

3rd Generation Intel® Xeon® Scalable processors

Accelerating AI and analytics from cloud to the edge

As we enter a new data-fueled decade, companies need the ability to create future-ready products, services, and experiences. **3rd Generation Intel® Xeon® Scalable processors** evolve the previous-generation 4 to 8-socket processor foundation for today's AI-infused, data-intensive digital services.

UP TO **28 CORES** PER SOCKET

UP TO **224 CORES** IN AN
8-SOCKET CONFIGURATION

ONLY CURRENT GENERATION
X86 PLATFORM OFFERING
SCALABILITY UP TO **8 SOCKETS**



Built-in acceleration

Get faster insights from data-intensive workloads with built-in AI acceleration and massive memory capacity.

Trusted protection

Multilayer security helps optimize service delivery, thwart malicious exploits while maintaining workload integrity, and drive encryption efficiency at rest, in use, and in-flight.

Enhanced platform

Consistently deliver amazing experiences with hardware-enhanced virtualization across compute, network, and storage.

3rd Generation Intel Xeon Scalable processors are designed for:



DEEP LEARNING



VM DENSITY



IN-MEMORY
DATABASES



ANALYTICS-INTENSIVE
CLOUD SERVICES

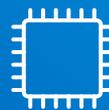
Pervasive performance



Industry's first x86 support of brain floating-point 16-bit (bfloat16) instructions



Up to **4.0 GHz**
Intel Turbo Boost
Technology



Up to **18TB** total system
memory (Intel® Optane™
persistent memory
200 series + DRAM)¹



Up to **3200 MT/s** &
16 Gbs DIMMs

Business resilience



New Intel Platform
Firmware Resilience



Intel Security Libraries
for Data Center



Intel Security Essentials

Agile service delivery



Intel Speed Select Technology



Intel Infrastructure
Management Technologies



Application Device Queues

Key selling points

Intel Deep Learning Boost (Intel DL Boost)

Enhanced Intel Deep Learning Boost uplevels inference and training performance to deliver AI readiness across the data center, to the edge and back.

- Accelerates AI/Deep Learning workloads, including image classification, speech recognition, language translation, and object detection.
- **Vector Neural Network Instructions (VNNI):** Delivers **significant performance improvement** by combining three instructions into one.
- **bfloat16:** Speeds AI performance for workloads requiring **high** compute intensity and **lower** need for precision.

INTEL DEEP LEARNING BOOST

**VNNI COMBINES
3 INSTRUCTIONS INTO 1**
TO ACCELERATE AI PERFORMANCE.

**bfloat16 DELIVERS
2X PEAK THROUGHPUT/CYCLE**
COMPARED TO FLOATING-POINT
32 FRAMEWORK.

DELIVERS OVER
225X FASTER CPU ACCESS
TO PERSISTENT DATA
THAN READING FROM A
MAINSTREAM NAND SSD.²

AVERAGE OF
38% HIGHER
MEMORY BANDWIDTH
IN MOST CONFIGURATIONS
COMPARED TO THE FIRST GENERATION.

Intel Optane persistent memory 200 series

Next-generation persistent memory modules enable massive memory capacity and lightning-fast access to persistently stored data.

- Extract **more value from larger datasets**, drive app innovation, and scale services delivery at a compelling TCO.
- Unlike conventional DDR memory, retain data even if power is lost unexpectedly.
- Will be validated for SAP HANA when deployed with 3rd Gen Intel Xeon Scalable processors.

Breakthrough DDR4 memory speed & capacity

Memory subsystem enhancements bring support for faster memory performance and unprecedented system memory capacity.

- Support for up to 6 channels of DDR4-3200 MT/s.
- 16Gb DIMM support enables up to 256GB DDR4 DIMMs per socket.
- Support for up to 18TB total system memory when combined with Intel Optane persistent memory.

UP TO
3200 MT/S DRAM SPEED
ON INTEL XEON
PLATINUM PROCESSORS

Unified, standards-based programming for diverse workloads

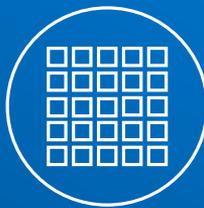
Using the Intel oneAPI ecosystem of cross-architecture tools, developers can deliver uncompromised performance and efficient development for AI, analytics, HPC, and other compute-intensive applications on Intel CPUs, GPUs, FPGAs, and other accelerators.



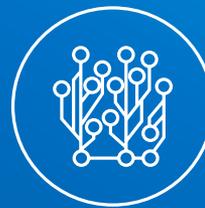
CPU



GPU



AI



FPGA

Using the latest platform optimizations, developers can:

- Accelerate AI workloads with advanced capabilities including Intel AVX-512, bfloat16, and VNNI.
- Future-proof code for new and upcoming generations of Intel processors and accelerators.



