TECHNOLOGY WHITE PAPER

NEI / 25 Dan Road Canton, MA / 02021-2817 / telephone: 781 332 1000 / fax: 781 770 2000 www.nei.com

Transitioning to the Intel® Xeon® 5500 (Nehalem) Series Microarchitecture

A Guide to Successfully Implementing the Industry's Next-Generation Server Processors

Introduction

Intel's new Xeon 5500 series of processors, based on the Nehalem microarchitecture, is creating genuine excitement in the computing world and with application developers deployed on x86-based hardware platforms. Is your company's software development team prepared? Are they properly planning to unleash the power of these new devices? This paper provides a guide to the critical thinking that software developers, OEMs and service providers should undertake to ensure success.

The Core of the Issue

The Xeon 5500 series is composed of about a dozen new multi-core, quad-core 45-nanometer processors designed for dual-socket servers and workstations.

The Nehalem microarchitecture offers up to 2.5 times more memory bandwidth than previous generation Intel architecture platforms, as well as the ability to complete more work per clock cycle. Faster processing provides the means to an end for many latency dependency applications; and the ability to access memory faster solves problems related to the growing trend for real-time encoded data processing. Many applications can therefore improve performance two times over by optimizing their application for the Nehalem architecture.

Beyond improved processor performance, memory capacity and network throughput, other advantages come into play. These quad-core chips boast as much as twice the energy efficiency benefits of their predecessors and they enable more "work per Watt." Better performance for the same power budget means as much as 20 to 50 percent more work may be done at the same power consumption level, depending on the application.

The general consensus among industry observers is that software developers should plan now to transition to the new technology. This is particularly true for security, enterprise communications, telecommunications and storage applications – any type of solution involving transcoding or decryption, including most video-related applications, VoIP and secure endpoint communications.

Lets take a closer look at the benefits of the new microprocessor, and offer answers to six commonly-asked questions about making the move to this architecture.

Common Questions, Common-Sense Answers

The transition to the Nehalem architecture and Xeon 5500 processing is not unlike other major new silicon releases – that is, it's not a overly complicated process. But like others throughout time, there are technical challenges and ways to foresee and overcome potential roadblocks. Start by looking at your company's product and service offering(s) in the context of some commonly asked implementation questions.

- **Can my application gain by moving from proprietary customized hardware to a more COTS-oriented solution?**

  The need to use dedicated (often proprietary) I/O accelerator boards should be reconsidered. In wire-speed communications and deep packet inspection/security applications, for example, DSP-based boards were extremely common for voice processing and accelerating the encryption and decryption process for real- or near-real-time response.

  In such cases, most of the work cycles went into decrypting the input data and encrypting it again before output to ensure a secure connection and assign data path instructions. As Intel's microarchitectures became faster and more efficient, they offered gradual enhancements for those types of functions. Applications that had required numerous DSP boards only five years ago can run on a single dual-processor system today at a much lower cost. Using the Xeon 5500, more applications may no longer require a DSP solution; in fact, it may be sufficient to multi-thread high-load functions across Xeon cores with or without some amount of DSP emulation software.

- **Do I need to change my source code?**

  Many software developers are constrained by their companies from making substantive changes to their application source code. Businesses that already have deployed devices in the field may be concerned that optimizing for this new architecture will make the coding incompatible with the previous architecture.

  With no real changes to your coding, you will likely see an immediate increase in performance when transitioning to Nehalem. Of course, any time you switch to a new piece of hardware, you should perform regression testing to ensure system stability.

  Additionally, in rewriting software for a new processing core, it is important to understand how the processor's clock/instruction cycles and processing speeds may affect code timing and interaction with peripheral (off-chip) jobs. This is common in applications where events

Key Features of Xeon 5500

The new Xeon 5500 series microprocessors offer many distinct advantages for product designers and developers. Here is a glimpse of some of the most compelling features:

- **PCI Express Gen 2**

  Backwards compatible to generation 1, this version of PCI Express offers 5 GHz signaling versus 2.5 GHz in the first generation products (resulting in a throughput capacity improvement of 500 MB/s per lane versus 250 MB/s).

