

The Intel® Xeon® 6 Processor Family



Addressing the needs of today's data centers

Data center infrastructure is one of the most important investments an organization makes. IT leaders need to implement resources that foster business growth while balancing security, energy efficiency, manageability, and other factors affecting total cost of ownership (TCO).

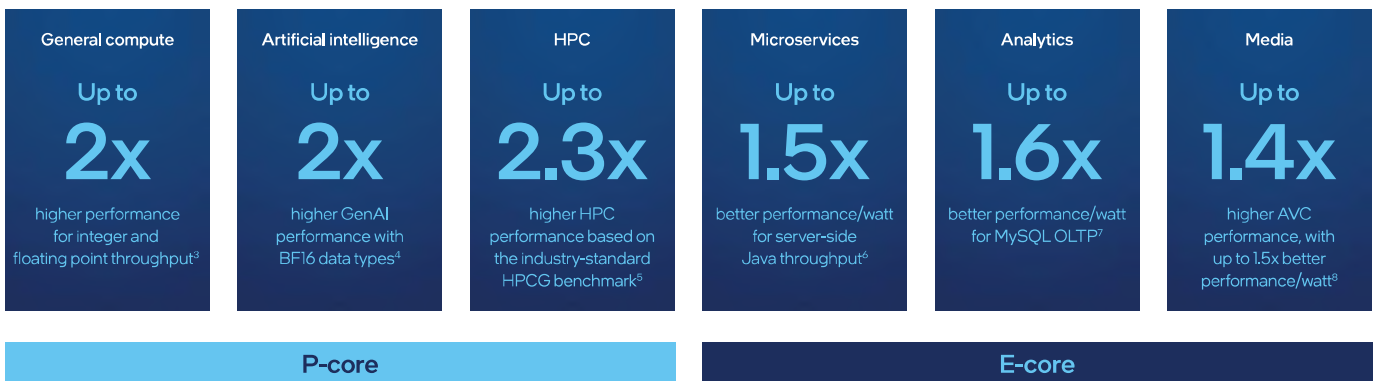
Perhaps most importantly, IT leaders must be cognizant of market factors driving rapidly evolving data center demands. For example, the AI market size and opportunity are expected to grow by 4x in the next five years, fueled by changes like workload automation and generative AI.¹ A growing cohort of enterprise applications is adding inferencing code, which requires processing large vectors of data with data-parallel computing requirements that favor performance per core. In that same time frame, the cloud microservices market is expected to grow by 5x, fueled by the redesign of monolithic applications with cloud-native principles.² These workloads are task-parallel. Thus, they benefit more from efficient, scalar processing than from added complex compute capabilities.

The Intel Xeon 6 processor family introduces a robust computing platform that excels at both performance and efficiency, crucial for meeting the evolving demands of modern data centers. From powering compute-intensive AI to enabling scalable cloud-native microservices, the processor family provides versatility for diverse operational requirements.

Faster business results across the spectrum of workloads

With more cores, flexible microarchitecture, additional memory bandwidth, and exceptional input/output (I/O), the Intel Xeon 6 processor family delivers new degrees of performance and efficiency across a range of workloads. New features and built-in accelerators give an additional boost to targeted workloads for even greater performance and efficiency.

Intel® Xeon® 6 processor compared to 5th Gen Intel® Xeon® Scalable processor



Intel® Xeon® 6 processor compared to 2nd Gen Intel® Xeon® Scalable processor



Exemplary user experience

Intel Xeon 6 processors provide the high level of quality and reliability that customers appreciate from Intel® products. Maintaining continuous operation and minimizing the time needed to service a system are fundamental to managing a data center’s service-level agreements (SLAs) and overall TCO. Intel reliability, availability, and serviceability (RAS) features encompass a suite of capabilities that help increase system uptime, reduce unplanned downtime, and maintain data integrity.

Security is important to user experience and customer satisfaction. IT teams must protect against a growing number of security threats and remain compliant with privacy regulations, whether on-premises or in the cloud. To protect data in use, Intel Xeon processors allow you to pick the confidential computing technologies that best meet your business and regulatory requirements. Intel® Trust Domain Extensions (Intel® TDX) offers isolation and confidentiality at the virtual machine (VM) level, while Intel® Software Guard Extensions (Intel® SGX) provides application-level isolation.

