

3rd Generation Intel Xeon Scalable Processors

Empowering transformation in the data-centric era

Across an evolving digital world, disruptive and emerging technology trends increasingly impact the world's economies. As we enter a new decade, companies will depend on the ability to create digitally enhanced products, services, and experiences, and large organizations will embrace the power of advancements in technology and usage models they enable.

This global transformation is rapidly scaling the demands for flexible compute, networking, and storage. Future workloads will necessitate infrastructures that can seamlessly scale to support immediate responsiveness and widely diverse performance requirements. The exponential growth of data generation and consumption, the rapid expansion of cloud-scale computing and 5G networks, and the convergence of high-performance computing (HPC) and artificial intelligence (AI) into new usages requires that today's data centers and networks evolve now—or be left behind in a highly competitive environment. These demands are driving the architecture of modernized, future-ready data centers and networks that can quickly flex and scale.

The Intel Xeon Scalable platform provides the foundation for an evolutionary leap forward in data center agility and scalability. Disruptive by design, this innovative processor sets a new level of platform convergence and capabilities across compute, storage, memory, network, and security. Enterprises, cloud, and communications service providers can now accelerate their most ambitious digital initiatives with a feature-rich, highly versatile platform.



Enabling greater efficiencies and lower TCO

Across infrastructures, from enterprise to technical computing applications, the Intel Xeon Scalable platform is designed for data center modernization to drive operational efficiencies that lead to improved total cost of ownership (TCO) and higher productivity for users. Systems built on the Intel Xeon Scalable platform are designed to deliver agile services with enhanced performance and groundbreaking capabilities.

Pervasive performance to propel insights

3rd Gen Intel Xeon Scalable processors deliver industry-leading, workload-optimized platforms with built-in AI acceleration, providing a seamless performance foundation to help speed data's transformative impact, from the multi-cloud to the intelligent edge and back.

- **Enhanced performance:** Delivers multi-socket core count density with up to 28 cores per processor and up to 224 cores per platform in an 8-socket configuration, driving enhanced performance, throughput, and CPU frequencies compared to 2nd Gen Intel Xeon Scalable processors.
- **Enhanced Intel Deep Learning Boost (Intel DL Boost) with VNNI and new bfloat16:** Enhanced Intel Deep Learning Boost with the industry's first x86 support of Brain Floating Point 16-bit (bfloat16) and Vector Neural Network Instructions (VNNI) brings enhanced artificial intelligence inference and training performance, delivering up to 1.93% more AI training performance vs. the previous generation.¹
- **More Intel Ultra Path Interconnect (Intel UPI):** Up to six Intel UPI channels increase platform scalability and improve inter-CPU bandwidth for I/O-intensive workloads versus the previous generation, offering the perfect balance between improved throughput and energy efficiency.
- **Increased DDR4 memory speed & capacity:** Memory subsystem enhancements include support for up to 6 channels of DDR4-3200 MT/s and 16Gb DIMMs, with up to 256GB DDR4 DIMMs per socket.
- **Intel Advanced Vector Extensions 512 (Intel AVX-512):** Intel AVX-512 boosts performance and throughput for the most demanding computational tasks in applications, such as modeling and simulation, data analytics and machine learning, data compression, visualization, and digital content creation. Now available with up to 2 Fused Multiply Add (FMA) instructions, beginning with Intel Xeon Gold 5300 processors.
- **Support for Intel Optane SSDs and Intel QLC 3D NAND SSDs:** Delivers industry-leading combination of high throughput, low latency, high QoS, and ultra-high endurance to break through data access bottlenecks. With low-latency NVMe* interconnect, Intel's low-overhead storage software, and our most advanced TLC 3D NAND, the all-new Intel SSD D7-P5500 & P5600 series are designed to support the intense I/O requirements of AI and analytics workloads in data center and edge deployments.



UP TO **1.93X MORE**
AI TRAINING PERFORMANCE
FOR IMAGE CLASSIFICATION
VS. PRIOR GENERATION¹



UP TO **1.92X HIGHER**
CLOUD DATA ANALYTICS
PERFORMANCE
VS. 5-YEAR OLD 4-SOCKET PLATFORM²

Business resilience

3rd Gen Intel Xeon Scalable processors with hardware-enhanced security can help thwart malicious exploits while maintaining workload integrity, with reduced performance overhead. Provide trusted service delivery with high availability and encryption efficiency at rest, in-use, and in-flight.