Processor levels and features

3rd Generation Intel Xeon Scalable processors

Intel Xeon Platinum 83XX processor

Performance leadership for AI training/inference and in-memory database

Intel Xeon Gold 63XX processor

Optimized for high core density/rack for lower TCO

Strong perf/W at a compelling TDP

Intel Xeon Gold 53XX processor

Solid balance of core count / frequency / price

| | | | |
|---|---|---|---|
| INTEL DEEP LEARNING BOOST (AI TRAINING) | ✓ | | |
| INTEL DEEP LEARNING BOOST (AI INFERENCE) | ✓ | | |
| VM DENSITY | | ✓ | |
| LATENCY SENSITIVE/HIGH FREQUENCY | ✓ | ✓ | |
| IAAS / DB SW COST OPTIMIZED / GENERAL COMPUTE | | ✓ | ✓ |
| LARGE MEMORY | ✓ | ✓ | |
| INTEL SPEED SELECT TECHNOLOGY | | ✓ | ✓ |

Qualification Questions

- Are you currently using large servers (≥ 4 sockets)?
- Do your workloads require large in-memory databases?
- Do you use common AI inferencing use cases, such as computer vision, image recognition, speech recognition, and/or anomalous behavior recognition?

Frequently Asked Questions

Question: How will 3rd Gen Intel Xeon Scalable processors support my customer's high-demand AI and deep learning use cases?

Answer: 3rd Gen Intel Xeon Scalable processors feature enhanced Intel Deep Learning Boost with the industry's first x86 support of bfloat16 instructions and VNNI, bringing enhanced artificial intelligence inference and training performance. New bfloat16 instructions benefit AI training workloads in healthcare, financial services, and retail where throughput and accuracy are key criteria, like vision, natural language processing, and reinforcement learning.

Question: My customers are looking for ways to process larger data volumes in real time so they can quickly roll out apps and services. What's the best way for them to deploy 3rd Gen Intel Xeon Scalable processors?

Answer: 3rd Gen Intel Xeon Scalable processors will be certified on Intel Select Solutions for SAP HANA to provide a simplified, optimized deployment based on high-capacity, non-volatile memory. This pre-configured, workload-driven hardware configuration also consists of Intel Optane persistent memory, Intel Solid State Drives, and Intel Ethernet Network Adapters. Designed to speed time to innovation and increase confidence for customers running in-memory databases, Intel Select Solutions for SAP HANA are verified by Intel against strict quality, performance, and security standards.

Question: My customer is concerned about security. Does the 3rd Gen Intel Xeon Scalable processor feature any notable security updates?

Answer: Yes. In addition to trusted Intel Security Essentials and Intel Security Libraries for Data Center, which support the building of a trusted, secure, and controlled cloud with hardware-enhanced root of trust, the 3rd Gen Intel Xeon Scalable platform introduces new Intel Platform Firmware Resilience, an Intel FPGA-based solution that can protect platform firmware, detect corruptions, and restore back to a known good state.³

Sales Resources for Intel Xeon Scalable processors

Drive IT Transformation

<https://www.intel.com/content/www/us/en/technology-provider/products-and-solutions/3rd-gen-xeon-scalable-family/overview.html>

Scale IT Up

scaleitup.intel.com

Use this sales tool to compare configurations side by side and see the benefits of IT modernization across the infrastructure. The latest Intel processor, storage, and network technologies are optimized to work together for maximum performance and TCO.

Intel Select Solutions

<https://www.intel.com/content/www/us/en/architecture-and-technology/intel-select-solutions-overview.html>

Explore how fast and easy infrastructure deployment can be with workload optimized data center solutions rigorously benchmark tested and verified on Intel architecture.

Intel Data Center TCO Tool

xeonprocessoradvisor.intel.com

Explore TCO for specific data center solutions and see recommended alternatives.

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

No product or component can be absolutely secure.

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 intel.com/benchmarks.

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 intel.com/go/turbo.

Cost reduction scenarios described are intended as examples of how a given Intel-based product, in the specified circumstances and configurations, may affect future costs and provide cost savings. Circumstances will vary. Intel does not guarantee any costs or cost reduction.

1. 6 x 512GB Optane persistent memory (3,072 GB) + 6 x 256GB DDR4 DRAM (1,536 GB) = 4,608 GB total memory per socket.

2. Intel® Optane persistent memory idle read latency of 340 nanoseconds. Intel® SSD DC P4610 Series Solid State Drive (TLC NAND) idle read latency of 77 microseconds.

3. No computer system can be absolutely secure.

