Product Brief

5th Gen Intel® Xeon® Processors

With faster memory and a larger last-level cache, 5th Gen Intel Xeon processors deliver up to 42 percent better AI performance than the previous generation and up to 14x better AI performance than 3rd Gen Intel Xeon processors.[1,2]

Trusted performance. Exceptional efficiency.

The digital universe is buzzing with data. From telemetry data collected at the edge to online transactions to cloud data stores, our world is brimming with information. But raw data alone, regardless of its volume, is not what propels us into the future. To truly innovate, we must convert this abundance of data into intelligence.

Intel enables this transformation and helps businesses solve some of the world's most complex problems with its software and hardware, accelerating time to actionable insights. Customers access the performance and power efficiency they need to optimally run diverse and demanding workloads, using one of the most extensive product portfolios on the market. Together with its broad ecosystem of partners, Intel enables businesses to meet every challenge head-on.

Now Intel's customers can harness the power of 5th Gen Intel Xeon processors to turn data into knowledge and drive innovation forward. These processors deliver impressive performance-per-watt gains across all workloads, plus outsized performance and lower total cost of ownership (TCO) for AI, databases, networking, storage, and high-performance computing (HPC).[3] They offer more compute, larger shared last-level cache, and faster memory at the same power envelope as the previous generation.[4] They are also software- and platform-compatible with the previous generation of 4th Gen Intel Xeon processors, so you can minimize testing and validation when deploying new systems for AI and other workloads.

The processor designed for AI

AI is transforming how we work. Businesses save money and time and create new opportunities using state-of-the-art AI models for language, text-to-image, and code generation. With AI acceleration in every core, 5th Gen Intel Xeon processors with built-in AI accelerators like Intel® Advanced Matrix Extensions (Intel® AMX), faster memory, and larger last-level cache are ready to handle demanding AI workloads with improved performance for deep learning (DL) inference and training.

Benefit from up to 14x higher performance on PyTorch for inference and training with 5th Gen Intel Xeon processors, compared to 3rd Gen Intel Xeon processors.[5] Realize up to 2.5x higher performance per watt with 5th Gen Intel Xeon processors with Intel AMX, compared to 4th Generation AMD EPYC processors.[6]

With nimble large-language models (LLMs) of up to 20 billion parameters, 5th Gen Intel Xeon processors meet the service-level agreements (SLAs) for delivering real-time user experiences with less than 100 ms second-token latency on LLMs.[7]

Software tools and ecosystem to bring AI everywhere

Deploy AI fast with Intel Xeon processors offering the largest ecosystem and supply availability, a familiar toolset, and the ability to scale across the cloud, the data center, and the edge. Optimizations for Intel Xeon processors are already integrated into the mainstream distributions of popular AI frameworks, including TensorFlow and PyTorch. Dozens of pre-trained, optimized Intel AI models are ready to use and easily customized. Using these resources, developers can easily migrate code from different hardware environments.

5th Gen Intel Xeon processors excel in a range of AI use cases, including:

- **Generative AI:** Run generative AI models on your CPU foundation, like LLMs and text-to-image generation.
- **Recommender systems:** Deliver fast, personalized products or content recommendations that don't slow down the user experience.
- **Natural language processing (NLP):** Enable more responsive smart assistants, chatbots, and predictive text with a performance leap in NLP inference.
- **Machine learning (ML):** Get results faster when you apply classical ML to HPC and AI applications.
- **Image classification:** Boost productivity and efficiency by detecting defects on the manufacturing line, identifying products in retail applications, and much more.

Other use cases include the following:

General compute

Boost performance and increase return on investment (ROI) with low-latency systems created using 5th Gen Intel Xeon processors. Accelerators can offload tasks from CPU cores, reducing the number of cores needed per workload. This allows you to expand capacity or run more applications on each server.

1.84x average performance gain with 5th Gen Intel Xeon processors, compared to 3rd Gen Intel Xeon processors[8]

AI

Achieve AI performance that's unmatched by any other CPU, on-premises or in your cloud instance, with 5th Gen Intel Xeon processors featuring built-in Intel® AI Engines. Use Intel AMX, a built-in accelerator that significantly improves DL training and inference. Processor optimizations are integrated into popular AI frameworks, including TensorFlow and PyTorch.