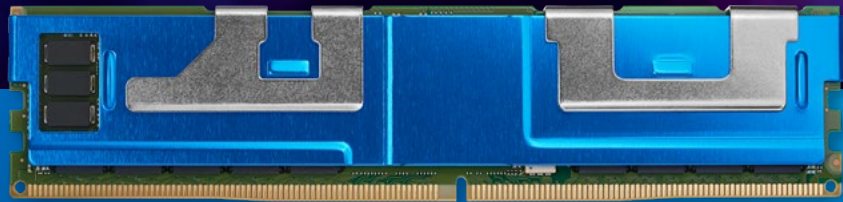
- **New Intel Platform Firmware Resilience (Intel PFR):** An Intel FPGA-based solution that can protect platform firmware, detect corruptions, and restore back to a known-good state.³



- **Intel Security Essentials and Intel Security Libraries for Data Center:** Supports the building of a trusted, secure, and controlled cloud with hardware-enhanced root of trust.³



- **Integrated Intel QuickAssist Technology (Intel QAT):** Chipset-based hardware acceleration for growing compression and cryptographic workloads enables greater efficiency while delivering up to 100 GB/s enhanced data transport and protection across server, storage, and network infrastructure.³ Available on select Intel C620A series chipsets.



Intel Optane persistent memory 200 series

New Intel Optane persistent memory 200 series represents a groundbreaking technology innovation. Delivered with all-new 3rd Gen Intel Xeon Scalable processors, this workload-optimized technology will help businesses extract more actionable insights from data—from cloud and databases, to in-memory analytics and more.

Compared to the first generation, the Intel Optane persistent memory 200 series delivers an average of 25% higher memory bandwidth.⁴

Learn more at intel.com/optanepersistentmemory

Agile service delivery

The 3rd Gen Intel Xeon Scalable platform brings innovations and hardware-enhanced virtualization across compute, network, storage, and persistent memory that help fuel cost-efficient, flexible, and scalable multi-cloud to consistently deliver amazing business-to-business and business-to-consumer experiences.

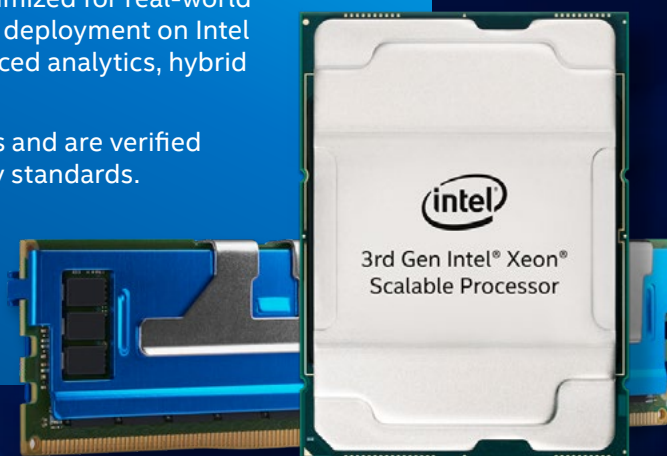
- **Intel Speed Select Technology:** A collection of configurable processor features, Intel Speed Select Technology (Intel SST) enables the optimization of processing resources to enhance workload performance, increase utilization, and help to optimize platform TCO. Available on select 3rd Gen Intel Xeon Scalable processors.
- **Support for Intel Optane persistent memory 200 series (Intel Optane PMem):** Complements DRAM by affordably enabling unprecedented system memory capacity to accelerate workload processing and service delivery. With module capacities up to 512 GB, Intel Optane persistent memory 200 series—supported on 4-socket 3rd Gen Intel Xeon Scalable platform designs—can deliver up to up to 3TB of persistent memory per socket.
 - For breakthrough system memory capacity, Intel Optane persistent memory enables up to 4.5TB total memory per socket and 18TB total system memory when combined with DRAM (in supported 4-socket designs only).
 - Compared to the first generation, the Intel Optane persistent memory 200 series delivers an average of 25% higher memory bandwidth.⁴
- **Intel Infrastructure Management Technologies (Intel IMT):** A framework for resource management, Intel Infrastructure Management Technologies include **Intel Resource Director Technology (Intel RDT)** and **Intel Virtualization Technologies (Intel VT-x)** for platform-level detection, reporting, and configuration. This hardware-enhanced monitoring, management, and control of resources can help enable greater data center resource efficiency, utilization and security.
- **Application Device Queues (ADQ):** An open technology on Intel Ethernet 800 series network adapters, Application Device Queues offer application-specific, uncontended, smooth-flowing traffic because there is no sharing of traffic by other applications on these queues.
- **Intel Stratix 10 NX FPGA:** Targeted for AI applications that require many variables to be evaluated in real-time across multiple nodes, the Intel Stratix 10 NX FPGA delivers exceptional performance for workloads such as natural language processing and financial fraud detection. Delivering accelerated AI compute with in-package high-bandwidth memory, accelerated matrix and vector operations, and integrated high-throughput networking, this flexible FPGA can be reprogrammed and optimized as models change and scale.

Faster time to value with Intel Select Solutions

In today's complex data center, hardware and software infrastructure are not "one size fits all." Intel Select Solutions eliminate guesswork with rigorously benchmark tested and verified solutions optimized for real-world performance. These solutions accelerate infrastructure deployment on Intel Xeon processors for today's critical workloads in advanced analytics, hybrid cloud, storage, and network communications.

