THE SOLARIS™ OS AND INTEL® NEHALEM-EX

THE SOLARIS OPERATING SYSTEM WITH THE NEXT GENERATION INTEL® XEON® PROCESSOR, FOR THE EXPANDABLE SERVER SEGMENT (CODENAME NEHALEM-EX)

White Paper
September 2009

Table of Contents

Intended Audience

This white paper offers a high-level overview of enterprise capabilities of the Solaris™ Operating System (OS), and how it enables users to utilize the new features of the next generation Intel® Xeon® processor for the expandable server segment (code-name "Nehalem-EX"). This new processor is designed for mission- and business-critical applications and data in demanding enterprise environments. This document is intended mainly for business decision makers, developers and IT professionals—people who make systems, architecture, platforms, and IT decisions in medium and large companies.

Executive Summary

Deployed on powerful, highly reliable servers, the Solaris OS powers the applications that drive today's enterprise—serving thousands of global users without interruption. The Solaris OS is the strategic platform for today's demanding enterprise, with proven results running everything from mission-critical enterprise databases to high performance Web farms, and scaling from single-processing systems to large, multi-processor systems with terabytes of memory. Leveraging more than 20 years of SMP expertise, the Solaris OS is the critical nexus of applications and hardware resources—including processors, memory, networking, and file system—that maximizes throughput, reliability, and efficiency for enterprise computing.

Deploying mission critical and enterprise applications is now possible in cost-effective rack-mount and blade servers based on Intel Xeon processors. The next generation Intel Xeon processor provides greater computing power—processing cores and memory—reliability, and other mainframe-style features. Until now, this class of systems has been primarily limited to two- and four-processors—which have not been powerful, scalable, or reliable enough to deploy in these situations.

Designing powerful servers with significant processing capabilities that deliver enterprise and mission critical requirements demands unprecedented collaboration between the processor architecture and operating system environment. Solaris 10 features are specifically engineered for multi-core systems running multi-threaded applications. Sun Microsystems and Intel have been collaborating to optimize how the Solaris ecosystem and the next generation Intel Xeon processor work together, creating an intelligent and adaptive computing environment that includes:

- Scalable performance: The Solaris OS enables the Intel® Xeon multi-processor capabilities—up to eight processors, each with up to eight cores and two threads per core—as well as Intel® Hyper-Threading Technology, Intel® QuickPath Technology, and Intel® Turbo Boost Technology. The Solaris OS and the next generation Intel Xeon processor are designed to address enterprise IT applications.

- Advanced reliability: The Solaris Fault Management Architecture (FMA) integrates with the Intel® Machine Check Architecture (MCA) recovery features, enabling systems to automatically monitor, report, and recover from hardware errors to maintain data integrity and keep mission-critical applications and services online.
- Power efficiency and utilization: The Solaris OS takes advantage of performance-enhanced dynamic power management capabilities of the next generation Intel Xeon processor.

Sun and Intel, as part of a broad strategic alliance, have been working together—from design and architecture through implementation—to optimize the Solaris OS and unleash the power and capabilities of current and future Intel Xeon processors. Over the first two years of the strategic alliance, Sun and Intel made significant advances to optimize the Solaris OS for current Intel Xeon processor-based systems. Now, Sun and Intel are working closely to enable new capabilities that are part of the next generation Intel Xeon processor.

The Solaris ecosystem, including the Solaris OS, Sun Studio development tools, and the OpenSolaris™ community, will offer a compelling value for developers and users who embrace the breakthrough capabilities of next generation Intel Xeon processor. The result will deliver a robust choice for both established enterprise datacenter and leading edge applications.