Up to 14x higher performance on PyTorch for inference and training with 5th Gen Intel Xeon processors, compared to 3rd Gen Intel Xeon processors[5]

HPC

Improve the performance of HPC applications, from manufacturing simulations to earth systems modeling. 5th Gen Intel Xeon processors with larger last-level cache and faster memory provide high levels of precision while speeding up workloads. Intel® Advanced Vector Extensions 512 (Intel® AVX-512), a built-in accelerator with ultra-wide 512-bit vector operations capabilities, is especially suited for the demanding computational tasks in the HPC segment.

Up to 2.1x higher average HPC performance with 5th Gen Intel Xeon processors, compared to 3rd Gen Intel Xeon processors[9]

Web and microservices

Deliver responsive web experiences that engage online customers. Increase CPU efficiency by utilizing Intel® Accelerator Engines such as Intel® QuickAssist Technology (Intel® QAT) and Intel® Dynamic Load Balancer (Intel® DLB). These built-in accelerators help improve microservices networking and storage application performance.

Up to 1.3x higher throughput with 5th Gen Intel Xeon processors, compared to 4th Gen Intel Xeon processors, on DeathStarBench—Social Network Microservices workload (Read Home Timeline) with 100 ms SLA[10]

Database and analytics

Accelerate insights and productivity with fast databases and analytics powered by 5th Gen Intel Xeon processors and Intel® Analytics Engines. Activate Intel® In-Memory Analytics Accelerator (Intel® IAA) to help increase query throughput. Harness the capabilities of Intel® Data Streaming Accelerator (Intel® DSA), which takes on data copy and transformation operations to free CPU cycles. Together with faster memory and a larger last-level cache, these two accelerators enhance performance for in-memory databases, big data analytics, and data warehousing.

Up to 3.7x higher RocksDB performance with 5th Gen Intel Xeon processors using integrated Intel IAA, compared to 3rd Gen Intel Xeon processors using Zstd[11]

Networking

Add capacity to your networks by speeding up data movement, encryption, and compression. Employ 5th Gen Intel Xeon processors with Intel® Network Engines like Intel QAT and Intel DLB to allow for efficient network data placement and accelerated network encryption to help secure more web connections per second.

Up to 1.7x higher NGINX TLS handshake performance with 5th Gen Intel Xeon processors with integrated Intel QAT, compared to 3rd Gen Intel Xeon processors[12]

Improving performance and efficiency while reducing costs

Energy efficiency is a growing concern for organizations of all kinds, especially as they work to advance their sustainability initiatives. This makes it a priority to reduce power consumption in technology infrastructures.

5th Gen Intel Xeon processors offer 34 percent out-of-box performance/power improvement versus the previous generation.[13] You can extend power efficiency and savings further by enabling Optimized Power Mode in the platform BIOS for workloads that can benefit.

5th Gen Intel Xeon processors with Intel Accelerator Engines deliver up to 10x higher performance per watt on targeted workloads.[14] With more built-in accelerators than any other CPU on the market, 5th Gen Intel Xeon processors deliver outsized performance and TCO for AI, database, networking, and HPC workloads.[15]

Integrate 5th Gen Intel Xeon processors into your server refresh strategy to reduce power consumption at the same time you efficiently modernize your data center. Moving to the latest generation processor, you can realize performance gains and lower costs to support your organization's current and future needs. By delivering more cores and more performance per core, 5th Gen Intel Xeon processors help reduce the number of servers needed, which reduces power consumption and operating costs while still meeting performance requirements.[3]

Confidential computing

Confidential computing with trusted execution environments (TEEs) helps protect data and AI models. With 5th Gen Intel Xeon processors, you can choose from the most researched and updated confidential computing options in data centers on the market today. Intel® Software Guard Extensions (Intel® SGX) provides application isolation and is designed to enhance data protection at rest, in motion, and in use. Intel® Trust Domain Extensions (Intel® TDX) offers isolation and confidentiality at the virtual machine (VM) level. Customers can independently verify the integrity of their Intel confidential computing environments across multiple data centers, cloud providers, and edge locations using Intel® Trust Authority attestation services.