Intel Select Solutions are delivered by solution partners and are verified by Intel against strict quality, performance, and security standards.

Learn more about Intel Select Solutions featuring new 3rd Gen Intel Xeon Scalable processors at intel.com/selectsolutions



Built-in AI performance with enhanced Intel Deep Learning Boost, now with bfloat16

Today's scientific discoveries are fueled by innovative algorithms, new sources and volumes of data, and advances in compute and storage. Machine learning, deep learning, and AI converge the capabilities of massive compute with the flood of data to drive next-generation applications, such as autonomous systems and self-driving vehicles.

3rd Gen Intel Xeon Scalable processors are built specifically for the flexibility to run complex AI workloads on the same hardware as your existing workloads. Enhanced Intel Deep Learning Boost, with the industry's first x86 support of Brain Floating Point 16-bit (bfloat16) numeric format and Vector Neural Network Instructions (VNNI), brings enhanced artificial intelligence inference and training performance, with up to 1.93X more AI training performance and 1.87X more AI inference performance for image classification vs. the prior generation.^{1,5} New bfloat16 processing support benefits AI training workloads in healthcare, financial services, and retail where throughput and accuracy are key criteria, like vision, natural language processing (NLP), and reinforcement learning (RL). Intel Deep Learning Boost with bfloat16 delivers 1.7X more AI training performance for natural language processing vs. the prior generation.⁶ 3rd Gen Intel Xeon Scalable processors help to deliver AI readiness across the data center, to the edge and back.

Segments



ENTERPRISE &
GOVERNMENT



CLOUD SERVICE
PROVIDERS



HEALTHCARE



FINANCIAL
SERVICES



RETAIL

Highlights for AI innovation

- Intel Xeon Platinum 8300 processors
- Intel Optane persistent memory 200 series
- Enhanced DDR performance up to 3200 MT/s
- Intel Deep Learning Boost, now with bfloat16
- More Intel Ultra Path Interconnect vs. previous generation
- Intel Advanced Vector Extensions 512
- Intel Omni-Path Architecture Host Fabric Interface
- Intel Optane SSDs

Learn more about built-in AI performance and acceleration with Intel Xeon Scalable processors at ai.intel.com

UP TO 1.87X MORE
AI INFERENCE PERFORMANCE
FOR IMAGE CLASSIFICATION
VS. PRIOR GENERATION⁵

UP TO 1.7X MORE
AI TRAINING PERFORMANCE FOR
NATURAL LANGUAGE PROCESSING
VS. PRIOR GENERATION⁶

UP TO 1.9X MORE
AI INFERENCE PERFORMANCE FOR
NATURAL LANGUAGE PROCESSING
VS. PRIOR GENERATION⁷

Delivering enhanced per-node performance for cloud, analytics, and mission critical

Cloud and enterprise customers are keen to extract value from the exploding data streams being presented to them for rapid insights that can shape their business initiatives. Traditional and emerging applications in the enterprise, including predictive analytics, machine learning, HPC, and mission critical usages require new levels of powerful compute capabilities and massive tiered data storage volumes. The modernized data center is being architected using a converged and holistic approach that can flexibly deliver new services and improve TCO across infrastructure assets today while providing the most seamless and scalable on-ramp to a self-governing, hybrid data center.

The Intel Xeon Scalable platform delivers next-generation capabilities to businesses through a future-ready platform that can serve the hybrid-cloud, data-fueled era while helping to improve day-to-day operations. This versatile platform brings disruptive levels of compute performance, coupled with memory and I/O advances, to compute-hungry and latency-sensitive applications. Combined with innovative Intel Optane SSDs and Intel QLC 3D NAND SSDs to manage large data volumes across storage, caching, and memory, platforms built on the Intel Xeon Scalable platform are ready to handle the intense demands of the data and cloud era.

Segments



ENTERPRISE &
GOVERNMENT



CLOUD SERVICE
PROVIDERS



MANUFACTURING



LIFE SCIENCES



OIL AND GAS



FINANCIAL
SERVICES

Highlights for cloud innovation

- Intel Xeon Platinum 8300 and Gold 6300 processors
- Intel Optane persistent memory 200 series
- More Intel Ultra Path Interconnect vs. previous generation
- Intel Deep Learning Boost, now with bfloat16
- Up to 224 cores per node in an 8-socket configuration
- Intel Advanced Vector Extensions 512, now with 2 FMA in Intel Xeon Gold 5300 processors
- Intel Omni-Path Architecture Host Fabric Interface
- Intel Optane SSDs

Learn more about the benefits 3rd Gen Intel Xeon Scalable processors bring to Cloud, Analytics and Mission Critical usages at intel.com/cloud

PROCESS UP TO
1.98X MORE OLTP DATABASE
TRANSACTIONS PER MINUTE
VS. 5-YEAR-OLD 4-SOCKET PLATFORM⁸

PROCESS LARGE DATASETS
WITH UP TO **31.2 BILLION**
INITIAL RECORDS ON AN 8-SOCKET
SINGLE-NODE PLATFORM
USING THE LATEST 3RD GEN INTEL XEON SCALABLE
PROCESSOR AND INTEL OPTANE PMEM 200 SERIES⁹

Enhanced VM density capabilities per node in a scalable design

Rapid growth of digital services is driving demand for more virtual machines (VMs) in public and private data centers. To increase data center capacity and efficiency, IT organizations continually seek to increase VM performance and density, loading more and more VMs per server.