Solaris™ Leadership

The Solaris ecosystem consists of the Solaris and OpenSolaris OS releases, as well as the Sun Studio development tools. The free and open Solaris OS is a proven, industry leading operating system with features designed to save time and money in business-critical operations. The Solaris OS provides stability, virtualization, massive scalability, high performance, and a business-oriented support lifecycle. Intel has embraced Solaris as a mainstream OS and the enterprise class, mission critical UNIX® OS for Intel Xeon processor-based servers.[1]

The Solaris OS includes many innovative features—some of the highest regarded include: ZFS™, Solaris Dynamic Tracing (DTrace), Predictive Self Healing, virtualization capabilities, power and performance optimizations, and more. The unique features of the Solaris OS are designed to deliver breakthrough enterprise functionality as well as flexibility in a wide range of business environments. Solaris advantages are not just on the technical side; it also provides operational benefits to businesses: unique among operating system vendors, Sun protects your IT investments by guaranteeing that existing Solaris applications will run unmodified on Solaris 10. Solaris 10 leverages the innovation of the OpenSolaris community, including significant investments by Intel and other partners. For example, The Sun Fire X4275 server

1. *sun.com/aboutsun/pr/2007-01/sunflash.20070122.1.xml*

powered by Intel Xeon X5570 processors and the Solaris OS set a new world record for SPECmail2009, designed to simulate real-world corporate e-mail environments.[2]

The OpenSolaris community is where the next generation of Solaris is being built, and where the latest innovations from Sun and Intel can usually be found first. The OpenSolaris OS release offers cutting-edge features contributed by a global development community, which provides accelerated time to market and support for the latest technologies and innovation in an environment familiar to GNU/Linux and UNIX developers and administrators. Solaris OS releases feature an extended support lifecycle as needed in today's demanding datacenters.[3]

The Solaris ecosystem, in conjunction with the next generation Intel Xeon processors, will be a superior choice for virtually all forms of enterprise computing, and leading edge applications such as Web 2.0 and high performance computing. Organizations can leverage Solaris as the mission-critical enterprise class operating system on systems based on next generation Intel Xeon processors. The Solaris OS is supported on over 1,200 systems, and OpenSolaris on over 4,500 systems.[4]

Next Generation Intel® Xeon® Processor (codename Nehalem-EX)

The next generation Intel Xeon processor is the next step in Intel's continuing success in leading the industry for enterprise computing. It delivers intelligent performance that automatically adapts to the diverse needs of a virtualized environment, including advanced reliability and scalablity. The next generation Intel Xeon processor continues Intel's philosophy of focusing on improvements in how the processor uses available clock cycles and power, rather than just increasing clock speeds, which raises energy requirements. In nearly every way, the next generation Intel Xeon processor delivers more capability—scalable performance, advanced reliability, and flexible virtualization—compared to previous-generation Intel® Xeon® processor 7400 series.

The scalable and modular processor architecture enables Intel to easily provide versions that are optimized for medium and large scale enterprise computing. An eight-socket system based on next generation Intel Xeon processor will have 64 processor cores and 128 threads, and supports terabytes of main memory—enterprise-class computing. When combined with the Solaris OS, it dynamically manages cores, threads, cache, interfaces, and power to deliver outstanding energy efficiency, performance, and scalability on demand.

2. *http://www.sun.com/servers/x64/x4275/performance.xml*
3. *sun.com/solaris/lifecycle.html*
4. *http://www.sun.com/bigadmin/hcl/*

Enterprise Computing: Solaris and Next Generation Intel Xeon Processor

With Intel and Sun collaborating on the Solaris OS for next generation Intel Xeon processors, powerful innovations in computing are at hand. Building on over two years of collaboration, the design goals of both the next generation Intel Xeon processor and Solaris OS have been closely aligned to provide an environment that delivers advanced functionality for enterprise applications, creating a platform for faster and more efficient application execution, and with business-class manageability and availability.

