

5 ways Intel® Xeon® Scalable processors with built-in accelerators can solve your most rigorous workload challenges

Workloads are evolving — and so is computer architecture. Traditionally, adding more cores to your CPU or choosing a higher-frequency CPU would improve workload efficiency. But these techniques no longer guarantee the performance efficiency benefits that they achieved in the past. Today, integrated accelerators purpose-built for particular functions in AI, security, HPC, networking, storage and data analytics can deliver considerably more value.

Current and future Intel® Xeon® Scalable processors support the broadest and most unique set of built-in hardware accelerators for modern workloads in both cloud and enterprise deployments. Whether you're looking to improve performance, reduce costs or increase energy efficiency, **here are five ways Intel Xeon Scalable processors with built-in accelerators can help your business solve the most rigorous workload challenges.**

01

Reduced need for additional hardware

More hardware means higher system costs. It can also mean leaving businesses more vulnerable to inefficiencies due to potential bottlenecks when adding devices. Intel Xeon Scalable processors have accelerators built into the CPU to give users the performance they need right out of the box. This spares businesses from having to purchase and integrate additional hardware, which can be costly.

While certain workload requirements warrant additional specialized hardware, in many cases our built-in accelerators will enable you to run your workloads effectively as is.

For instance, AI accelerators built into Intel Xeon Scalable processors are designed to run complex workloads — such as AI training and inferencing, plus many classical machine learning applications — on the same hardware as other critical tasks. Intel Xeon Scalable processors are already optimized for the most popular AI frameworks used by data scientists, such as TensorFlow and PyTorch.



02

Faster workload completion compared to AMD CPUs

Intel has worked directly with the most popular AI tools, frameworks and solutions to optimize their performance on Intel products. This helps enterprises achieve the best training and inference from a CPU. Intel Xeon Scalable processors feature Intel® Deep Learning Boost (Intel® DL Boost), a built-in accelerator designed to improve performance for common AI workloads.

Customers using Intel® optimization for TensorFlow and DL Boost will gain over 11 times more AI inference performance on 3rd Gen Intel Xeon Scalable processors compared to 2nd Gen Intel Xeon Scalable processors.¹ The upcoming launch of 4th Gen Intel Xeon Scalable processors with Intel® Advanced Matrix Extensions (Intel® AMX) will deliver 4.5 times INT8 image inference per second compared to the prior generation.²

Compared to AMD Milan FP32, 3rd Gen Intel Xeon Scalable processors supporting Intel DL Boost deliver 1.5 times higher AI performance.³



11x

more AI inference performance compared to 2nd Gen Intel Xeon Scalable processors



4.5x

INT8 image inference per second compared to the prior generation



1.5x

higher AI performance

03

Increased energy efficiency

Because Intel built-in accelerators help maximize energy efficiency, you can increase your performance across various workloads without adding discrete accelerators to your server rack.

This benefit was recently highlighted by Patrick Kennedy from ServeTheHome, an online guide for IT professionals. While running AI workloads on Intel DL Boost, he noted, "We saw relatively similar power consumption, but we also saw a huge delta in terms of performance. The net impact was a huge efficiency delta."⁴

“We saw relatively similar power consumption, but we also saw a huge delta in terms of performance. The net impact was a huge efficiency delta.”

Patrick Kennedy
ServeTheHome

04

High performance maintained while protecting your most sensitive data

Intel Xeon Scalable processors enable confidential computing solutions that better protect your data on prem, at the edge and in the cloud. Intel® Software Guard Extensions (Intel® SGX), a built-in accelerator for security, is the most deployed, researched and battle-tested confidential computing technology in data centers on the market today.

Intel SGX helps protect sensitive data and application code while it's actively in use. This helps defend against breaches, leaks or attacks that could halt business operations, compromise critical data or break compliance. Confidential computing technologies with larger attack surfaces put sensitive data at greater risk. Intel SGX provides the smallest attack surface within a system.⁵

Additionally, Intel® Crypto Acceleration — an instruction set in the Xeon core architecture — uses single instruction, multiple data (SIMD) techniques to process more encryption operations in every clock cycle, making things go faster. This can increase the total throughput of applications that require strong data encryption, with minimal impact on performance and user experience.



More processing capacity and bandwidth delivered for your HPC workloads

The need for greater computing performance in business, science and academia has never been higher. Intel has helped organizations design system architectures that execute the most demanding of workloads. Whether these organizations are trying to crack the biggest challenges in medicine, economics or engineering, the HPC accelerators built into Intel Xeon Scalable processors will increase workload performance to conquer highly advanced computational tasks — and do so faster than before.

For data-intensive workloads such as modeling and simulation, Intel Xeon Scalable processors not only allow code to take advantage of Intel® Advanced Vector Extensions 512 (Intel® AVX-512), but also offer high system memory capacity and bandwidth. This helps accelerate complex workloads on existing hardware. For example, 3rd Gen Intel® Xeon® Platinum 8358 CPUs using Intel® AVX-512 perform 23% better when compared to AMD's EPYC 7543 processors.⁶



Designed and optimized for specific market segments, built-in accelerators for Intel Xeon Scalable processors enable high-performance, hardware-based workload acceleration with excellent cost and power efficiency.

The 4th Gen Intel Xeon Scalable processors will feature built-in accelerators that include Intel® Advanced Matrix Extensions (Intel® AMX), Intel® QuickAssist Technology (Intel® QAT) and Intel® Data Streaming Accelerator (Intel® DSA). These will continue to boost performance and results, giving customers the capabilities they need to be faster, more sustainable and more successful than before.

Learn more about Intel Xeon Scalable processors by visiting <https://www.intel.com/xeonscalable>.

05

¹See [118] at <https://www.intel.com/3gen-xeon-config>
²See Session Benchmark #41 and #42 at <https://edc.intel.com/content/www/us/en/products/performance/benchmarks/vision-2022/>. Results may vary.
³See [11] at <https://www.intel.com/3gen-xeon-config>
⁴Kennedy, Patrick. "Deep Dive into Lowering Server Power Consumption." Serve The Home. February 21, 2022. <https://www.servethehome.com/deep-dive-into-lowering-server-power-consumption-intel-inspur-hpe-dell-emc/>
⁵See <https://www.intel.com/content/www/us/en/architecture-and-technology/software-guard-extensions-enhanced-data-protection.html>
⁶See [104] at <https://www.intel.com/3gen-xeon-config>

Notices & Disclaimers

Performance varies by use, configuration and other factors. Learn more on the [Performance Index site](#).

Performance results are based on testing as of dates shown in configurations and may not reflect all publicly available updates. See backup for configuration details. No product or component can be absolutely secure.

Your costs and results may vary.

Intel technologies may require enabled hardware, software or service activation.

© Intel Corporation. Intel, the Intel logo, and other Intel marks are trademarks of Intel Corporation or its subsidiaries. Other names and brands may be claimed as the property of others.

Intel does not control or audit third-party data. You should consult other sources to evaluate accuracy.