Technology overview

5th Gen Intel Xeon processors, built on a shared architectural platform with 4th Gen Intel Xeon processors, offer improved performance and performance per watt, TCO enhancements, and silicon-based security capabilities.[16] They can drive enhanced return on investment (ROI), with 21 percent better overall performance at the same thermal design power (TDP), compared to the previous generation.[8] 5th Gen Intel Xeon processors also boost performance for memory-bound and latency-sensitive workloads with faster memory and larger last-level cache, compared to the previous generation.[17]

Choose from a wide range of SKUs with more cores and CPU cache than the previous generation for higher overall performance in 1- or 2-socket server designs.[8] 5th Gen Intel Xeon processors provide accompanying advances in the memory and input/output (I/O) subsystems that let you scale and adapt as workload requirements change. These processors also enable you to:

- Get faster DDR5 memory, supporting up to 5,600 megatransfers per second (MT/s) (1 DPC), a 16 percent increase over 4th Gen Intel Xeon processors.[18]
- Increase shared last-level cache, with up to 320 MB last-level cache shared across all cores—an up to 3x increase over the prior generation for select SKUs.[19]
- Increase inter-socket bandwidth with Intel® Ultra Path Interconnect (Intel® UPI) 2.0, with up to 20 gigatransfers per second (GT/s)—a 25 percent increase over the prior generation.[20]
- Expand your confidential computing to VM-level workload isolation with Intel TDX, while continuing to implement application-level workload isolation with Intel SGX.
- Facilitate low-latency, coherent memory communication with the CPU by using Compute Express Link (CXL) Type 1 or 2 devices. Type 1 devices are smart network interface cards (NICs) and accelerators, while Type 2 devices are accelerators with cache.
- Expand memory capacity with CXL Type 3 memory devices to increase in-system memory bandwidth with targeted hyperscaler enablement.
- Eliminate the need for a traditional hardware RAID host bus adapter (HBA) card with Intel® Virtual RAID on CPU (Intel® VROC), an enterprise RAID solution specifically designed for NVM Express (NVMe) solid state drives (SSDs).
- Configure your CPU to meet specific workload needs with Intel® Speed Select Technology (Intel® SST).
- Improve availability and maintain data reliability across all platform architecture elements by increasing system uptime, reducing the duration of unplanned downtime, and maintaining data integrity with advanced reliability, availability, and serviceability (RAS) capabilities.
- Use Intel® Ethernet 800 Series Network Adapters to accelerate high-priority applications, packet processing, and latency-sensitive workloads.
- Solve for operational efficiency challenges with features such as seamless firmware updates, Intel® Platform Monitoring Technology, and Intel® Resource Director Technology (Intel® RDT).
- Flexibly add more acceleration or security when you need it, using Intel® On Demand, a service available on most 5th Gen Intel Xeon processors that allows you to upgrade accelerators or hardware-enhanced features.

Memory and I/O technologies featured in 5th Gen Intel Xeon processors include:

DDR5 support

Improve compute performance by overcoming data bottlenecks with high memory bandwidth. DDR5 offers up to 1.7x bandwidth improvement over DDR4,[21] enabling opportunities to improve performance, capacity, power efficiency, and cost. 5th Gen Intel Xeon processors boost memory speeds up to 5,600 MT/s (1 DPC) or 4,400 MT/s (2 DPC) with DDR5, increasing performance for memory-bound and latency-sensitive workloads.

PCIe 5.0 support

Double the I/O bandwidth versus PCIe 4.0, with opportunities to enable the highest possible throughput between the CPU and connected devices. 5th Gen Intel Xeon processors have 80 lanes of PCIe 5.0—ideal for fast networking, high-bandwidth accelerators, and high-performance storage devices. PCIe 5.0 doubles the I/O bandwidth from PCIe 4.0,[22] maintains backward compatibility, and provides foundational slots for CXL. The growing PCIe ecosystem gives customers the flexibility to customize hardware and expand performance.

CXL support

Reduce compute latency in the data center and help lower TCO with CXL 1.1 for next-generation workloads. CXL is an alternate protocol that runs across the standard PCIe physical layer and can support both standard PCIe devices and CXL devices on the same link. CXL provides a critical capability to create a unified, coherent memory space between CPUs and accelerators, and it will revolutionize how data center server architectures will be built for years to come.