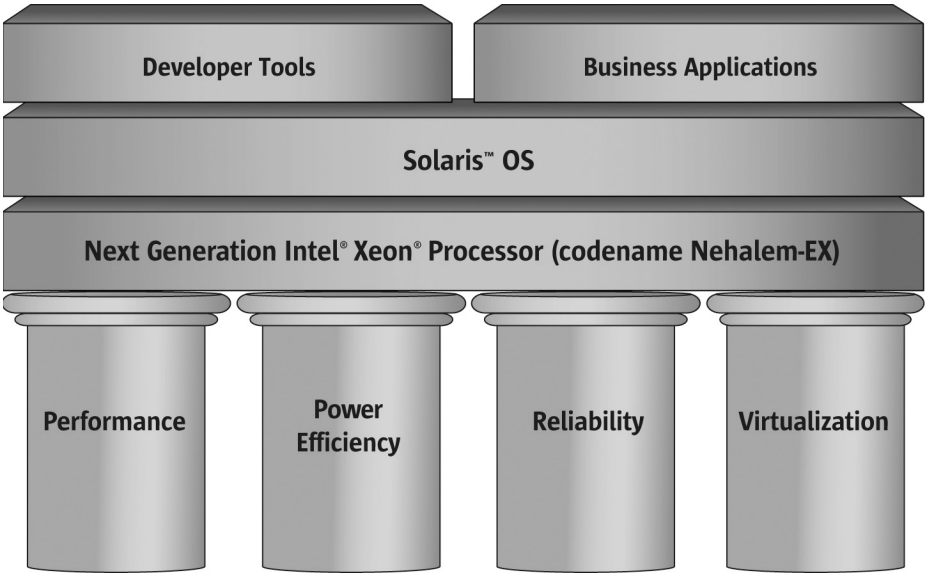

Figure 1: The Solaris ecosystem will be ready for systems built with the next generation Intel Xeon processor.

The following sections describe several of the new features and capabilities of next generation Intel Xeon processor, and how the Solaris OS enables mission-critical functionality for developers, administrators, and end-users alike.

Performance and Scalability

Building on a dozens[5] of world records, Solaris will be ready to take advantage of the groundbreaking performance capabilities of the next generation Intel Xeon processor. The most significant performance innovations come from the ability to scale well beyond 64 multi-threaded cores, and address large amounts of physical memory. In addition to other performance and scaling improvements, the Solaris OS and the next generation Intel Xeon processor will maximize enterprise application throughput.

5. *sun.com/x64/benchmarks.jsp*

The next generation Intel Xeon processor is able to take advantage of the proven Solaris scalability. Initial testing shows that the next generation Intel Xeon processor delivers more than 9x memory bandwidth, 2.5x database performance, 1.7x integer throughput, and 2.2x floating point throughput when compared to Xeon processor 7400 series.[6]

There are significant areas where Solaris will provide an environment for maximum application performance by taking advantage of the next generation Intel Xeon processor features.

- The next generation Intel Xeon processor features Intel Hyper-Threading Technology, which enables a more energy-efficient means of increasing performance for multi-threaded workloads. The next generation Intel Xeon processor will have two hardware threads per core, or up to 16 threads per 8-core processor—*up to 128 threads in eight-socket systems.*

  Solaris has an outstanding threading model for commercial workloads, outperforming the competition on customer applications as well as industry-standard benchmarks, running n production systems with dozens of sockets and multiple terabytes of memory. With specific optimizations for the next generation Intel Xeon processor, the Solaris OS enables new levels of performance as applications incorporate multi-threaded design, increasing throughput, responsiveness, efficiency, scalability, and overall performance.
- The Solaris OS will leverage the capabilities of Intel QuickPath Technology with capabilities such as an optimized scheduler and memory placement optimization (MPO) capability that have proven performance benefits with non-uniform memory access (NUMA) architecture systems. Intel QuickPath Technology connects processors, memory, I/O, and other components with up to four high-speed point-to-point interconnects, and enables the doubling of memory capacity with up to 16 memory slots per processor socket. Intel QuickPath Technology includes a new Intel® Scalable Memory Interconnect with Buffers and integrated memory controller (IMC). The IMC has the significant advantage of being coupled with large high-performance caches. This relieves pressure on the memory subsystem and lowers overall latency, resulting in dramatically improved throughput.
- Intel Turbo Boost Technology converts any available power headroom into higher frequencies. In those situations where the Solaris OS determines that maximum processing power is required, the next generation Intel Xeon processor increases the frequency in the active core when conditions such as load, power consumption and temperature permit it. By utilizing thermal and power headroom as a performance boost, the Solaris OS and the next generation Intel Xeon processor can deliver more work with reduced power consumption . The Solaris scheduler is “power aware” (see Power Efficiency, below) and works with the next generation Intel Xeon processor to maximize performance while minimizing power consumption.

