

2nd Generation Intel[®] Xeon[®] Scalable Processors Maintain Leadership Performance for Today's AI Infused HPC Platforms

2nd Gen Intel Xeon Scalable processors offer outstanding real-world application performance for a new era of HPC in which traditional HPC workloads are enhanced with new analytical workloads including deep learning.

New Generation of HPC Performance Leadership: Intel Xeon Platinum 9282 processors deliver an average of **31% higher performance** than AMD EPYC "Rome" 7742¹



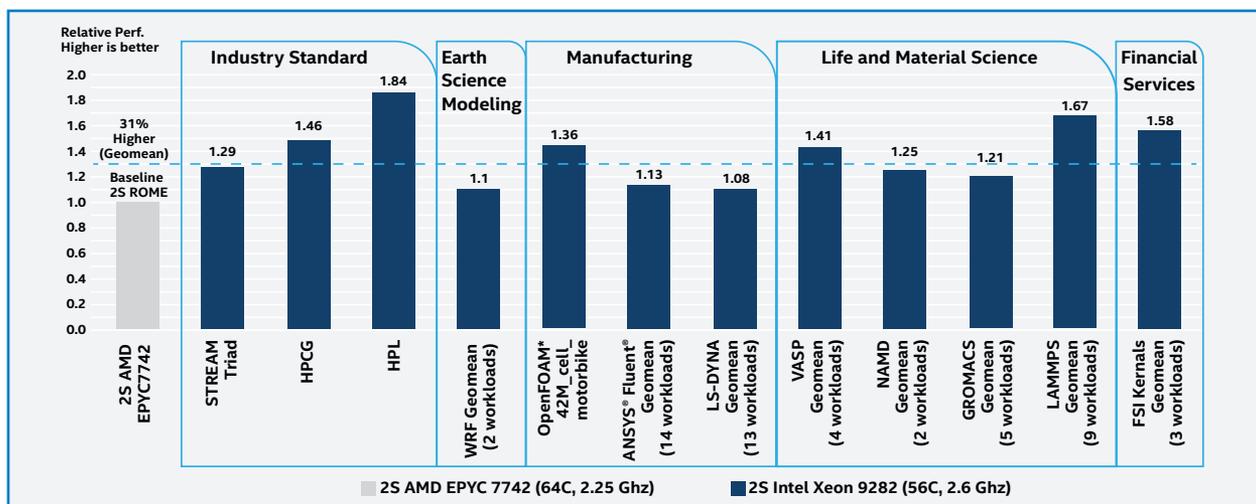
Executive Summary

Systems based on the 2nd Gen Intel Xeon Scalable processor are enhanced with *outstanding* processing performance coupled with *outstanding* memory bandwidth per core, per node, and at scale to serve a broad set of real-world HPC, data analytics, and AI applications. This powerful platform will increase the performance of systems ranging from the smallest high-performance computing (HPC) clusters to the world's largest supercomputers. Support of Intel Optane™ DC persistent memory allows substitution of some, or all, DDR memory to solve larger problems at lower cost.

Great scale, performance, and more advanced HPC capabilities are unlocked by combining the latest processors, Intel Optane DC persistent memory, Intel Optane DC SSDs, optimized libraries, frameworks, software tools, and platform specifications.

Providing for the convergence of AI and HPC, the 2nd Gen Intel Xeon Scalable is the only platform with HPC *and* AI acceleration built in. The 2nd Gen Intel Xeon Scalable processor delivers world-class performance capability for a broad range of HPC and AI applications enabled by technologies such as Intel AVX-512, Intel Deep Learning Boost (Intel DL Boost) for built-in inference acceleration, a balanced compute I/O and memory, and architecture delivering predictable latency.

HPC Performance Leadership



31% Higher Performance with 2S Intel Xeon-AP vs 2S AMD EPYC "Rome" 7742¹

*This offering is not approved or endorsed by OpenCFD Limited, producer and distributor of the OpenFOAM software via www.openfoam.com, and owner of the OPENFOAM® and OpenCFD® trademark.

