and www.oracle.com with assumed discounts applied

Table 13. Two clustered Sun Fire 6800 servers compare well with eight clustered Sun Fire V480 servers when both configurations are licensed for Oracle 9i RAC to provide greater availability

Even with the different total numbers of processors, the TCA for two Sun Fire 6800 servers is roughly equivalent to that of the Sun Fire V480 servers when both are equipped with Oracle 9i RAC licenses. As a result, a real-world deployment decision will likely depend on factors such as ease of management, lower complexity, and ability to grow over time. As previously discussed, adding a ninth node to the horizontal cluster may not deliver much scalability while adding more processors to the half-populated Sun Fire 6800 servers will likely deliver higher performance.

For smaller databases, the TCA for horizontal systems will be less than for two larger SMP servers. However, the complexity of managing multiple servers may make the management costs higher. This criteria has to be evaluated by the data center staff prior to making a platform decision.

Chapter 6
Conclusion

For successful and sustainable deployments, data center architecture must be aligned with the needs of both the application and the organization. Horizontal, diagonal, and vertical architectures are all viable approaches and all three can be shown to benefit applications depending on the circumstances. While horizontally-scalable architectures show the most promise at the Web tier of the data center, their use should not be blindly adopted in other tiers without careful analysis of the implications of scalability, reliability, availability, and cost.

With the availability of new distributed applications, vertically- and diagonally-scalable architectures can deliver multiple benefits at the application layer. Consolidation, both within individual tiers, and across the database and application tiers can help to simplify operations and management while avoiding performance hot-spots and improving utilization and availability. In addition, the better utilization rates provided by vertically- or diagonally-scalable architectures can equate to acquisition costs that are comparable or better than those of horizontally-scalable architectures.

At the database layer, vertically-scalable systems also have inherent advantages. The superior scalability of vertical solutions means that fewer processors and systems are required to run a given database goal. Fewer systems mean less complexity and lower management costs. When this scalability is taken into account along with the high licensing costs of distributed database technology, vertically-scalable database systems can provide acquisition costs that are substantially lower than those for horizontally-scalable systems. In fact, this analysis showed that even with a substantial discount, the license costs of Oracle 9i RAC alone for a horizontally-scalable configuration exceeded the combined cost of vertically-scalable hardware and Oracle 9i combined— often by hundreds of thousands of dollars.

Sun is dedicated to providing effective data center systems for all architectures whether vertical, horizontal, or diagonal. Along with Sun's proven and tested Reference Architectures and broad line of binary-compatible servers, Sun Services takes the time to understand the priorities of the business, helping organizations deploy the right architecture the first time. With the power and flexibility of the Solaris Operating System and innovative technologies such as the chip-multithreaded UltraSPARC IV processor and N1 Grid Containers, Sun preserves customers investments in technology, while helping them to make the most of their people and other resources.

Chapter 7

References

Sun Microsystems posts product information in the form of data sheets, specifications, and white papers on its Internet World Wide Web Home page at: *http://www.sun.com*. Look for the these and other Sun technology white papers:

1. *Vertical and Horizontal Computing Architectures* — Trends and Attributes, White Paper, Tom Atwood, SUPerG, Berlin 2003, Sun Microsystems
2. *Oracle 9I RAC Performance and Best Practices*, White Paper, David J. Miller, SUPerG, Berlin, 2003, Sun Microsystems
3. *Deploying Oracle in the SunPlex Environment*, White Paper, Jim Sangster, SUPerG, Berlin, 2003, Sun Microsystems
4. *Consolidation in the Data Center*, Sun BluePrints Book, David Hornby and Ken Pepple, ISBN 0-13-145495-8, Sun Microsystems Press
5. *Consolidated SAP Solutions — A Review of Costs*, White Paper, Dr. Christian Tolkes, Sun Microsystems GmbH

Sun Microsystems, Inc. 4150 Network Circle, Santa Clara, CA 95054 USA **Phone** 1-650-960-1300 or 1-800-555-9SUN **Web** sun.com

The Network is the Computer™

Printed in USA 00/00 XX0000-0/#K