6. *intel.com/pressroom/archive/releases/20090526comp.htm*

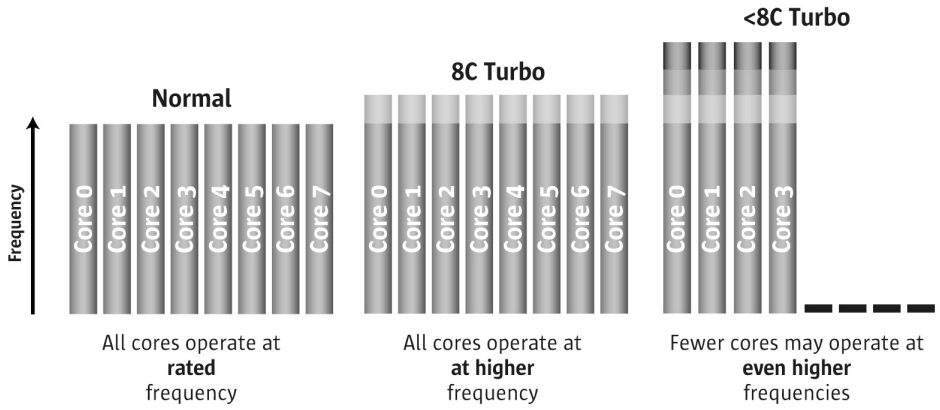

Figure 2: CPUs typically operate at a fixed maximum frequency regardless of the workload; however, most applications allow the CPU to operate below maximum power. Headroom may also be available if some cores are in idle mode, as pictured. Intel Turbo Boost Technology speeds up the CPU to utilize available power headroom, as needed.

- Intel® Smart Cache adds up to 24 MB of shared L3 (last-level) memory, improving speed and throughput in multi-threaded solutions. In addition to this cache being shared across all cores, the inclusive shared L3 cache can increase performance while reducing traffic to the processor cores. Multi-level shared cache improves performance and efficiency by reducing latency to frequently used data. Solaris is being optimized to take advantage of this capability in the next generation Intel Xeon processor.
- Innovative extensions to the Intel® Streaming SIMD Extensions 4 (SSE4.2)—and in particular its string and text processing instruction extensions—will enhance XML, string, and text processing, resulting in improved performance in areas such as Web 2.0 applications and virus scanning.

The Solaris OS offers additional performance advantages that help improve overall throughput. For example, the high-performance I/O acceleration features available in the next generation Intel Xeon processor is enabled by the turbocharged TCP/IP stack in the Solaris OS. Sun development tools will be integrated to take advantage of specific features in the next generation Intel Xeon processor, such as internal counters. The Solaris OS also includes a highly integrated facility for troubleshooting and tuning applications in real time. Solaris Dynamic Tracing (DTrace) technology provides very detailed tracing and observability, making it possible to detect performance bottlenecks in remarkably short periods of time.

Reliability and Availability

Increased scalability brings reliability to the forefront of any enterprise IT operation. The Solaris OS provides a proven architecture for building and deploying systems and services capable of *Predictive Self Healing*, which automatically diagnoses, isolates, and aids in recovery from hardware and application faults. Solaris Fault Manager is a key component of Predictive Self Healing.

Solaris Fault Manager receives data relating to hardware and software errors and automatically diagnoses the underlying problem. Once diagnosed, Solaris Fault Manager automatically responds by offlining faulty components. Sun and Intel are working together to extend these capabilities to systems based on the next generation Intel Xeon processor, including chipsets and memory subsystems.

