SPARC/Solaris: the Optimum Platform for Mid-Range and High-End Servers

Sun Microsystems
White Paper
February 2004

Table of Contents

Executive Summary

For a customer-centric company like Sun Microsystems, the major determinant of high-end server architecture is the customer workloads that servers must tackle. For workloads from applications such as On Line Transaction Processing (OLTP), Customer Resource Management (CRM), Enterprise Resource Planning (ERP) and other large network-based applications, the SPARC/Solaris platform from Sun is the optimum platform. Over the years, both the SPARC® processor and the Solaris™ operating system have been deliberately designed to work together to support the highest performance and scalability in systems ranging from desktop workstations to large, mission-critical servers. Sun™ high-end servers with as many as 72 SPARC processors provide the massive scalability needed to handle today's large data center workloads.

The SPARC/Solaris platform has been developed and refined over the years with continuous technological advances and improvements, always with the goal of meeting and anticipating customer needs. The latest versions contain new capabilities such as chip multithreading (CMT) and throughput computing, high throughput interconnects, predictive self-healing, and automatic error detection, diagnosis and recovery of systems. Throughput computing alone increases application performance and workload throughput not by incremental percentages but by double. This would enable the Sun Fire™ E25K server with 72 SPARC processors to operate on 144 threads at one time. Today, these and other innovations have given the SPARC/Solaris platform the lead in high-end server performance and capabilities.

Meeting Increasing Data Center Workloads

Today's mounting workloads, driven by the constantly expanding use of the Internet and huge, ever growing databases, have placed increasingly heavy demands on servers. As a result, companies must scale up (vertical scaling) their data centers with ever more capable high-end, SMP (symmetric multiprocessor) servers, or scale out (horizontal scaling) in enlarging clusters of small servers.

Some vendors claim that clustered small servers and blade servers are displacing larger, vertically scalable, high-end servers in the data center. However, the facts dispute this claim. Dataquest reports that large SMP servers are actually a growing portion of total server sales. From 2000 to 2002, the percentage of sales of midrange to high-end servers priced above $500,000 jumped from 38 percent to 52 percent. IDC believes that midrange to high-end servers will continue to be the dominant data center platform. However, other IDC data indicates growth as well in numbers of small (two CPU) servers shipped, so small servers in clusters will still complement large servers in the data center.

The fact is that data centers need both horizontal and vertical systems. Certain workloads require horizontally scalable systems, while others demand the vertical scalability that can only be deployed on midrange and high-end SMP servers. This situation can be seen in the division of the data center into three computing tiers. The presentation layer (tier 1) receives transactions that are then sent to the application layer (tier 2) for processing. When data must be read or written, the transaction is passed to the database layer (tier 3). Horizontal architecture (clustered small servers) normally controls tier 1, while vertical architecture (high-end SMP servers) dominates tier 3. Tier 2 can be either horizontal or vertical, usually depending on the size of the workload.

A variety of ever increasing workloads, typically best suited for high-end SMP servers, is driving sales of these servers. These workloads include:

- Mission-critical workloads that can be more efficiently and cost-effectively consolidated onto larger, multi-domain servers.
- Key applications, such as high-volume OLTP, Business Intelligence and Data Warehousing, and large database applications that can only be deployed on vertically scalable SMP servers
- Mission-critical, enterprise-class application tier (tier 2) and database tier (tier 3) workloads
- Some HPTC (High Performance Technical Computing) applications that require a combination of large numbers of powerful processors, large memory and high-throughput interconnects

Sun's family of massive, vertically scalable, highly reliable SMP servers, based on its SPARC/Solaris platform, is ideal for these workloads.

Some System Characteristics of Sun's High-End Servers

A Sun high-end SMP server based on the SPARC/Solaris platform can be configured with either one or multiple domains. If configured with a single domain, the server has one instance of the Solaris operating system that covers the number of processors, memory and I/O connections installed in the server. Sun Fire servers from the E2900 to the E25K can be configured with multiple Dynamic System Domains. Each fault-isolated domain contains a separate instance of Solaris for the number of processors, memory and I/O connections configured into it.

Perhaps the most unique feature of Sun's high-end SMP servers is the industry's first and only triple redundant crossbar interconnect. The heart of the system, the interconnect contains three redundant crossbars. Separate address, response, and data paths allow two-way traffic, which enables multiple, simultaneous transfer of data between processor, memory and I/O system.

The crossbar interconnect is a key enabler of Sun's large, scalable systems. Sun combines the SPARC processor architecture with the crossbar switch to create a parallel bus, where most other processor buses are serial buses that only allow one processor to access the bus at a time. Sun's crossbar interconnect enables multiple processors to access 32 bytes of data at the exact same time, significantly speeding performance.