Overview of 5th Gen Intel Xeon processors

Intel Xeon Platinum 8500 processors are the foundation for security-enabled, agile data centers—both on-premises and in the multicloud. They are designed for AI, advanced data analytics, high-density infrastructure, and multicloud workloads. These processors deliver high levels of performance, increased platform capabilities, and industry-leading workload acceleration with up to two-socket scalability. They offer enhanced hardware-based security and exceptional multi-socket processing performance. With trusted, hardware-enhanced data-service delivery and newer I/O and connectivity technologies, these processors deliver improvements in I/O, memory, storage, and network capabilities to harness actionable insights from the increasingly data-fueled world, including:

- Up to 64 cores per processor
- 8 memory channels per processor at up to 5,600 MT/s (1 DPC)
- AI acceleration with Intel AMX for a giant leap in DL inference and training performance

With up to two-socket scalability, **Intel Xeon Gold 6500 and Intel Xeon Gold 5500 processors** are optimized for demanding mainstream data center, multicloud compute, and network and storage workloads. With support for higher memory speeds and enhanced memory capacity, these processors deliver improved performance and superior memory capabilities compared to previous generations. They also offer hardware-enhanced security and workload acceleration.

Intel Xeon Silver 4500 processors deliver essential performance, improved memory speed, and increased power efficiency. They offer the hardware-enhanced performance required for entry-level data center compute, network, and storage. Additionally, edge-enhanced (EE) SKUs offer power efficiency for edge use cases.

Learn more

For more information on how these processors can advance your business, visit intel.com/5thgenxeon.

Discover more about Intel Accelerator Engines.

Explore the Intel® Developer Cloud and Intel® Developer Zone for additional resources.

Learn more about Intel® Ethernet products.

Up to two-socket scalability

Four Intel® Ultra Path Interconnect (Intel® UPI) ports at 20 GT/s

80 lanes of PCIe 5.0/CXL 1.1

DDR5 at up to 5,600 MT/s (1 DIMM per channel) or 4,400 MT/s (2 DIMMs per channel)

Intel AVX-512 (two 512-bit FMAs)

Intel® Hyper-Threading Technology (Intel® HT Technology) and Intel® Turbo Boost Technology

Intel AMX

Intel SST (Intel SST-TF, Intel SST-BF, and Intel SST-CF)

Advanced RAS capabilities

Intel SGX up to 512 GB maximum enclave size

Intel TDX

Intel DSA, 1 device

Workload acceleration with Intel QAT, Intel DLB, and Intel IAA on select SKUs

Up to two-socket scalability

Three Intel UPI ports at 20 GT/s

80 lanes of PCIe 5.0/CXL 1.1

DDR5 at up to 5,200 MT/s (1 DIMM per channel) or 4,400 MT/s (2 DIMMs per channel)

Intel AVX-512 (two 512-bit FMAs)

Intel HT Technology and Intel Turbo Boost Technology

Intel AMX

Intel SST (Intel SST-TF, Intel SST-BF, and Intel SST-CF)

Advanced RAS capabilities

Intel SGX up to 128 GB maximum enclave size

Intel TDX

Intel DSA, 1 device

Workload acceleration with Intel QAT, Intel DLB, and Intel IAA on select SKUs

Up to two-socket scalability

Two Intel UPI ports at 16 GT/s

80 lanes of PCIe 5.0/CXL 1.1

DDR5 at up to 4,400 MT/s (1 DIMM per channel) or 4,400 MT/s (2 DIMMs per channel)

Intel AVX-512 (two 512-bit FMAs)