- **Integrated Memory Controller on CPU**

  Supporting up to 1333 MHz DDR3 memory, Nehalem accommodates three memory channels per CPU, for 2.5 x the memory bandwidth over previous memory technology -- ideal for higher performance for data-intensive applications.

- **QPI (QuickPath Interconnect)**

  With the elimination of front side bus (FSB) connectivity, the architecture allows for direct serial interconnect between devices.

- **Hyperthreading**

  A dramatic improvement to parallel computations on PC microprocessors, each physical core represents two logical cores. The operating system sees two virtual processors, and shares the workload between them.

- **Improved IOAT**

  I/O Acceleration Technology allows for faster networking and virtualization through hardware acceleration. It minimizes the bottlenecks that can result from the large data flows of virtualized workloads and multi-port Gigabit Ethernet and 10 Gigabit Ethernet network adapters.

- **Dynamic Power Technology (DPT)**

  The server can throttle power as required to stay under predetermined limits.

- **Enhanced reliability and manageability**

  Many memory controller features combine to help improve platform reliability versus previous-generation platforms. New features include Error Correcting Code (ECC) system bus, new memory mirroring and I/O hot-plug.

need to occur in a particular sequence (where "A" must occur before "B" can run, for example), code timing is often addressed by building in cycle wait time. That wait time may need to change in a new processing environment where more work can be done faster.

On the whole, however, if your base code works on the previous generation architecture, it should work on the new architecture with the added benefit of increased performance. More importantly, if you specifically optimize your coding around the Nehalem architecture, those benefits and performance gains will be even greater.

- **Is it time to refresh my OS distribution?**

Some developers have expressed concern about driver compatibility with the introduction of the new Xeon processor. For those who use a standard Linux distribution, driver compatibility will not be a big issue – no matter the hardware architecture, there are almost always new drivers to supplement the base image and run the hardware.

If you are maintaining a custom Linux design, you need to look carefully at the security patches and drivers back into your environment. Can you get into the new kernels without sacrificing performance? Re-writing new drivers into custom kernels can cost literally hundreds of thousands of dollars. Is the maintenance of a custom Linux design worth continuing as Nehalem takes hold in hardware systems?

For customers who may be deploying on an older operating system (OS), it is important to consider whether it is time to update to the most current Linux or Windows distribution, for example. The newer OS distributions will have the tools needed to take advantage of multiple core supervision, as well as hyper-threading, memory management and virtualization improvements.

It's typically best to make this type of OS migration at the same time as a hardware migration, since a complete requalification must be conducted and tested. Therefore, planning OS migration should be closely considered when planning to release a new application.

- **How can I plan ahead for virtualization?**

As the Xeon 5500 series platform rolls out into more data centers, the demand to run more applications virtually will also increase. Most developers should already be addressing the likelihood of having to deliver a virtual solution in the next two to four years. With Xeon processing power aboard, perhaps your virtualization roadmap and/or schedule changes.

The Xeon 5500 microarchitecture is enhanced to improve the capabilities and performance of virtual machines, allowing more virtual machines to run more efficiently. Intel Virtualization Technology (VT) combined with software-based virtualization solutions can consolidate multiple environments into a single server or PC. Intel enables the integration of multiple generations of Intel Xeon processor-based servers, improving flexibility for failover, load balancing and disaster recovery.

By virtualizing different generations of Intel Xeon processor-based servers in the same pool, the architecture is able to deliver peak performance during high-use periods and automatically reduce energy cost during low-use periods for the best possible system utilization.

In small to medium markets, the benefits of virtualization are particularly important. Consider security, for example. A small-office solution today may include several security applications, local processing and some local storage. These could all be consolidated onto a single Nehalem-based server running in a virtual environment.

For larger enterprises, developers will need to consider separation of storage from processing in order to create new virtual machines and provide adequate load-sharing among these virtual devices.

By accommodating virtualization in your planning now, your organization will be able to develop solutions that allow your customers to reduce costs, increase management efficiency, strengthen security, and enjoy a more robust, resilient network operation in the event of a disaster.