The next generation Intel Xeon processor will add new, mainframe-inspired reliability, availability, and serviceability features such as Machine Check Architecture (MCA) recovery. Intel MCA recovery enables the system to detect and correct errors in memory and cache that were previously "uncorrectable" through ECC or other means. MCA accomplishes this by first detecting and containing errors before the data is consumed by an application, then works in conjunction with Solaris to determine the best course of action to keep the system and application running. This advanced recovery capability means that systems based on the next generation Intel Xeon processor will be able to recover and keep running in situations where other x86-based systems would not.

Sun and Intel are working together to use status information from the next generation Intel Xeon processor to help the Solaris Fault Manager diagnose a hardware fault correctly. This verification ensures that users running the Solaris OS on Intel Xeon processor-based systems will get a correct diagnosis and recovery should a hardware fault occur.

One critical element to enterprise-class availability is reliable data subsystems. Solaris ZFS provides unparalleled data integrity, capacity, performance, and manageability. Solaris ZFS provides high-resiliency features such as metadata logging to guarantee data integrity and speed recovery in the event of system failure. ZFS dramatically simplifies file system administration to help increase protection against administrative error. Solaris ZFS was introduced in the Solaris 10 6/06 Update, and is the default file system in the current OpenSolaris OS release and all future Solaris releases.

Power Efficiency

Intel® Intelligent Power Technology enables policy-based control that allows processors to operate at optimal frequency and power. The Solaris OS can make this determination automatically, or administrators can designate which applications require high-frequency processing and which should be executed at lower frequencies to conserve power.

Automated low-power states automatically put processor and memory into the lowest available power states that will meet the requirement of the current workload. Because Next generation Intel Xeon processors are enhanced with more and lower CPU power states, and the memory and I/O controllers have new power management features, the degree to which power can be minimized is now greatly enhanced.

Systems based on these processors can lower the voltage and frequency in steps without affecting overall throughput, and provide considerable energy savings at the same time. For the next generation Intel Xeon processor, Sun and Intel have worked to enhance power management capabilities in the Solaris OS, including:

- Adjusting processor power requirements in response to utilization, and enabling the system to go into the deepest power saving state possible when a processor is idle.
- Working to make the Solaris kernel "tickless," which means it will not periodically wake up to process a clock tick, but rather stay idle until an event of interest occurs. This can save considerable energy.
- Creating tools for developers that show where applications may be waking the system up unnecessarily and preventing it from going into a power-saving sleep mode. To identify applications or programming systems which may be responsible for polling behavior, the Intel and Sun engineering teams have developed the PowerTOP application for Solaris. The PowerTOP application points the developer to areas that consume high power within the application.

The Solaris kernel dispatcher—the part of the kernel that decides where threads should run—is integrated with the power management subsystem of the next generation Intel Xeon processor. The Solaris Power Aware Dispatcher (PAD) has increased awareness of the next generation Intel Xeon processor, allowing threads to be scheduled according to the power state of the processor. Workloads can be efficiently utilized on available hardware threads, with benefits for shared pipelines, shared caches, and shared sockets. The Solaris PAD is able to communicate what processor resources are being used by the operating system, and which are not. The Solaris kernel now has the ability to utilize those parts of the processor that are active, and continue to avoid doing work on those parts that are powered down. With support for deep idle CPU power management (deep C-states) in the next generation Intel Xeon processor, Solaris can dynamically place uninitialized cores into a state where they consume very little power. Any available power headroom can be applied to increase the performance of other cores through Intel Turbo Boost Technology.

Flexible Virtualization

Enterprise users need choice when it comes to server virtualization and consolidation, and flexibility with respect to application, OS, and network virtualization methods. As strong proponents of open source software, Intel and Sun collaborate with the Xen community to enable the Xen hypervisor to take advantage of the latest Intel® Virtualization Technology capabilities. The xVM hypervisor, part of OpenSolaris release today, offers comprehensive support for virtualization capabilities in the next generation Intel Xeon processor.