Intel HT Technology and Intel Turbo Boost Technology

Intel AMX

Standard RAS capabilities

Intel SGX up to 64 GB maximum enclave size

Intel TDX

Intel DSA, 1 device

Workload acceleration with Intel QAT, Intel DLB, and Intel IAA on select SKUs

intel.
XEON®

[1] See [A17] at intel.com/processorclaims: 5th Gen Intel Xeon processors. Results may vary.
[2] See [A15] at intel.com/processorclaims: 5th Gen Intel Xeon processors. Results may vary.
[3] See [G1, T1] at intel.com/processorclaims: 5th Gen Intel Xeon processors. Results may vary.
[4] See [G5] at intel.com/processorclaims: 5th Gen Intel Xeon processors. Results may vary.
[5] See [A15, A16] intel.com/processorclaims: 5th Gen Intel Xeon processors. Results may vary.
[6] See [A209, 210] at intel.com/processorclaims: 5th Gen Intel Xeon processors. Results may vary.
[7] Based on Intel internal modeling as of December 2023.
[8] See [G1] at intel.com/processorclaims: 5th Gen Intel Xeon processors. Results may vary.
[9] See [H1] at intel.com/processorclaims: 5th Gen Intel Xeon processors. Results may vary.
[10] See [W1] at intel.com/processorclaims: 5th Gen Intel Xeon processors. Results may vary.
[11] See [D1] at intel.com/processorclaims: 5th Gen Intel Xeon processors. Results may vary.
[12] See [N15] at intel.com/processorclaims: 5th Gen Intel Xeon processors. Results may vary.
[13] Estimated performance on 1-node, pre-production platform with 2 x 5th Gen Intel Xeon Platinum 8592+ processor (64 cores, 350 W TDP); 1,024 GB total memory (16 x 64 GB DDR5 5,600 MT/s), Intel Turbo Boost Technology on, Intel HT Technology on, BIOS version: EGSDCRB1.SYS.0105.D74.2308261931, kernel: 6.2.0-emr.bkc.6.2.13.3.43.x86_64, microcode: 0x21000161; OS: CentOS Stream 9; software: JDK 1.11; 2 x Intel Ethernet® Controller I225-LM; workload: Power efficiency. Compared to: 1-node, 2 x 4th Gen Intel Xeon Platinum 8480+ processor (56 cores, 350 W TDP); 1,024 GB total memory (16 x 64 GB DDR5 4,800 MT/s), Intel Turbo Boost Technology on, Intel HT Technology on, BIOS version: EGSDCRB1.SYS.0105.D59.2308191339, kernel: 5.15.0-spr.bkc.pc.16.4.24.x86_64, microcode: 0x2b000541, OS: CentOS Stream 8, software: JDK 1.11; 2 x Intel Ethernet Controller I225-LM; workload: Power efficiency. Tested by Intel on 9/14/2023.
[14] Based on performance-per-watt gains of 1.46x to 10.6x with built-in accelerators on a range of AI, database, and networking workloads. See [A19–A25, D1, D2, D5, N16] at intel.com/processorclaims: 5th Gen Intel Xeon processors. Results may vary.
[15] See [A19–A25, D1, D2, D5, N16] at intel.com/processorclaims: 5th Gen Intel Xeon processors. Results may vary.
[16] See [G1, S1, T1] at intel.com/processorclaims: 5th Gen Intel Xeon processors. Results may vary.
[17] See [G11, G12] at intel.com/processorclaims: 5th Gen Intel Xeon processors. Results may vary.
[18] See [G12] at intel.com/processorclaims: 5th Gen Intel Xeon processors. Results may vary.
[19] See [G11] at intel.com/processorclaims: 5th Gen Intel Xeon processors. Results may vary.
[20] As measured by performance on SPDK NVMe TCP vs. 4th Gen AMD EPYC 9554. See [N201] at intel.com/processorclaims: 5th Gen Intel Xeon processors. Results may vary.
[21] 5th Gen Intel Xeon processor: 8 channels DDR5, up to 4,800 MT/s (1 DPC) vs. 3rd Gen Intel Xeon processor: 8 channels DDR4, 3,200 MT/s (2 DPC).
[22] Intel. "What Are PCIe 4.0 and 5.0?" www.intel.com/content/www/us/en/gaming/resources/what-is-pcie-4-and-why-does-it-matter.html.

Availability of accelerators varies depending on SKU. Visit the Intel Product Specifications page for additional product details.

Performance varies by use, configuration and other factors. Learn more at www.Intel.com/PerformanceIndex.

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

Statements in this document that refer to future plans or expectations are forward-looking statements. These statements are based on current expectations and involve many risks and uncertainties that could cause actual results to differ materially from those expressed or implied in such statements. For more information on the factors that could cause actual results to differ materially, see our most recent earnings release and SEC filings at www.intc.com.

© 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.

Printed in USA 1223/MG/PRW/PDF Please Recycle 356228-001US