Performance and efficiency without compromise

The Intel Xeon 6 processor family introduces an innovative modular x86 architecture that allows data center architects to configure and deploy infrastructures that are purpose-built for your unique needs and workloads across private, public, and hybrid clouds. As shown in Table 1, Intel Xeon 6 processors are available in four different series, offering tiered capabilities from entry-level to demanding workloads through options for an increased number of cores, larger cache, faster and higher-capacity memory, and improved I/O over previous generations.

For the ultimate versatility, Intel Xeon 6 processors allow for the choice of two different CPU microarchitectures: Performance-cores (P-cores) and Efficient-cores (E-cores). Both core types use a compatible x86 instruction set architecture (ISA) and a common hardware platform, including CPU socket type. Furthermore, Intel has teamed with industry partners to help ensure seamless use of both core types with common operating systems, compilers, libraries, and frameworks. With this shared software stack and a global ecosystem of hardware and software vendors, solutions can be matched to every business need.

Table 1. The Intel Xeon 6 processor family encompasses four series of processors

| Series | Designed for |
|-----------------------------------|---|
| Intel Xeon 6900-series processors | Maximum performance ideal for the most demanding cloud, AI, and HPC environments |
| Intel Xeon 6700-series processors | Enhanced performance ideal for a wide array of data center and telco environments |
| Intel Xeon 6500-series processors | Essential performance ideal for mainstream server and edge environments |
| Intel Xeon 6300-series processors | Entry-level performance ideal for small/medium business environments |

Intel Xeon 6 processors with Performance-cores (P-cores)

P-cores are optimized for high performance per core and excel at the widest range of workloads, including better AI performance than any other general-purpose CPU. In comparison to 5th Gen Intel Xeon processors, which are the CPUs commonly referenced for newer compute-intensive solutions, Intel Xeon 6 processors with P-cores can provide up to 2x better performance.¹²

- Enable AI everywhere with AI acceleration in every core. Intel® Advanced Matrix Extensions (Intel® AMX) speeds up inferencing for int8 and BF16 and offers new support for FP16-trained models with up to 2,048 floating point operations per second (FLOPS) for int8 and 1,024 FLOPS for BF16/FP16.
- Improve memory throughput with the fastest DDR5 memory available, Multiplexer Combined Ranks (MCR) DIMMs. These can deliver more than 37 percent greater memory bandwidth than RDIMMs,¹³ with an expected data transfer rate of up to 8,800 megatransfers per second (MT/s).
- Take advantage of up to 128 cores per socket with L3 cache as large as 480 MB and with exceptionally low latency at large L3 access sizes.

Intel Xeon 6 processors with Efficient-cores (E-cores)

E-cores are optimized for high core density and exceptional performance per watt, delivering distinct advantages for cloud-scale workloads that demand high task-parallel throughput. In comparison to 2nd Gen Intel Xeon Scalable processors, which are within systems making up the majority of today's data center footprint and are excellent candidates for performance per watt upgrades, Intel Xeon 6 processors with E-cores can deliver more than 2.7x better results.¹¹ This efficient performance is also ideal where power, space, and cooling are limited. Intel Xeon 6 processors with E-cores can:

- Replace 4.3 2nd Gen Intel Xeon Scalable processor-based servers with a single server at similar performance.¹⁴
- Consolidate three racks of 2nd Gen Intel Xeon Scalable processor-based systems into a single rack.¹¹
- Accommodate AI inferencing and vector-oriented operations with Intel® Advanced Vector Extensions 2 (Intel® AVX2) and new enhancements such as Vector Neural Network Instructions (VNNI) and fast-convert for BF16 and FP16.
- Offer up to 288 cores per socket, with as much as 216 MB L3 cache, and with exceptionally low latency at large L3 access sizes.

The efficiency of Intel Xeon 6 processors with P-cores and E-cores is highlighted by their ability to provide scalable performance per watt as server utilization increases, delivering nearly linear power-performance consumption across the load line. For performance-demanding workloads, this means power is efficiently utilized at high loads to finish jobs faster. For a scalable implementation, common to cloud or shared computing environments, this level of efficiency means that servers are only consuming the power they need when under load, reducing costs when instances are not fully utilized.

The sustainability of these processors is further enhanced through system-wide power management and telemetry capabilities. These capabilities allow for increased performance per watt on a per-application basis to help with lowering overall energy consumption.