The system interconnect in a Sun high-end server is tightly coupled, low latency, high bandwidth and cache coherent (i.e., it maintains information on the location of all data regardless of cache or memory location). If needed, more resources can be added by inserting system boards into the server. Assuming no partitioning, all added resources (memory, processor, I/O connections) are controlled by one instance of Solaris. On these high-end servers, memory is shared, which means that all processors and I/O connections have equal access to all memory, which appears as one large unified resource.

With Sun's N1 initiative – the company's architecture for the next generation data center – customers can treat the whole data center as a single system. N1 aggregates widely distributed computing resources (servers, storage, software and networking) and enables them to operate as a single, powerful, unified entity that can be shared by different applications. When fully implemented in the future, N1 will automate the tasks of resource configuration and dynamic provisioning of capacity to meet fluctuations in the customer's computing load.

In some ways, a Sun Fire multi-domain server can be considered as "N1 in a Box." Many enabling technologies key to the complete N1 vision are already in use as system and resource management tools for Sun Fire servers. These are all based on the latest versions of the SPARC/Solaris platform and include:

- Dynamic System Domains: Fault-isolated, flexible partitions that allow customers to run multiple copies of Solaris on a single server. For flexible resource management, Dynamic System Domains can be adapted to meet customers' changing application environments.

- Dynamic Reconfiguration: As part of the enabling technology supporting Dynamic System Domains, Dynamic Reconfiguration allows customers to add, remove or reconfigure hardware resources on the fly, without the need to reboot. Dynamic Reconfiguration helps customers keep applications running while cost-effectively managing IT resources.

- Third Party Interface: By supporting the WBEM (Web-Based Enterprise Management) standard, Sun improved Dynamic Reconfiguration by introducing an application program interface that enables independent software vendors to develop advanced resource management software. Third-party solutions like BMC Patrol from BMC Software promise to improve functionality, flexibility and ease-of-use when monitoring and allocating system resources.

- Solaris Resource Manager/Solaris Containers: Solaris Resource Manager/Solaris containers isolate software applications or services using flexible, software-defined boundaries. Solaris containers create computing environments within a single instance of Solaris, and provide full resource containment, fault isolation and security isolation.

These technologies transform Sun Fire servers into pools of resources that can be managed and provisioned dynamically – and soon automatically – to meet changing workloads. Such capabilities promise to help data centers better utilize resources, reduce complexity and reduce total cost of ownership (TCO).

Additional System Attributes of SPARC

The UltraSPARC® processor family is a key element of Sun's customer-driven design philosophy, which, together with the Solaris operating environment, provides the foundation for a robust, scalable computing platform. With customer workloads increasing, Sun has made significant modifications in the latest SPARC designs that considerably increase system RAS (reliability, availability and serviceability) and make it an especially capable processor for high-end servers. Sun designed SPARC specifically to be able to scale the interconnect to support large coherent memory systems, large configurations and large workloads. Sun further optimized the interconnect to allow the system to contain multiple processors.

In addition, the SPARC interconnect architecture supports the directory-based cache coherency scheme that allows for expansion of the coherency protocols and enables the building of larger systems. Without this capability, the largest system possible would be the Sun Fire E6900 with 24 CPUs. With this interconnect architecture, Sun can build systems with 100 or more processors. Currently, the largest server in Sun's product line is the Sun Fire E25K with 72 processors.

The interconnect system also has a robust error detection capability through an independent service processor. If an error occurs within the system, the system may or may not be able to determine what that error is without destroying the error data, or recreating the same error. Given access to all the error detection logic within the system, the service processor independently determines what the failure is and reports on it back to the operating system. This service processor basically brings the system back up, tests it and makes sure it is functionally correct – and, in the process, significantly improves the system's RAS.

Sun also optimized the infrastructure in the SPARC architecture that supports Solaris for large systems. This is seen in aspects of the system such as interrupt handlers, page tables, I/O structures, etc. SPARC handles page tables more efficiently, as well as larger page sizes, which means fewer entries with which to be concerned. The system is optimized for both large memories and a large number of processors.

Yet, with all these advances and improvements, the most fundamental and important change in SPARC architecture today is the chip's new multithreading (CMT) technology that makes throughput computing possible. CMT is a design concept that allows the processor to execute multiple threads simultaneously, radically increasing application throughput. Not only are data center workloads growing rapidly, but they are also becoming increasingly multithreaded with network-based applications such as OLTP, ERP, videoconferencing, decision support and Web services. These applications depend less on the execution of a single thread (a set of software instructions that executes independently) than on overall throughput.

Solaris is already considered the best commercial operating system for multithreaded SMP-based systems, which it has supported since 1992. Now Solaris, together with the latest UltraSPARC processors and their CMT capabilities, enable high-end servers based on the SPARC/Solaris platform to offer greater overall throughput.