The 3rd Gen Intel Xeon Scalable platform contains processors designed specifically to deliver enhanced VM density per node and seamless VM migration across five generations of Intel Xeon processors and platforms. This versatile platform enables higher resource utilization across processor, memory, storage, and I/O while meeting service-level agreements with fail-over and recovery solutions. Businesses can leverage the Intel Xeon Scalable platform to improve TCO by consolidating more applications to higher performing servers to optimize space, power, cooling, and maintenance costs. Combined with Intel Optane persistent memory 200 series to rapidly provision VMs for improved infrastructure agility and scalability, platforms built on the Intel Xeon Scalable platform enable enterprises to improve configuration flexibility for load balancing, peak workload management, test and development, and system maintenance.

Segments



**ENTERPRISE &
GOVERNMENT**



**CLOUD SERVICE
PROVIDERS**

Highlights for VM density

- Intel Xeon Platinum 8300 processors
- Intel Optane persistent memory 200 series
- Enhanced DDR performance with up to 3200 MT/s
- More Intel Ultra Path Interconnect vs. previous generation
- Intel Deep Learning Boost, now with bfloat16
- Up to 224 cores per node in an 8-socket configuration
- Intel Advanced Vector Extensions 512
- Intel Omni-Path Architecture Host Fabric Interface
- Intel Optane SSDs

Learn more about the benefits 3rd Gen Intel Xeon Scalable processors bring to VM Density usages at intel.com/csp

**UP TO
2.2X MORE VMS
VS. 5-YEAR-OLD 4-SOCKET PLATFORM¹⁰**



Overview of 3rd Gen Intel Xeon Scalable processors

Intel Xeon Platinum 8300 processors are the foundation for secure, agile, hybrid-cloud data centers. With enhanced hardware-based security and exceptional multi-socket processing performance, these processors are built for mission-critical, real-time analytics, machine learning, artificial intelligence, and multi-cloud workloads.³ With trusted, hardware-enhanced data service delivery, these processors deliver improvements in I/O, memory, storage, and network technologies to harness actionable insights from our increasingly data-fueled world. Designed for advanced analytics, artificial intelligence, and high-density infrastructure, new Intel Xeon Platinum 8300 processors deliver new levels of performance, platform capabilities and industry-leading workload acceleration.

- Up to 28 cores per Intel Xeon Scalable processor
- 6 memory channels per processor at up to 3200 MT/s (1 DPC)
- Features Intel Deep Learning Boost, now with bfloat16 and VNNI for enhanced AI inference acceleration and performance

With support for higher memory speeds, enhanced memory capacity, and up to four-socket scalability, **Intel Xeon Gold 6300 and 5300 processors** deliver improved performance, enhanced memory capabilities, hardware-enhanced security, and workload acceleration. These processors are optimized for demanding mainstream data center, multi-cloud compute, and network and storage workloads. With up to four-socket scalability, they are suitable for an expanded range of workloads.

Intel Xeon Scalable processors for one- and two-socket platforms

New 3rd Gen Intel Xeon Scalable processors are introduced first on platforms supporting four- and eight-socket designs. Coming soon, Intel will announce additional 3rd Gen Intel Xeon Scalable processors to support one- and two-socket platform designs. These processors will deliver generational platform and technology advancements in 2020 and beyond for mainstream server use cases.

Helping to deliver continued enhancements in two-socket server performance, Intel introduced new 2nd Gen Intel Xeon Scalable processors in February 2020, offering enhanced performance and value for Intel Xeon Gold, Silver, and Bronze processors. Please visit newsroom.intel.com/news/xeon-scalable-5g-network-portfolio-launch/ for more information regarding new 2nd Gen Intel Xeon Scalable processors.

Workload guidance

3rd Gen Intel Xeon Scalable processors are suitable for a wide variety of usages. Processor characteristics, including core count, base and turbo frequencies, memory configuration, and workload accelerators—like Intel Deep Learning Boost and Intel Speed Select Technology—support enhanced performance for certain workloads and services. This workload guidance is provided as a resource to help guide your journey as you explore new 3rd Gen Intel Xeon Scalable processors.