The next generation Intel Xeon processor includes next generation Intel Virtualization Technology (Intel® VT) that delivers even more efficient virtual machine operation with optimizations that enable a higher level of performance and reliability. These new features will enable improved software performance, security, and reliability in virtualized environments. The next generation Intel Xeon processor delivers faster virtualization performance by reducing transition latency into and out of virtual machine mode, as well a reducing the number of required transitions into and out of the virtual machine mode. In addition, the next generation Intel Xeon processor will offer improved I/O virtualization as part of the core logic chipset, and enhanced I/O performance through direct assignment of a device to a virtual machine. Fast, reliable, comprehensive, hardware-assisted virtualization capabilities contribute to better power efficiency, improved reliability, and increased asset utilization.

Many of the unique, built-in features of the Solaris OS contribute to the value of virtualization running on next generation Intel Xeon processors. Solaris offers a top-to-bottom engineered approach to virtualization where the hardware, the hypervisor, the OS, networking, and the ZFS file system are all designed to deliver optimal performance and manageability. In addition, Solaris offers features that help improve overall IT efficiency. For example, Solaris Containers safely isolate resources, allowing enterprise admins to maintain the one-application per-server deployment model when consolidating many single servers into a larger system based on the next generation Intel Xeon processor. Sun Ops Center software provides a unified management interface for both Solaris Containers and other virtualization technologies—a single management framework provides a view into managing and monitoring physical machines and the operating systems on both physical and virtual nodes.

Solaris Containers and the Solaris xVM hypervisor provide complementary virtualization solutions for heterogeneous environments. These solutions can be used in different environments, or even together in the same environment, depending on the given workloads and operating systems. They provide flexibility and choice in virtualization technologies.

Network Virtualization

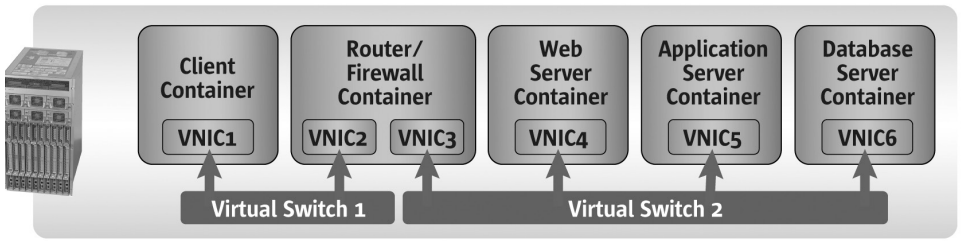

Figure 3: OpenSolaris Project Crossbow provides high-performance network virtualization capabilities, for comprehensive management of infrastructure resources.

OpenSolaris 2009.06 incorporates Project Crossbow, a multi-year engineering project to rearchitect the network stack, particularly the data link layer, in order to accomplish three goals.

- Enable server consolidation projects involving entire network topologies at vastly superior economics when compared to building out physical network topologies. Crossbow enables network-in-a-box functionality supporting virtual NICs, switches, routing and firewalls.
- Allow organizations to specify and meet quality of service goals for networking through management capabilities such as setting bandwidth limits on physical and virtual network interfaces, supporting traffic priorities, and assignment of CPU resource limits for servicing the network interfaces.
- Improve system networking performance by taking advantage of the capabilities built into Intel NICs.

Project Crossbow takes advantage of the capabilities of Intel networking chips to deliver the above new functionality without additional overhead. For example, in a network resource management scenario, the traditional way to add Quality of Service enforcement is to create another layer of software in the stack, which creates problems with higher-speed networking. Instead Crossbow relies on the capabilities of Intel network interface controller chips for incorporation into the overall stack architecture. In the traditional approach, the OS brings in a packet and then, depending on a variety of factors, makes the decision on whether to drop or process it. Unfortunately, just by bringing that packet into the OS means the system incurs a significant performance cost to process it—upwards of 60% of the total cost in CPU cycles. With newer NICs like the Intel® 82598EB ("Oplin") and Intel® 82599EB ("Niantic") 10 Gigabit Ethernet Controllers, the transmit and receive ring buffers become the point of management.