Versatility and complementary advantages of Intel Xeon 6 processors

At the extreme ends of the workload spectrum, P-cores offer the best solution for compute-intensive, vector-based workloads such as AI. E-cores are best for task-parallel, scalar-based workloads such as microservices. Between these extremes, the two microarchitectures combine to allow for highly versatile and complementary solutions. For example, systems with Intel Xeon 6 processors with E-cores can be used to conserve power so that it is available for AI and scientific workloads running on nodes with Intel Xeon 6 processors with P-cores. Data centers designed with a mix of Intel Xeon 6 processors with P-cores and Intel Xeon 6 processors with E-cores can take advantage of their platform commonality to transition workloads from one core type to the other depending on performance and power needs. The wide mix of options makes it easy for the data center to scale as the business grows.

As another example of the complementary nature of Intel Xeon 6 processors, a data center can easily mix servers with Intel Xeon 6 processors with P-cores and Intel Xeon 6 processors with E-cores to support business needs that require databases of different structures. Relational databases, which are characterized by complex data relationships, complex queries, joins, and aggregations, can benefit from the parallel data processing capabilities of Intel Xeon 6 processors with P-cores. Non-relational databases with numerous small, independent requests for data retrieval, such as key-value stores, can benefit from the task-parallel design of Intel Xeon 6 processors with E-cores.

Highlight technologies

The innovative P-core and E-core microarchitectures of the Intel Xeon 6 processor family deliver the following advanced features and benefits:

- Up to 288 cores in a single socket for Intel Xeon 6 processors with E-cores—or up to 128 cores in a single socket for Intel Xeon 6 processors with P-cores—enabling ultra-high-density compute performance and scalability.
 - Intel AMX provides up to 16x more multiply accumulate (MAC) operations than Intel AVX-512 for BF16 and FP16-based models to enhance AI performance (P-core-only feature).
 - Intel AVX-512 encompasses unique instructions and two 512-bit fused-multiply add (FMA) units per core, boosting the speed of vector mathematics common to AI, HPC, and database workloads (P-core-only feature).
 - Intel AVX2 with VNNI instructions and fast up/down convert for BF16 and FP16 enables better AI compatibility for Intel Xeon 6 processors with E-cores.
 - MCR DIMMs are capable of providing more than 37 percent additional memory bandwidth compared to standard DDR5 DIMMs, supporting bandwidth-constrained use cases found in AI and HPC (P-core-only feature).¹³
 - Up to 12 memory channels, further supporting higher memory bandwidth.
 - Intel® Ultra Path Interconnect (Intel® UPI) 2.0 provides up to 24 gigatransfers per second (GT/s) of inter-socket bandwidth—a 20 percent increase over the prior generation.
 - Up to 188 lanes of PCIe Gen 5 for two-socket servers with options of up to 136 lanes for one-socket server designs to allow for significant I/O add-in components including accelerators, network adapters, storage controllers, and storage.
 - Up to 64 lanes of Compute Express Link (CXL) 2.0 with data transfer rates up to 32 GT/s per lane, supporting CXL capabilities including memory expansion and sharing, including Type 3 devices.
 - Flat Memory Mode helps expand system memory and improve TCO when using lower-cost memory, such as DDR4 with CXL 2.0.
 - Intel® QuickAssist Technology (Intel® QAT) allows offload of bulk cryptography and compression to accelerate networking and storage.
 - Intel® Data Streaming Accelerator (Intel® DSA) 2.0 enables offload of data movement and transform operations such as move, fill, compare, cyclic redundancy checking (CRC), data integrity field (DIF), delta, and flush.
 - Intel® In-Memory Analytics Accelerator (Intel® IAA) allows offload of memory compression and decompression, scan and filter functions, and CRC.
 - Intel® Dynamic Load Balancer (Intel® DLB) enables the dynamic distribution of network packet processing and offload of reordering operations.
 - Intel TDX upgrades with AES-256 and 2,048 encryption keys enhance confidential computing for protection of sensitive business data.
 - The Intel® On Demand service lets your hardware provider enable select CPU-based features and capabilities. It is offered either as a one-time license-based feature activation or in a metering-based consumption model.
- To learn more about Intel Xeon 6 processors, including the listed features above, visit [intel.com/xeon](https://www.intel.com/xeon).