	Intel Deep Learning Boost (AI training)	Intel Deep Learning Boost (AI inference)	In-memory database	VM density	Latency sensitive/ high frequency	IaaS / DB SW cost optimized / general compute	Large memory Up to 4.5TB/ socket	Intel Speed Select Technology
Intel Xeon Platinum 8380HL processor			•				•	
Intel Xeon Platinum 8380H processor	•							
Intel Xeon Platinum 8376HL processor			•				•	
Intel Xeon Platinum 8376H processor	•							
Intel Xeon Platinum 8354H processor	•	•						
Intel Xeon Platinum 8353H processor								
Intel Xeon Gold 6348H processor				•				
Intel Xeon Gold 6328HL processor							•	•
Intel Xeon Gold 6328H processor					•			•
Intel Xeon Gold 5320H processor						•		•
Intel Xeon Gold 5318H processor						•		

New Intel Speed Select Technology (Intel SST) features

Intel Speed Select Technology is a collection of features giving you more granular control over processor performance for optimized total cost of ownership. With Intel SST, one server can do more. Lots more.

Intel Speed Select Technology – Core Power (Intel SST-CP)

Designed for workloads that benefit from higher base frequencies on a subset of processor cores, and lower base frequencies on the remaining cores, all while maintaining max turbo frequencies across all cores.

Intel Speed Select Technology – Turbo Frequency (Intel SST-TF)

Designed for workloads that benefit from higher turbo frequencies on a subset of processor cores, and lower turbo frequencies on the remaining cores, all while maintaining base frequencies across all cores.

Intel SST-CP and Intel SST-TF features are supported on new Intel Xeon Gold 6328HL, 6328H, and 5320H processors.

3rd Gen Intel Xeon Scalable processor SKUs

For the most up-to-date information, please [visit intel.com/xeon](https://www.intel.com/xeon) or ark.intel.com

Processor identifier	Cores /threads	TDP (watts)	Processor cache	Processor base frequency (GHz)	Max Turbo frequency rate (GHz)	Cache (MB)	Memory capacities HL/H SKU suffix
PLATINUM 8380HL PLATINUM 8380H	28/56	250	38.5 MB	2.9	4.3	38.5	4.5 TB / 1.12 TB
PLATINUM 8376HL PLATINUM 8376H	28/56	205	35.75 MB	2.6	4.3	38.5	4.5 TB / 1.12 TB
PLATINUM 8354H	18/36	205	24.75 MB	3.1	4.3	24.75	1.12 TB
PLATINUM 8353H	18/36	150	24.75 MB	2.5	3.8	24.75	1.12 TB
GOLD 6348H	24/48	165	33 MB	2.3	4.2	33	1.12 TB
GOLD 6328HL GOLD 6328H	16/32	165	22 MB	2.8	4.3	22	4.5 TB / 1.12 TB
GOLD 5320H	20/40	150	27.5 MB	2.4	4.2	27.5	1.12 TB
GOLD 5318H	18/36	150	24.75 MB	2.5	3.8	24.75	1.12 TB

	Intel Xeon Gold processor (5300 series)	Intel Xeon Gold processor (6300 series)	Intel Xeon Platinum processor (8300 series)
Highest Core Count Supported	20 cores	24 cores	28 cores
Highest Supported Frequency	3.3 GHz	3.7 GHz	4.3 GHz
Number of CPU Sockets Supported	Up to 4	Up to 4	Up to 8
Intel Ultra Path Interconnect (Intel UPI)	6	6	6
Intel UPI Speed	10.4 GT/s	10.4 GT/s	10.4 GT/s
Intel Advanced Vector Extensions 512 (Intel AVX-512)	2 FMA	2 FMA	2 FMA
Highest Memory Speed Support (DDR4)	2666 MT/s	2933 MT/s	3200 MT/s (1 DPC) 2933 MT/s (2 DPC)
Highest Memory Capacity Supported Per Socket ^o	1.12 TB	1.12 TB, 4.5 TB	1.12 TB, 4.5 TB
Intel Deep Learning Boost (Intel DL Boost) with Vector Neural Network Instructions (VNNI) and Brain Floating Point 16-bit (bfloat16) numeric format Processing	•	•	•
Intel Optane Persistent Memory Module Support (4-socket designs only)	•	•	•
Intel Omni-Path Architecture (Discrete PCI Express card)	•	•	•
Intel QuickAssist Technology (Integrated in select chipsets)	•	•	•
Intel QuickAssist Technology Support (Discrete PCI Express card)	•	•	•
Intel Optane SSDs Support	•	•	•
Intel DC SSD Support (3D NAND)	•	•	•
PCI Express 3 (48 lanes per processor)	•	•	•
Intel QuickData Technology (CBDMA)	•	•	•
Non-Transparent Bridge (NTB)	•	•	•
Intel Turbo Boost Technology 2.0	•	•	•
Intel Hyper-Threading Technology (Intel HT Technology)	•	•	•
Node Controller Support		•	•
Reliability, Availability, and Serviceability (RAS) Capability	Standard	Advanced	Advanced
Intel Run Sure Technology		•	•
Intel Speed Select Technology (Intel SST) – Features include Intel SST Core Power (SST-CP) and Intel SST Turbo Frequency (SST-TF)	•	•	•
Intel Infrastructure Management Technologies (Intel IMT)	•	•	•
Intel Resource Director Technology (Intel RDT)	•	•	•
Intel Volume Management Device (Intel VMD)	•	•	•
Intel Virtualization Technology (Intel VT)	•	•	•
Intel Speed Shift Technology	•	•	•
Intel Node Manager 4.0	•	•	•
Mode-Based Execute Control	•	•	•
Timestamp Counter Scaling (TSC) for Virtualization	•	•	•
Intel Security Libraries for Data Center (Intel ISeCL-DC)	•	•	•
Intel Advanced Vector Extensions 512 (Intel AVX-512)	•	•	•
Intel Key Protection Technology (KPT) with Integrated Intel QAT	•	•	•
Intel Platform Trust Technology (PTT)	•	•	•
Intel Trusted Execution Technology (Intel TXT) with One-Touch Activation (OTA)	•	•	•

