

## WHITE PAPER

---

# SPARC-Based Systems Support Vertical Computing Architectures

---

Sponsored by: Sun

---

Vernon Turner  
February 2004

## EXECUTIVE SUMMARY

IT executives plan their computing infrastructures with care and consideration for the long term. In particular, application and database life cycles play a fundamental part in the selection of any server platform. Today, the server market offers a wide spectrum of choices based on many key criteria beyond the price of hardware. However, organizations are asking their IT executives to choose between vertical scaling systems (those with more than eight processors) and horizontal scaling systems (those with up to four processors). Often, IT executives are pressured to treat this issue as a binary decision or a black-and-white choice — that is, they must pick one system with little or no consideration for the fundamental requirements of the application or workload.

Although there are benefits to each system, the best IT infrastructure will maximize the benefits of both. Best practices show that the optimization of application, data, and systems is achieved by running the workloads on the technology that can fulfill all of their life-cycle needs. Therefore, it is important for companies to have an environment that has both vertical scaling systems and horizontal scaling systems. This white paper highlights business and technical requirements to inform the reader of the capabilities of vertical scaling systems while paying respect to horizontal scaling systems.

## INTRODUCTION

IT executives continually make trade-offs to deal with the growing pressure of supporting a broader array of application, platforms, and business-unit demands. As the industry examines the need to improve systemwide utilization through a variety of processes such as server and storage consolidation or "utility-like" computing, one trend that continually arises is that of using medium-sized to large servers or choosing smaller servers linked together in clusters or placed in racks. This white paper seeks to highlight the critical differences between using medium-sized to large servers to deploy a "scale-up," or vertical scaling systems, strategy and using smaller servers to deploy a "scale-out" or horizontal scaling systems strategy. The document's objective is to provide IT executives with the right level of information to allow them to make sound platform decisions as they attempt to match their application portfolios with their server needs.

---

## **Market Trends**

The IT infrastructure is a very dynamic market with many significant ongoing changes. These changes, while frequently hyped by the media, are often taken out of context. For example, market hype would suggest that there is an industrywide trend to move applications and workloads off both large and medium-sized servers and their operating systems to smaller and different operating systems. Although there are normal market churns of technology, research continues to show that enterprise computing is much more conservative than pundits would like mainly due to several key demands such as investment protection, application integration, and platform stability. As a result, mission-critical computing being performed in the application and database layers continues to be supported by vertical systems.

Another key trend is the renewed interest in IT consolidation (i.e., the balance of server and storage consolidation), where today's new platforms are much more capable of supporting, managing, and allowing multiple applications to grow and share resources in a more efficient manner. IT managers are looking to reduce the complexity of their infrastructures by pursuing consolidation and using large and medium-sized servers to accomplish this goal.

In conclusion, IT executives continue to realize the value of investing in medium-sized and large servers. Tactically, these servers continue to provide the capability and capacity to support workloads and applications that both grow and have unpredictable resource requirements in a safe and secure environment. Strategically, these systems are providing the rest of the IT infrastructure with well-tested innovation to support initiatives such as partitioning, processor designs, and operating system management.