Throughput Computing With SPARC/Solaris

Sun's CMT technology, which enables a microprocessor to concurrently execute multiple threads, makes throughput computing possible. To work successfully, a CMT processor must be combined with a hardware platform that can supply the necessary memory capacity – as provided by 64-bit addressing capability – and an operating system like Solaris that can efficiently handle and schedule a heavily threaded environment. This results in a significant increase in application throughput or aggregate useful work done by the processor.

Sun designed the new UltraSPARC IV processor specifically to maximize throughput for network computing workloads. This processor supports multithreading at the chip level by processing multiple threads simultaneously.

The UltraSPARC IV tackles latency, or memory delay, in a very innovative way. When one thread waits for memory to deliver data, the processor switches to another thread that is ready to run. This process then repeats for several threads. The processor then switches back to the first thread when that thread receives its data and is ready to execute. The cycle then repeats. Using CMT, a single processor can process multiple threads simultaneously, considerably increasing the amount of data processed in a given amount of time. UltraSPARC IV will potentially double the throughput of Sun's high-end and midrange systems.

The UltraSPARC IV is a dual-threaded processor. Built with 130 nanometer (nm) process technology, the processor will operate at an initial frequency of 1.2 GHz, moving higher over time in later family members. Depending on customers' application workloads, performance improvements will range from 1.6 to 2.0 times the throughput of today's 1.2 GHz UltraSPARC III processor. Future UltraSPARC IV processors, to be built with 90nm technology, will increase throughput 3 to 4 times that of today's UltraSPARC III.

Where SMP delivers throughput computing at the system level, CMT does so at the processor level, providing a significant reduction in the cost of network computing. Since many Sun customers already use multithreaded SMP-based systems, they will not need to change their software model. Throughput computing can be directly applied to their software-installed base, and no recompiling or reprogramming is necessary. The Solaris operating environment and Java technology are already designed to support multithreading.

Since a large amount of memory is needed to support this computational capacity and to scale with the power of the CPU, 64-bit addressability by the processor is required, as opposed to the limited address space imposed by 32-bit addressing. To facilitate this addressability, the number of bits allotted to accommodate addresses in directories on the UltraSPARC IV processor is 43. In addition, internal structures like Transition Lookaside Buffers (TLBs) in the memory table have been sized to contain large virtual and physical addresses.

The new UltraSPARC IV also maintains Sun's tradition of binary compatibility that protects customer investments in development tools and application software. These innovative CMT processors also give the customer a price/performance advantage by significantly increasing throughput performance without increasing chip costs. Since CPU costs are normally based on volume and die size, CMT chip costs are expected to be comparable to those of traditional processors.

Most important, an operating system is needed that can efficiently handle and schedule such a heavily threaded environment. Neither Windows nor Linux has this capability. Solaris is regarded as having the best threading model among commercial operating systems today. Using the UltraSPARC IV processor, Solaris can run upwards of 200 simultaneous threads on an SMP system.

Additional System Attributes of Solaris

The Solaris operating system has been supporting SMP environments since the late 1980s. All technologies that Solaris uses to support large SMP systems are directly applicable to CMT, making the operating system ideal for hosting multithreaded processes. Solaris can schedule multiple concurrent tasks both inside and outside the Solaris kernel. Solaris' unrivaled scalability allows it to maintain a very high threading level and run many threads efficiently.

The technologies honed through consecutive releases of the Solaris operating system have resulted in an extremely fine-tuned, multithreaded kernel for running on multiprocessor machines like the Sun Fire E25K. With Java performance also tightly integrated into it, Solaris produces the highest performance on Java, where even the smallest Java application contains many threads. If numerous Java programs are run on a system, the ability to deal with many threads intelligently is crucial. All associated functionality libraries have also been fine-tuned, an ongoing process for the last 12 years.

The Solaris kernel also contains technologies such as memory positioning optimization that automatically locates both the optimal CPUs and the optimal memory space for an application on which to run. This is critical for the new UltraSPARC IV processor that is essentially a "server on a chip" – instead of multiple processors running multiple threads, each individual processor runs multiple threads. For example, a Sun Fire E25K containing 72 UltraSPARC IV processors could simultaneously run 144 threads.

Another important Solaris innovation is the Sun Reliance Architecture (SRA), of which the first phase will be included in Solaris 10 and is available in the new Sun Fire servers. SRA provides a new level of diagnosis and self-healing for both software and hardware faults. SRA predicts, detects and isolates errors before they occur – and before they can impact applications. SRA also identifies and tracks system faults as they occur, aggregates and analyzes them, and recommends actions such as taking a particular component offline to keep service levels up. SRA also logs the error, making the information available to Sun for remote diagnostics. If needed, SRA enables service personnel to be dispatched before the user even knows a problem exists.