Product Name	SKU	Compression	Encryption	RSA
Intel C621A Chipset	LBG-1G	N/A	N/A	N/A
Intel C627A Chipset with Intel QuickAssist Technology	LBG-T	~65 Gb/s	100 Gb/s	100K Ops/s
Intel C629A Chipset with Intel QuickAssist Technology	LGB-C	~75-80 GB/s	100 Gb/s	N/A

^oSupported on select SKUs only.

Notices & Disclaimers

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/benchmarks.

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.

1. Up to 1.93x higher AI training performance with 3rd Gen Intel® Xeon® Scalable processor supporting Intel® DL Boost with BF16 vs. prior generation on ResNet50 throughput for image classification – New: 1-node, 4x 3rd Gen Intel® Xeon® Platinum 8380H processor (pre-production 28C, 250W) on Intel Reference Platform (Cooper City) with 384 GB (24 slots / 16GB / 3200) total memory, ucode 0x700001b, HT on, Turbo on, with Ubuntu 20.04 LTS, Linux 5.4.0-26,28,29-generic, Intel 800GB SSD OS Drive, ResNet-50 v 1.5 Throughput, <https://github.com/Intel-tensorflow/tensorflow/-b bf16/base, commit#828738642760358b388d8f615ded0c213f10c99a>, Modelzoo: <https://github.com/IntelAI/models/-b v1.6.1, Imagenet dataset, oneDNN 1.4, BF16, BS=512, test by Intel on 5/18/2020>. Baseline: 1-node, 4x Intel® Xeon® Platinum 8280 processor on Intel Reference Platform (Lightning Ridge) with 768 GB (24 slots / 32GB / 2933) total memory, ucode 0x4002f00, HT on, Turbo on, with Ubuntu 20.04 LTS, Linux 5.4.0-26,28,29-generic, Intel 800GB SSD OS Drive, ResNet-50 v 1.5 Throughput, <https://github.com/Intel-tensorflow/tensorflow/-b bf16/base, commit#828738642760358b388d8f615ded0c213f10c99a>, Modelzoo: <https://github.com/IntelAI/models/-b v1.6.1, Imagenet dataset, oneDNN 1.4, FP32, BS=512, test by Intel on 5/18/2020>.
2. Up to 1.92x higher performance on cloud data analytics usage models with the new 3rd Gen Intel® Xeon® Scalable processor vs. 5-year old 4-socket platform – New: 1-node, 4x 3rd Gen Intel® Xeon® Platinum 8380H processor (pre-production 28C, 250W) on Intel Reference Platform (Cooper City) with 1536GB (48 slots / 32 GB / 3200 (@2933) total memory, microcode 0x700001b, HT on, Turbo on, with Ubuntu 18.04.4 LTS, 5.3.0-53-generic, 1x Intel 240GB SSD OS Drive, 4x P4610 3.2TB PCIe NVMe, 4 x 40 GbE x710 dual port, CloudXPRT vCP - Data Analytics, Kubernetes, Docker, Kafka, MinIO, Prometheus, XGBoost workload, Higgs dataset, test by Intel on 5/27/2020. Baseline: 1-node, 4x Intel® Xeon® processor E7-8890 v3 on Intel Reference Platform (Brickland) with 1024 GB (64 slots / 16GB / 1600) total memory, microcode 0x0000016, HT on, Turbo on, with Ubuntu 18.04.4 LTS, 5.3.0-53-generic, 1x Intel 400GB SSD OS Drive, 4x P3700 2TB PCIe NVMe, 4 x 40 GbE x710 dual port, CloudXPRT vCP - Data Analytics, Kubernetes, Docker, Kafka, MinIO, Prometheus, XGBoost workload, Higgs dataset, test by Intel on 5/27/2020.
3. No computer system can be absolutely secure.
4. Baseline: 1-node, 1x Intel® Xeon® 8280L 28C @ 2.7GHz processor on Neon City with Single PMem module config (6x32GB DRAM; 1x{128GB,256GB,512GB} Intel Optane PMem 100 Series module at 15W) ucode Rev: 04002F00 running Fedora 29 kernel 5.1.18-200.fc29.x86_64, and MLC ver 3.8 with App-Direct. Source: 2020ww18_CPX_BPS_DI. Tested by Intel, on 27 Apr 2020. New configuration: 1-node, 1x Intel® Xeon® pre-production CPX6 28C @ 2.9GHz processor on Cooper City with Single PMem module config (6x32GB DRAM; 1x{128GB,256GB,512GB} Intel Optane PMem 200 Series module at 15W), ucode pre-production running Fedora 29 kernel 5.1.18-200.fc29.x86_64, and MLC ver 3.8 with App-Direct. Source: 2020ww18_CPX_BPS_BG. Tested by Intel, on 31 Mar 2020.