---

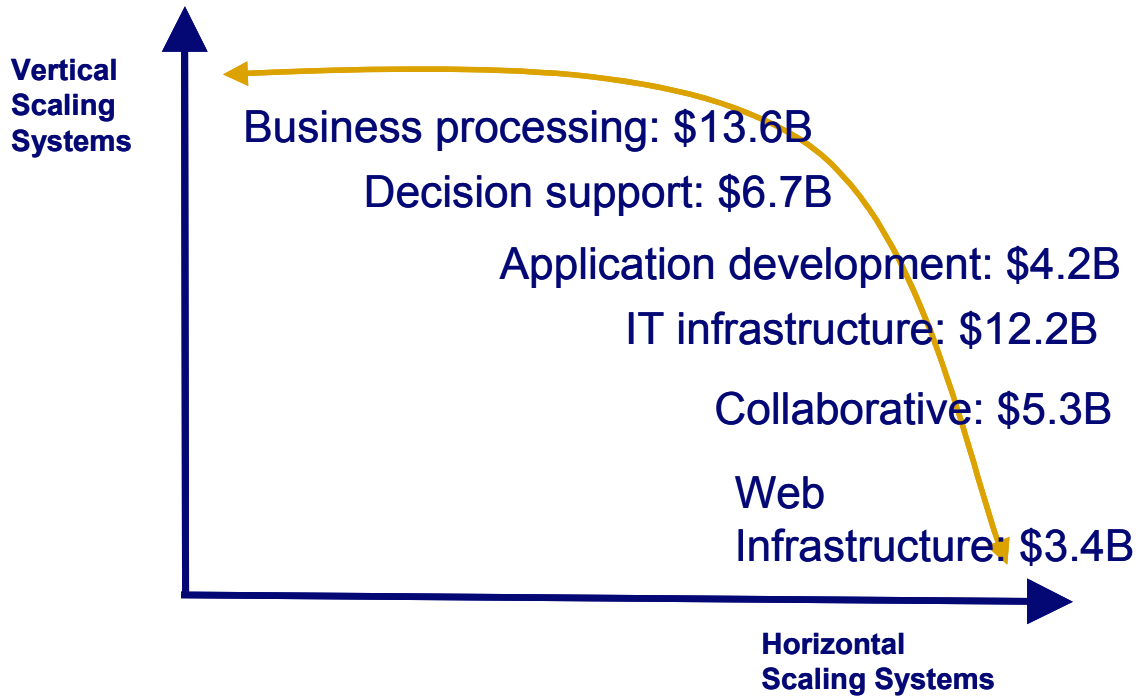
## **Driving Up or Out: Putting Mission-Critical Workloads on the Right Platform**

Application architects rely on their IT organizations to select the right platform for the right application or workload. For example, they assume that applications or groups of applications to form workloads that are stateless, small, and easily replicated are ideal for small servers in various form factors (e.g., server blades or rack-optimized servers). In contrast, workloads that are stateful and that support large amounts of data, many users, and large-scale internal data communication are ideal for vertical servers.

In 2003, IDC's research concluded that although horizontal systems shipped more individual servers than vertical systems, IT investment spending was still largely in the vertical systems segment to support the backbone of commerce (see Figure 1). These vertical systems are used for most of the business processing (consisting of transaction processing databases, batch processing, and ERP and CRM systems) and accounted for almost 28% (or more than \$13 billion) of all server sales. Horizontal systems succeeded in the Web infrastructure and collaborative markets where the system architectures make them a natural fit.

**FIGURE 1**

IT Infrastructure Investments: Customer Workloads Installed Base



Source: IDC, 2004

Vertical scaling systems are deployed extensively to support the most mission-critical workloads within an IT infrastructure. These workloads often need systems with the following attributes:

- Large shared memory space that is also cache coherent
- The ability to support many dependent execution threads for complex and highly integrated applications
- System configurations that can upgrade or add essential features such as faster processors, more RAM, and more and faster I/O connections
- A platform environment in which reliability, scalability, and availability are assumed features that work, require minimal support, and are built into the product and application architectures

## **Sun Microsystems' Vertical Systems Offerings**

### ***What Gives a Vertical Scaling System Its IT Endurance?***

A key point to remember when deploying mission-critical systems such as databases, data mining applications, or enterprisewide applications (e.g., ERP) is that these

applications have stringent processing requirements and will be installed in an infrastructure for a long time. To make these platforms run optimally requires very careful design from the processor to business recovery capabilities. Sun has recognized the advantages that it can give to its customers by optimizing every part of the IT architecture through its highly scalable SPARC-based platforms. The following is an extensive list of the proof points that Sun has made as part of its investment in its SPARC-based platforms. Along with these tangible investments are some of the very strong intangible benefits that come with this architecture:

- ☒ **Processor and system interconnect.** Building vertical scaling systems to support mission-critical workloads is often about creating balanced processing platforms that deliver on key qualities. Often, importance has been placed on processors' clock speeds, which is, in reality, a diversion for processors' real objective. Rather, processors should be measured on their ability to run at their maximum potential (often in a sustained compute-intensive mode). Furthermore, the more processors a parallel system has, the less important each one is. Now the important balanced system feature rests on the **system interconnect** to move data from disk, memory, and the network to the processors. In addition, the system interconnect supports any data addresses, keeping cache coherency in place and therefore avoiding system-based corruption. Sun continues to realize the importance of sustaining processors and system interconnects and plans to introduce faster interconnects (or pipes) that will enable larger (and more dense) compute-capable systems that take advantage of emerging multiple processors on a single silicon die.
- ☒ **Input and output.** The new commerce model created by the straight-through processing from pervasive computing devices to large-scale disk storage pools requires the server platform to support a very fast input/output (I/O) environment. The consequence of an I/O slowdown in any infrastructure will result in rapidly forming queues and delays, which translate to highly visible customer application slowdowns. Any I/O bottleneck will have a ripple effect on the best processors and system interconnects and render them ineffective. To that end, maintaining a balanced system that includes the I/O interface is critical.
- ☒ **Investment protection path.** Vertical scaling systems provide the IT backbone for database and application processing. These workloads typically have very long life cycles and therefore need to be placed on platforms that will sustain their growth or changing function needs. Customers are very aware of the need to maintain platform stability through long-term system upgrades that are timely and nondisruptive. Sun is addressing this need by delivering systems that have nondisruptive upgrades that allow its customers to take advantage of price/performance offerings and therefore support a strong investment protection path.
- ☒ **Operating system.** The choice of operating system is viewed as a critical decision when new applications are being written. However, the scalability of an operating system is often taken for granted because it is a nonentity in horizontal scaling systems. This is because individual instances of the operating system are usually confined to four processors or less. However, any mission-critical platform will be rendered ineffective if the operating system is not scalable (and

preferably optimized for performance) to support large single-system configurations that often extend beyond 32 and 64 processors. Sun understands this vital requirement and has optimized its Solaris operating system to meet it.

- ☒ **System availability.** System availability can be an IT organization's Achilles heel. Today's commerce model is built around continuous processing and high availability. Vertical scaling systems and horizontal scaling systems approach this challenge in different ways. Vertical scaling systems build availability into each server to improve planned uptime while minimizing unplanned downtime. For example, Sun has achieved many levels of nondisruptive maintenance, upgrades, and system replacements in a safe IT operational environment. Additionally, Sun's hardware and software partitioning allows applications and workloads to perform uninterrupted while other partitions are being worked on. Horizontal scaling systems typically do not receive this investment; rather, they obtain high system availability through replication and the deployment of many more servers. Although this solution may meet the requirement, it introduces more elements into an infrastructure and by default increases the overall single points of failure.
- ☒ **Optimized application.** As we mentioned earlier, vertical scaling systems have hardware and operating platforms that are optimized for maximum performance. These benefits are often passed directly to application architects by enabling the applications to be easily optimized. Consequently, Sun has recognized the challenge to undo these benefits by migrating the applications onto horizontal scaling systems. Often, these attempts are costly or simply ineffective and the original benefits for system selection are lost.
- ☒ **Consolidation.** IT consolidation (the process of keeping a storage and server consolidation initiative in sync) is fast becoming a mandate with CIOs. The need to share resources, raise overall system utilization, reduce platform complexity, and break down business-unit processing barriers can be readily met by the use of vertical scaling servers. Sun's investment in technologies such as server partitioning and nondisruptive upgrades allows IT organizations to embark on IT consolidation with a high degree of success.
- ☒ **"Utility computing."** Utility computing (or optimized resource sharing) is receiving a lot of investment from the leading systems companies (Sun with its N1 initiative) and interest from curious CIOs. IDC research has shown that a vertical scaling system with strong application and platform management capabilities can create a scalable server environment to meet unpredictable consumer demand. Sun's vertical scaling servers advance the vendor's N1 strategy via its binary-compatible features by enabling an IT organization to provision and virtualize an application on the most appropriate and cost-effective server.