Addressing HPC Challenges

HPC delivers significant value to a broad range of organizations by accelerating discovery and innovation. Whether it is scientific research such as the study of the cosmos, discovery of new materials, and personalized medicine for cancer, or business transformation enabled by reducing cost and accelerating time to market for new innovations, organizations rely on HPC systems to maintain their competitive edge.

Traditional HPC workloads (e.g., workloads that use simulation and modeling) across sciences, academia and industry have historically each placed unique demands on an HPC system. The convergence of traditional HPC, data analytics, and AI—across both infrastructure and workloads—is further driving increased demands on the various components of an HPC system. Systems built around the 2nd Gen Intel Xeon Scalable processor not only offer world-class performance for traditional HPC workloads, but also offer uniquely compelling solutions for integrating data analytics and AI capabilities in a single HPC system.

The high performance of Intel AVX-512, including dual fused multiply-adders (FMAs), supports the most demanding traditional HPC workloads, while the addition of Intel DL Boost accelerates some algorithms including deep learning. Together, these capabilities offer unprecedented agility, flexibility, and performance for HPC and HPC/AI converged workloads now and into the future.

In the past few years, prior software gaps caused by uneven optimization efforts have largely disappeared thanks to many enhancements in popular deep learning frameworks and the libraries that support them. The 2nd Gen Intel Xeon Platinum 9282 processor can achieve up to 30x inference throughput over the first generation processors when it was launched.⁴

The 2nd Gen Intel Xeon Scalable processor, with the addition of Intel DL Boost is especially useful for deep learning when computational capacity is an issue. This is a particular win for HPC/AI converged applications where large models, such as those using medical images, benefit from direct access to the memory capacity of the processor.

Intel® Xeon® Platinum 9200 processors included in the Intel® Server System S9200WK deliver:

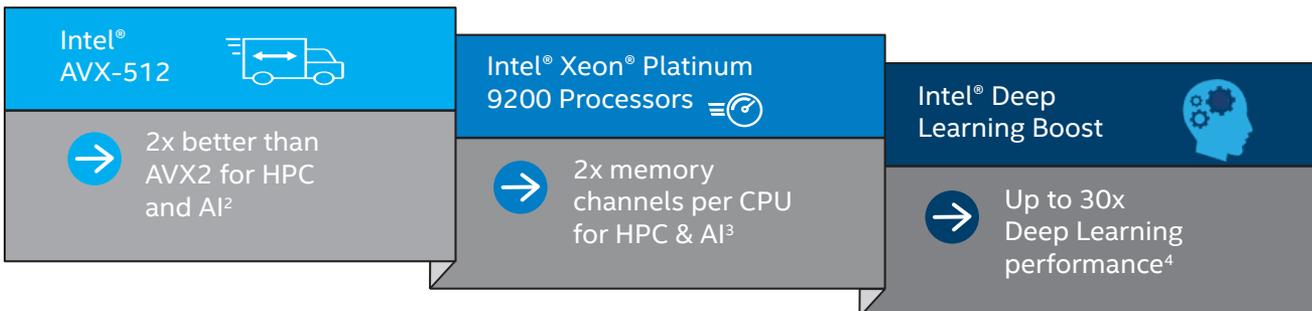
Amazing Performance in a Density Optimized Solution Ideal for HPC and AI



Designed for HPC, advanced analytics, AI, and high-density infrastructures, Intel Xeon Platinum 9200 processors feature up to 56 cores and 12 memory channels to deliver breakthrough levels of performance with high Intel architecture FLOPS per rack, and the highest DDR4 native memory bandwidth support of any Intel Xeon processor platform. Learn more at <https://www.intel.com/content/www/us/en/products/docs/servers/server-boards/server-board-s9200wk-brief.html>