- **What is the benefit of "Turbo Boost" mode?**

Some solutions require either many applications or a few well-threaded applications to run simultaneously. The dual processor 5500 series architecture is designed to simultaneously accommodate as many as 16 threads. Some applications don't thread very well, however, and for that reason the new architecture includes a "Turbo Boost" mode.

Intel's Turbo Boost Technology increases processor frequency and enables faster speeds when conditions allow. In Turbo mode, cores that are not needed are essentially "throttled back" into an idling state, enabling the remaining cores to run faster.

Customers whose applications do not take advantage of multiple parallel processes can benefit particularly well from Turbo Boost mode and still achieve new levels of overall performance.

- **Can reduced energy consumption make my application more competitive?**

Because servers spend much of their time in mid-to-low utilization states, significant enhancements were made on the Xeon 5500 technology to improve efficiency from mid-utilitization down to idle levels. Intel's Intelligent Power Technology lowers energy costs without hurting performance by automatically putting processor and memory into the lowest available power state adequate for the current workload.

Additionally, the Xeon server processors will automatically reduce frequencies on a core-by-core utilization basis, thereby minimizing the overall power consumption and enabling greener solutions with no degradation of performance.

Power and thermals characteristics go hand in hand and heat dissipation can be a key differentiator. Each watt of power consumed by a server platform and converted into heat can take as much as 1.2W of additional power to remove the heat from the server environment. Waste heat must be removed from the data center to ensure the reliable operation of network devices, which requires energy to operate air conditioning systems and keep a constant temperature in the server room.

Most IT departments provision power for peak load characteristics, but systems are only at peak load times for a relatively small percentage of each day. The Nehalem architecture enables companies to use less energy to eliminate waste heat. That translates into reduced operating costs and more environmentally responsible network operations.

Conclusion

Regardless of your application, it may be important to begin deploying the innovative Xeon 5500 series processors as quickly as your planning process allows.

In low-latency applications, Nehalem architecture enables the CPU to run more quickly, with equally robust memory performance, virtually eliminating bottlenecks.

The new architecture is well suited to memory intensive database applications, such as billing or authentication for telecommunications, which can require thousands of verifications per second. It is also extremely useful for wire speed security analysis applications that require fast analysis, and for unified threat management, where success is directly proportional to the system's performance.

For Advanced Telecom Computing Architecture (ATCA) solutions, the Xeon 5500 series processor enables much greater memory access without increasing power consumption. This allows companies to increase capacity from 16 GB to 64 GB on a single blade.

Beyond the improvement in performance, however, companies incorporating Xeon's architecture into their product development plans stand to improve their market share by being part of the leading edge adopters of this new technology.

Switching to the Xeon 5500 series early will allow for increased product offerings with more flexibility and scalability. The ancillary benefit of reduced energy consumption will allow companies with these solutions to promote economically and environmentally responsible benefits.

NEI's Xeon 5500 Products

NEI's Xeon 5500-based platforms currently include the E-1800 R2 and E-2900 R2 medium density enterprise-class communication appliances. Later in 2009, NEI will launch the S-1900 R2, C-2500 R2 and E-4000 R4 dual processors and the S-1200 R2 and S-1450 single processor platforms. NEI can provide custom engineering for different form factors, power, and third-party add-on cards. And NEI's OEM relationships with Dell, HP, Intel and Sun Microsystems provide Tier 1 options for the widest range of compatible technologies. Contact NEI sales for complete details.

About NEI

NEI is a leading provider of application platforms, appliances and services for software developers, OEMs and service providers worldwide. NEI enables customers to more effectively deploy, manage and support application platforms and appliances using its comprehensive capabilities, including solution design, integration control, global logistics, smart services, technical support and maintenance. Founded in 1997, NEI is headquartered in Canton, Massachusetts and trades on the NASDAQ exchange under the symbol NENG. For more information about NEI's products and services, visit **www.nei.com**.

Contact Us

For more information on the full range of appliance-related hardware, software and services NEI can provide to your company, please contact us by telephone +1 (781) 332-1295 or by email at **sales@nei.com**.