The combination of these investments enables Sun to provide its customers with an application development platform and datacenter environment that is highly reliable and scalable. Because the SPARC architecture has been developed to deliberately scale up, it can stress-test the processors under more challenging conditions than those originally developed from the desktop or PC world. In doing so, Sun has been

able to generate vertical scaling systems that tackle system latency bottlenecks, delays, and queuing in an aggressive way to continually deliver balanced systems. Finally, such a linear scaling compute model has tremendous advantages for application developers. For example, they are working on a quicker and more reliable platform because the system's memory can be accessed as a single entity, yet shared safely through cache coherency techniques. Developers understand this important feature, and consequently ISVs continue to invest in the SPARC platform and turn out more software for it.

## **CHALLENGES AND OPPORTUNITIES**

Sun Microsystems has been able to invest and deliver innovative systems for many years and will continue to face a growing number of challenges in the vertical systems market. Sun's customers will want to see ongoing investment protection programs while they continue to take advantage of Sun's commitment to R&D. Sun's competitors will challenge its internal cost structure to sustain its R&D by attempting to convince customers to migrate workloads off its vertical scaling systems platforms and onto horizontal scaling systems.

Sun has already responded to these market conditions by expanding its range of horizontal scaling systems to encompass internally developed SPARC-based systems and partner-developed systems to support a mix of 32-bit and 64-bit applications. For more than 11 years, Sun has developed a rich legacy in building vertical scaling systems beyond 64 processors (even up to a 106-way server). One unique proposition that Sun has to support this scalable is its binary-compatible architecture using the SPARC platform to allow workloads and applications to be completely portable from an entry server system to a high-end mission-critical enterprise server.

In addition, Sun needs to demonstrate that the price/performance equation for vertical scaling systems is as effective as that of the horizontal scaling model. The benefits of Sun's investment in delivering a complete compute ecosystem for mission-critical workloads need to be highlighted and balanced against those of the market alternatives. Sun is responding to this challenge in several ways. It is demonstrating to customers that its vertical scaling systems can take advantage of all of the compute cycles and run at very high utilization rates. Sun is achieving this goal with the optimized SPARC architecture and its Solaris operating system, allowing its tightly coupled systems to run in a very efficient and effective IT mode and therefore creating a compelling price/performance platform.

## **CONCLUSION**

Systems vendors have to be able to provide their customers with a full range of server solutions. IT executives will place their application portfolios on a broad range of technology. To meet this requirement, Sun has established a complementary server portfolio based on an optimized SPARC/Solaris platform. The vendor has established its horizontal scaling systems for applications such as Web hosting and collaboration and its vertical scaling systems to meet the needs of larger, more integrated processing environments. As IT executives consider their server platform

options for mission-critical and business processing-oriented needs, they should think about the following key points with regard to vertical scaling systems:

- ☒ Balanced computing environments that are rich in technology investment
- ☒ Provide application and workload life-cycle stability through sustained innovation
- ☒ Create the capability to take advantage of ongoing business processing improvements such as IT consolidation and utility computing

IT executives also should note that hardware advances such as aiming to reduce vertical scaling systems' functions onto silicon (such as work that Sun is planning to do for its SPARC road map) will sustain the innovation and performance of large enterprise server systems ahead of the rest of the market.

---

### **Copyright Notice**

External Publication of IDC Information and Data — Any IDC information that is to be used in advertising, press releases, or promotional materials requires prior written approval from the appropriate IDC Vice President or Country Manager. A draft of the proposed document should accompany any such request. IDC reserves the right to deny approval of external usage for any reason.

Copyright 2004 IDC. Reproduction without written permission is completely forbidden.