5. Up to 1.87x higher AI Inference performance with 3rd Gen Intel® Xeon® Scalable processor supporting Intel® DL Boost with BF16 vs. prior generation using FP32 on ResNet50 throughput for image classification – New: 1-node, 4x 3rd Gen Intel® Xeon® Platinum 8380H processor (pre-production 28C, 250W) on Intel Reference Platform (Cooper City) with 384 GB (24 slots / 16GB / 3200) total memory, ucode 0x700001b, HT on, Turbo on, with Ubuntu 20.04 LTS, Linux 5.4.0-26,28,29-generic, Intel 800GB SSD OS Drive, ResNet-50 v 1.5 Throughput, <https://github.com/Intel-tensorflow/tensorflow/-b bf16/base, commit#828738642760358b388d8f615ded0c213f10c99a>, Modelzoo: <https://github.com/IntelAI/models/-b v1.6.1, Imagenet dataset, oneDNN 1.4, BF16, BS=56, 4 instances, 28-cores/instance, test by Intel on 5/18/2020>. Baseline: 1-node, 4x Intel® Xeon® Platinum 8280 processor on Intel Reference Platform (Lightning Ridge) with 768 GB (24 slots / 32 GB / 2933) total memory, ucode 0x4002f00, HT on, Turbo on, with Ubuntu 20.04 LTS, Linux 5.4.0-26,28,29-generic, Intel 800GB SSD OS Drive, ResNet-50 v 1.5 Throughput, <https://github.com/Intel-tensorflow/tensorflow/-b bf16/base, commit#828738642760358b388d8f615ded0c213f10c99a>, Modelzoo: <https://github.com/IntelAI/models/-b v1.6.1, Imagenet dataset, oneDNN 1.4, FP32, BS=56, 4 instances, 28-cores/instance, test by Intel on 5/18/2020>.
6. Up to 1.7x more AI training performance with 3rd Gen Intel® Xeon® Scalable processor supporting Intel® DL Boost with BF16 vs. prior generation on BERT throughput for natural language processing – New: 1-node, 4x 3rd Gen Intel® Xeon® Platinum 8380H processor (pre-production 28C, 250W) on Intel Reference Platform (Cooper City) with 384 GB (24 slots / 16GB / 3200) total memory, ucode 0x700001b, HT on, Turbo on, with Ubuntu 20.04 LTS, Linux 5.4.0-26,28,29-generic, Intel 800GB SSD OS Drive, BERT-Large (QA) Throughput, <https://github.com/Intel-tensorflow/tensorflow/-b bf16/base, commit#828738642760358b388d8f615ded0c213f10c99a>, Modelzoo: <https://github.com/IntelAI/models/-b v1.6.1, Squad 1.1 dataset, oneDNN 1.4, BF16, BS=12, test by Intel on 5/18/2020>. Baseline: 1-node, 4x Intel® Xeon® Platinum 8280 processor on Intel Reference Platform (Lightning Ridge) with 768 GB (24 slots / 32GB / 2933) total memory, ucode 0x4002f00, HT on, Turbo on, with Ubuntu 20.04 LTS, Linux 5.4.0-26,28,29-generic, Intel 800GB SSD OS Drive, BERT-Large (QA) Throughput, <https://github.com/Intel-tensorflow/tensorflow/-b bf16/base, commit#828738642760358b388d8f615ded0c213f10c99a>, Modelzoo: <https://github.com/IntelAI/models/-b v1.6.1, Squad 1.1 dataset, oneDNN 1.4, FP32, BS=12, test by Intel on 5/18/2020>.