SRA reduces error diagnostics time by 50 percent, and its accurate diagnosis allows for a fast return to the normal state. SRA also provides on-the-fly non-disruptive maintenance. With dynamic reconfiguration and hot patching, system faults can be repaired without interruption – and without rebooting.

SRA further differentiates Solaris from other UNIX® operating systems that do not have the features to run large numbers of processors, nor to correct itself from faults. In how these operating systems handle errors, they could be referred to as "crash-and-reboot" operating systems. Solaris with SRA, on the other hand, handles errors more gracefully and avoids the need to reboot. With SRA, Solaris' infrastructure allows for handling not only more errors, but also more types of errors.

Many Solaris features and capabilities combine with those of the new UltraSPARC IV to make SPARC/Solaris the best choice for a high-end server platform today. Solaris is designed to deal with what customers need – better security, scalability, utilization and performance – and with fewer headaches through better manageability. Solaris, above all, is designed to make a system easier to maintain.

Summing Up

Sun is, above all, a customer-centric company. Sun studies its customers' workload needs and seeks solutions that answer those needs. Sun believes that its SPARC/Solaris platform can answer customers' needs in systems of all sizes – and, with its latest technological advances, especially in high-end servers.

Today's IT departments must deliver the highest reliability, availability, serviceability, throughput, scalability, compute density and upgradability. Sun Fire servers is a family of massive, vertically-scalable, highly reliable SMP servers that are ideal for server consolidation, mainframe rehosting, HPTC applications and enterprise-class application tier (tier 2) and database tier (tier 3) applications.

Sun high-end servers are tightly integrated and developed within a singular mind-set focused on overall performance and customer satisfaction. These servers:

- Are capable of massive scalability, incorporating a very large number of processors in a system – SPARC and Solaris are key enablers of scalability, and Sun leads the industry in optimizing hardware and software to increase the size of a system rapidly and reliably

- Offer binary compatibility to protect software investments – Sun systems have legacy software operating reliably in mission-critical applications

- Deliver mission-critical performance in every area that counts: scalability, application performance and processor bandwidth

Sun is a vertically integrated company that has focused on high-end servers for many years. In that time, Sun has developed exceptional expertise and engineering capabilities that allow it to design and integrate all parts of the system – the microprocessors, the operating system, the interconnects, etc. – and enable these elements to work together impeccably. By maintaining control of the whole system and its parts, Sun has achieved the optimum designs for its high-end servers based on the SPARC/Solaris platform.

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

We make the net work.

Sun Worldwide Sales Offices: Argentina +5411-4317-5600, Australia +61-2-9844-5000, Austria +43-1-60563-0, Belgium +32-2-704-8000, Brazil +55-11-5187-2100, Canada +905-477-6745, Chile +56-2-3724500, Colombia +571-629-2323, Commonwealth of Independent States +7-502-935-8411, Czech Republic +420-2-3300-9311, Denmark +45 4556 5000, Egypt +202-570-9442, Estonia +372-6-308-900, Finland +358-9-525-561, France +33-134-03-00-00, Germany +49-89-46008-0, Greece +30-1-618-8111, Hungary +36-1-489-8900, Iceland +354-563-3010, India–Bangalore +91-80-2298989/2295454; New Delhi +91-11-6106000; Mumbai +91-22-697-8111, Ireland +353-1-8055-666, Israel +972-9-9710500, Italy +39-02-641511, Japan +81-3-5717-5000, Kazakhstan +7-3272-466774, Korea +822-2193-5114, Latvia +371-750-3700, Lithuania +370-729-8468, Luxembourg +352-49 11 33 1, Malaysia +603-21161888, Mexico +52-5-258-6100, The Netherlands +00-31-33-45-15-000, New Zealand–Auckland +64-9-976-6800; Wellington +64-4-462-0780, Norway +47 23 36 96 00, People's Republic of China–Beijing +86-10-6803-5588; Chengdu +86-28-619-9333; Guangzhou +86-20-8755-5900; Shanghai +86-21-6466-1228; Hong Kong +852-2202-6688, Poland +48-22-8747800, Portugal +351-21-4134000, Russia +7-502-935-8411, Saudi Arabia +9661 273 4567, Singapore +65-6438-1888, Slovak Republic +421-2-4342-94-85, South Africa +27 11 256-6300, Spain +34-91-596-9900, Sweden +46-8-631-10-00, Switzerland–German 41-1-908-90-00; French 41-22-999-0444, Taiwan +886-2-8732-9933, Thailand +662-344-6888, Turkey +90-212-335-22-00, United Arab Emirates +9714-3366333, United Kingdom +44 0 1252 420000, United States +1-800-555-9SUN OR +1-650-960-1300, Venezuela +58-2-905-3800, or Online at sun.com/store