Overview of the Intel Xeon 6 processor family

Intel Xeon 6900-series processors are delivered in a new class of Intel server platform design, offering customers maximum performance, the highest memory bandwidth, and maximum throughput ideal for cloud, HPC, and AI environments. These processors feature higher core counts, more memory channels, and I/O lanes with thermal design points that are higher than the other series.

Intel Xeon 6700-series and Intel Xeon 6500-series processors are delivered in an updated server platform design featuring high performance with cost and power-efficient solutions ideal for the widest array of data center environments. These processors come in one-socket to eight-socket options with enhanced I/O and memory within established data center power and cooling footprints.

| Intel® Xeon® 6 processors | | Intel® Xeon® 6 CPUs with P-cores | Intel® Xeon® 6 CPUs with E-cores |
|---|---|---|--|
| Intel Xeon 6900-series processors | Maximum performance Introducing a new class of Intel server platform design, ideal for cloud computing, AI, HPC, software-as-a-service (SaaS), and infrastructure-as-a-service (IaaS) workloads. | <ul style="list-style-type: none"> Up to 128 cores (256 threads) per CPU Up to 500 W per CPU One- or two-socket servers 12 channel memory Up to 6,400 MT/s DDR5 8,800 MT/s MCR DIMMs Up to 96 PCIe 5.0 lanes Six Intel UPI 2.0 links Coming soon | <ul style="list-style-type: none"> Up to 288 cores (288 threads) per CPU Up to 500 W per CPU One- or two-socket servers 12 channel memory Up to 6,400 MT/s DDR5 Up to 96 PCIe 5.0 lanes Six Intel UPI 2.0 links Coming soon |
| Intel Xeon 6500-series/ 6700-series processors | Enhanced performance A significant upgrade to established Intel server platform designs. Mainstream servers from edge to cloud for enterprise IT, digital service providers, and telco. Ideal for AI, HPC, Networking and Media, Data Services, Infrastructure and Storage, Web, Applications, and Microservices. | <ul style="list-style-type: none"> Up to 86 cores (172 threads) per CPU Up to 350 W per CPU One-, two-, four-, or eight-socket servers Eight channel memory 6,400 MT/s DDR5 8,000 MT/s MCR DIMMs Up to 88 PCIe 5.0 lanes with up to 136 lanes for 1S designs Four Intel UPI 2.0 links Coming soon | <ul style="list-style-type: none"> Up to 144 cores (144 threads) per CPU Up to 330 W per CPU One- or two-socket server Eight channel memory 6,400 MT/s DDR5 Up to 88 PCIe 5.0 lanes Four Intel UPI 2.0 link Now available |



¹ Fortune Business Insights. "Artificial Intelligence Market Size, Share & Industry Analysis" April 2024. fortunebusinessinsights.com/industry-reports/artificial-intelligence-market-100114.

² Fortune Business Insights. "Cloud Microservices Market Size, Share & COVID-19 Impact Analysis" April 2024. fortunebusinessinsights.com/cloud-microservices-market-107793.

³ See [9G10] at intel.com/processorclaims: Intel® Xeon® 6. Results may vary.

⁴ See [9A10] at intel.com/processorclaims: Intel® Xeon® 6. Results may vary.

⁵ See [9H10] at intel.com/processorclaims: Intel® Xeon® 6. Results may vary.

⁶ See [7W4] at intel.com/processorclaims: Intel® Xeon® 6. Results may vary.

⁷ See [7D2] at intel.com/processorclaims: Intel® Xeon® 6. Results may vary.

⁸ See [7W2] at intel.com/processorclaims: Intel® Xeon® 6. Results may vary.

⁹ See [7D1] at intel.com/processorclaims: Intel® Xeon® 6. Results may vary.

¹⁰ See [7W1] at intel.com/processorclaims: Intel® Xeon® 6. Results may vary.

¹¹ See [7N1] at intel.com/processorclaims: Intel® Xeon® 6. Results may vary.

¹² Better performance versus 5th Gen Intel Xeon processors based on Intel architectural projections as of August 2023.

¹³ In comparison to DDR5 6,400 RDIMM.

¹⁴ Based on MySQL OLTP and server-side Java throughput with SLA. See intel.com/processorclaims: Intel® Xeon® 6. Results may vary.

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 configuration disclosure for additional 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.

© 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.