- Average 31% performance advantage for the Intel Xeon Platinum 9282 CPU vs. AMD "Rome".¹
- Optimized for density and performance
- 30x Deep Learning performance with Intel® DL Boost vs prior generation when launched.⁴
- Double the memory bandwidth³ for memory-intensive workloads with 12 memory channels per CPU
- 24 memory channels per compute module

Intel Generational Advancements



<p>Intel® Optane™ DC Persistent Memory for HPC</p>	<p>Lower Cost More Memory</p> <p>Use Intel® Optane™ DC persistent memory over DDR to solve larger problems at lower cost; Discover new use cases with Intel® Optane™ technology</p>
---	--

Systems built with the 2nd Gen Intel Xeon Scalable processor not only offer performance for a broad range of HPC workloads, they also offer uniquely compelling solutions for integrating data analytics and AI in a single system.

Latest Intel Technologies Accelerate Today's Supercomputers—and Tomorrow's

Today's HPC systems must balance the needs for higher performance and cost efficiency to meet the growing demands of organizations.

Innovative technologies in compute, memory, fabric, storage, and system software break through the performance, memory and I/O barriers of the past to unleash HPC innovation.

The 2nd Gen Intel Xeon Scalable processor offers platform improvements enhancing each of the following elements:

- **Extensive configurability:** comprehensive SKU support delivers choice from 1 socket to 8+ sockets to match application and deployment needs.
- **Versatile high throughput core design** with outstanding performance per core and more predictable latency thanks to an enhanced internal mesh/memory architecture.
- **Intel® Advanced Vector Extensions** (Intel AVX-512) boost performance for the most demanding computational workloads, with up to 2x the FLOP/s compared with AVX2.²
- **Intel Deep Learning Boost** (Intel DL Boost) extends Intel AVX-512 to collectively offer high-performance data type-support for HPC, data analytics, AI, and applications that utilize all three modalities.
- **Density and performance** come together in the Intel Xeon Platinum 9200 processors for the most demanding workloads with up to 56 cores, incredible memory bandwidth with 12 DDR4 memory channels per CPU at 2933MT/s, and high-speed interconnect capabilities via 80 PCIe Gen3 lanes per node.
- **Visualization** of very large data sets is an important part of many HPC workflows. Software Defined Visualization (SDVis) is a popular open source initiative that has helped to improve the visual fidelity, performance and efficiency of prominent visualization solutions. This benefits the rapidly growing Big Data usage in HPC without the memory limitations and additional cost of GPU based solutions.

Agile HPC Building Blocks

Cluster architectures may also draw on a variety of other Intel architecture building blocks, including:

- **Intel Optane DC Persistent Memory** is an innovative memory technology that delivers a unique combination of affordable large capacity, performance, and persistence (non-volatility). Supported by most of the 2nd Gen Intel Xeon Scalable processors, this technology bridges the gap between DDR and SSDs. By revolutionizing the memory-

storage hierarchy, it enables massive data sets to be stored closer to the CPU for faster time to insight, enables larger working sets to deliver higher resolution simulations, enhances performance for latency-sensitive workloads, and allows for more frequent local checkpointing capability.

- **Intel SSDs** provide a range of storage options that enable customers to create their own balance between cost and performance. Intel Optane technology delivers differentiated performance, while Intel 3D NAND Technology offers higher capacity while still providing very high performance.
- **Intel Interconnect products** include Intel Omni-Path Architecture, Intel's next-generation HPC fabric that supports up to 100 Gbps, as well as Intel Ethernet, an established industry leader with a broad array of options for speed, cable medium, and port count. Both are built to work smoothly in clusters with other fabrics such as InfiniBand.
- **Intel FPGA** solutions offer another useful complement to the 2nd Gen Intel Xeon Scalable processor family, providing world-class energy efficiency, lowest latencies for inferring, and outstanding reconfigurability for evolving HPC workloads—all contributing to lower TCO.
- **Intel Parallel Studio XE 2019**, a developer toolkit for HPC applications, includes performance libraries such as the Intel Math Kernel library (Intel MKL) for HPC workloads, Intel Math Kernel Library for Deep Neural Networks (Intel MKL-DNN) to accelerate deep learning frameworks on Intel architecture, and Intel Data Analytics Acceleration Library (Intel DAAL) to speed big data analytics.
- **High-performance library building blocks** harness the power of Intel hardware and high-productivity tools and platforms that simplify and streamline workflows.
- **Optimizations for deep learning frameworks** for Intel architecture including Neon, Caffe, Theano, Torch, and TensorFlow, deliver increased value and performance for data scientists on Intel platforms.