For example, consider the scenario where a Virtual NIC (VNIC) is set up and configured with a bandwidth limit less than the physical NIC can support. The OS will calculate how many packets should be picked up in a given time frame, which sets the limit for how many are removed from chip's receiver buffer. If the bandwidth on a particular virtual NIC is set low relative to the physical bandwidth capabilities, when traffic to this VNIC is high, packets will be dropped; however, the OS does not have to be involved. They are dropped at the chip level as the ring buffer fills to capacity.

Taking this example one step further, suppose the system encounters a denial of service attack. Crossbow can take advantage of another attribute of Intel networking chips—the availability of multiple transmit and receiver buffers. This enables the placement of network traffic into a dedicated lane from the moment it arrives at the chip. The chip does its own packet inspection as programmed by the OS, which can be based on a variety of factors, source IP, destination IP, type of traffic (HTTPS, HTTP, FTP, and so on), MAC address, and more. The code paths from the chip all the way to the application are independently based on this 'lane' architecture and enforced by classification at the chip level. Once the OS starts processing the packet there are no common queues, no common threads, no common locks, and no common counters as the packet (figuratively speaking) moves through the stack. In a denial of service example, once the attackers are identified by an IP or IP subnet address, it is possible to dial down the bandwidth on those lanes so that packets get dropped at the chip, and never adversely impact the OS.

Crossbow will be able to offer the same capabilities for a wide range of networking chips and when the chip does not support multiple buffers, this service must be supplied by the OS. Clearly when these services are provided by Intel's chips, Crossbow can extract the best performance and still deliver important new functionality like network resource management.

Conclusion

Solaris has continued to demonstrate great success as a mission-critical, enterprise-class OS for scalable performance, advanced reliability, and power efficiency in the data center. The upcoming introduction of the next generation Intel Xeon processor is creating an even greater demand for Solaris in the x86 volume server segment. Solaris is leveraging more than 20 years SMP expertise for proven performance in a very large multi-core processing environment. The next generation Intel Xeon processor is well positioned to take advantage of the Solaris 10 OS, with features specifically engineered for multi-core systems.

Intel Xeon processors and the Solaris OS are both widely recognized as the technologies of choice for enterprise and mission-critical applications. Whether serving large databases, high performance computing applications, or consolidating multiple lower-powered servers, systems must scale smoothly and intelligently.

The collaboration of Sun and Intel, including the joint engineering work from both companies, has already resulted in state-of-the-art development and deployment platforms for environments based on Intel Xeon processors. The next step is focusing on delivering the Solaris OS optimized for the next generation Intel Xeon processor—maximizing performance, reliability, virtualization, and power efficiency. The result will be an ideal platform for mission- and business-critical deployment.

More Information

Intel

intel.com/sunalliance/
intel.com/technology/architecture-silicon/next-gen/

Sun

Solaris OS:
sun.com/solaris/

OpenSolaris OS:
opensolaris.org, opensolaris.com

Project Crossbow:
opensolaris.com/learn/features/networking/networkcrossbow/index.html

Sun and Intel Alliance:
sun.com/solaris/intel

Sun Microsystems, Inc. 4150 Network Circle, Santa Clara, CA 95054 USA Phone 1-650-960-1300 or 1-800-555-9SUN (9786) Web sun.com

© 2009 Sun Microsystems, Inc. All rights reserved. Sun, Sun Microsystems, the Sun logo, Solaris, OpenSolaris, ZFS and "The Network is the Computer" are trademarks or registered trademarks of Sun Microsystems, Inc. or its subsidiaries in the United States and other countries.

Intel® and Xeon® are trademarks or registered trademark of Intel Corporation or its subsidiaries in the United States and other countries. All SPARC trademarks are used under license and are trademarks or registered trademarks of SPARC International, Inc. in the US and other countries. Products bearing SPARC trademarks are based upon an architecture developed by Sun Microsystems, Inc. Information subject to change without notice.

Printed in USA 09/2009