7. Up to 1.9x higher AI inference performance with 3rd Gen Intel® Xeon® Scalable processor supporting Intel® DL Boost with BF16 vs. prior generation with FP32 on BERT throughput for natural language processing – New: 1-node, 4x 3rd Gen Intel® Xeon® Platinum 8380H processor (pre-production 28C, 250W) on Intel Reference Platform (Cooper City) with 384 GB (24 slots / 16GB / 3200) total memory, ucode 0x700001b, HT on, Turbo on, with Ubuntu 20.04 LTS, Linux 5.4.0-26,28,29-generic, Intel 800GB SSD OS Drive, BERT-Large (QA) Throughput, <https://github.com/Intel-tensorflow/tensorflow/-b bf16/base, commit#828738642760358b388d8f615ded0c213f10c99a>, Modelzoo: <https://github.com/IntelAI/models/-b v1.6.1, Squad 1.1 dataset, oneDNN 1.4, BF16, BS=32, 4 instances, 28-cores/instance, test by Intel on 5/18/2020>. Baseline: 1-node, 4x Intel® Xeon® Platinum 8280 processor on Intel Reference Platform (Lightning Ridge) with 768 GB (24 slots / 32 GB / 2933) total memory, ucode 0x4002f00, HT on, Turbo on, with Ubuntu 20.04 LTS, Linux 5.4.0-26,28,29-generic, Intel 800GB SSD OS Drive, BERT-Large (QA) Throughput, <https://github.com/Intel-tensorflow/tensorflow/-b bf16/base, commit#828738642760358b388d8f615ded0c213f10c99a>, Modelzoo: <https://github.com/IntelAI/models/-b v1.6.1, Squad 1.1 dataset, oneDNN 1.4, FP32, BS=32, 4 instances, 28-cores/instance, test by Intel on 5/18/2020>.
8. Process up to 1.98x more OLTP database transactions per minute with the new 3rd Gen Intel® Xeon® Scalable platform vs. 5-year old 4-socket platform – New: 1-node, 4x 3rd Gen Intel® Xeon® Platinum 8380H processor (pre-production 28C, 250W) on Intel Reference Platform (Cooper City) with 768 GB (24 slots / 32 GB / 3200) total memory, microcode 0x700001b, HT on, Turbo on, with Redhat 8.1, 4.18.0-147.3.1.el8_1.x86_64, 1x Intel 240GB SSD OS Drive, 2x6.4T P4610 for DATA, 2x3.2T P4610 for REDO, 1Gbps NIC, HammerDB 3.2, Popular Commercial Database, test by Intel on 5/13/2020. Baseline: 1-node, 4x Intel® Xeon® processor E7-8890 v3 on Intel Reference Platform (Brickland) with 1024 GB (64 slots / 16GB / 1600) total memory, microcode 0x16, HT on, Turbo on, with Redhat 8.1, 4.18.0-147.3.1.el8_1.x86_64, 1x Intel 800GB SSD OS Drive, 1x1.6T P3700 for DATA, 1x1.6T P3700 for REDO, 1Gbps NIC, HammerDB 3.2, Popular Commercial Database, test by Intel on 4/20/2020.
9. The 3rd Generation Intel Xeon Scalable platform will enable processing large datasets up to 31.2 Billion initial records with the SAP BW edition for SAP HANA V3 benchmark on an 8-socket single node using the latest 3rd Gen Intel® Xeon® Scalable processor and Intel® Optane™ PMem 200 series, and upcoming Linux OS release.
10. Up to 2.2x more Virtual Machines with the new 3rd Gen Intel® Xeon® Scalable platform and Intel® SSD Data Center Family vs. 5-year old 4-socket platform – New: 1-node, 4x 3rd Gen Intel® Xeon® Platinum 8380H processor (pre-production 28C, 250W) on Intel Reference Platform (Cooper City) with 1536 GB (48 slots / 32 GB / 3200 (@2933)) total memory, microcode 0x700001b, HT on, Turbo on, with RHEL-8.1 GA, 4.18.0-147.3.1.el8_1.x86_64, 1x Intel 240GB SSD OS Drive, 4x P4610 3.2TB PCIe NVMe, 4 x 40 GbE x710 dual port, Virtualization workload, Qemu-kvm 2.12 (inbox), WebSphere 8.5.5, DB2 v9.7, Nginx 1.14.1, test by Intel on 5/20/2020. Baseline: 1-node, 4x Intel® Xeon® processor E7-8890 v3 on Intel Reference Platform (Brickland) with 1024 GB (64 slots / 16GB / 1600) total memory, microcode 0x0000016, HT on, Turbo on, with RHEL-8.1 GA, 4.18.0-147.3.1.el8_1.x86_64, 1x Intel 240GB SSD OS Drive, 4x P3700 2TB PCIe NVMe, 4 x 40 GbE x710 dual port, Virtualization workload, Qemu-kvm 2.12 (inbox), WebSphere 8.5.5, DB2 v9.7, Nginx 1.14.1, test by Intel on 5/20/2020.