HPC Select Solutions

Intel Select Solutions offer easy and quick-to-deploy infrastructure optimized to help you succeed.

- **Intel Select Solutions for Simulation & Visualization:** Assemble a solution optimized for highly scalable simulation and visualization applications, with both excellent performance and lower costs
- **Intel Select Solutions for Simulation and Modeling:** Deploy an optimized HPC cluster for simulation and modeling

- **Intel Select Solutions for Genomics Analytics:**
Access performance, scale, and ease of deployment for genomics insight and discovery
- **Intel Select Solutions for HPC & AI Converged:**
Deploy a solution that delivers the compute-intensive resources needed to run AI workloads on existing HPC clusters

Learn More About:

[Intel Select Solutions](#)

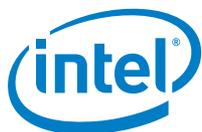
[Intel Xeon Scalable processors](#)

[Intel HPC solutions](#)

Intel: Delivering What You Need

Intel 2nd Gen Intel Xeon Scalable processors offers outstanding real-world application performance for HPC today in which traditional HPC workloads are enhanced with new analytical workloads including deep learning.

Platforms built around 2nd Gen Intel Xeon Scalable processors—the only processors with HPC and AI acceleration built-in—benefit from the agility of additional Intel HPC building blocks that help you solve the problems you need to tackle.



¹ 31% Higher Performance with 2S Intel Xeon-AP vs 2S AMD EPYC “Rome” 7742; For configuration details, visit www.intel.com/2019xeonconfigs/ (Intel® Xeon® Scalable Processors - claim #31).

² 2x better performance with AVX512 vs AVX2: Theoretical FLOPS when comparing Cascade Lake with AVX512 and 2 512 bit FMAs to older generation Broadwell with AVX2 and 2 256 bit FMAs.

³ 2x memory channels: Comparing 2nd Gen Intel Xeon Platinum 9200 processors with 12 memory channels per socket vs. Intel Xeon Scalable processors (Skylake) with 6 memory channels per socket.

⁴ 30x inference throughput improvement on Intel Xeon Platinum 9282 processor with Intel® DL Boost: For configuration details, visit www.intel.com/2019xeonconfigs/ (Intel® Xeon® Scalable Processors - claim #3)

Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products. For more complete information visit www.intel.com/2019xeonconfigs/.

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

Intel technologies' features and benefits depend on system configuration and may require enabled hardware, software or service activation.

Refer to <https://software.intel.com/en-us/articles/optimization-notice> for more information regarding performance and optimization choices in Intel software products.

Intel® Advanced Vector Extensions (Intel® AVX)* provides higher throughput to certain processor operations. Due to varying processor power characteristics, utilizing AVX instructions may cause a) some parts to operate at less than the rated frequency and b) some parts with Intel® Turbo Boost Technology 2.0 to not achieve any or maximum turbo frequencies. Performance varies depending on hardware, software, and system configuration and you can learn more at <https://www.intel.com/go/turbo>.

Intel, the Intel logo, Intel Optane, Xeon and other Intel marks are trademarks of Intel Corporation or its subsidiaries.

Other names and brands may be claimed as the property of others.

© 2019 Intel Corporation