Solution Brief

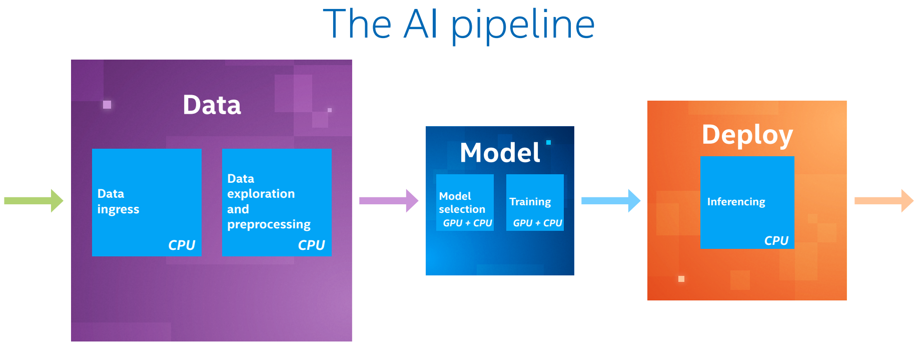
Artificial Intelligence (AI)
Intel® Advanced Matrix Extensions (Intel® AMX)

Accelerate Artificial Intelligence (AI) Workloads with Intel Advanced Matrix Extensions (Intel AMX)

Intel advances AI capabilities with 4th Gen Intel® Xeon® Scalable processors and Intel AMX, delivering 3x to 10x higher inference and training performance versus the previous generation.[1]

Optimizing the AI pipeline

Businesses can benefit from applying artificial intelligence (AI) in a variety of scenarios. These range from recommender systems for books and movies to retail digital software that drives large e-commerce sites to natural language processing (NLP) for chatbots and machine translation. But the attributes that make AI valuable—making sense of complex environments and massive datasets and solving previously impenetrable problems—have the potential to revolutionize business even further. According to one study, 90 percent of new enterprise application releases will include embedded AI functionality by 2025.[2]



The three outer boxes represent AI pipeline stages.
The five inner boxes represent AI workloads.
Box sizes indicate relative levels of processor activity within the AI pipline.

Figure 1. AI workloads and processor activity within the AI pipeline

To optimize AI pipelines, organizations can turn to 4th Gen Intel Xeon Scalable processors with Intel Advanced Matrix Extensions (Intel AMX), a built-in AI accelerator. Intel AMX was designed to balance inference, the most prominent use case for a CPU in AI applications, with more capabilities for training (see Figure 1).[3] With Intel Xeon Scalable processors representing 70 percent of the processor units (installed base) that are running AI inference workloads in the data center, selecting 4th Gen Intel Xeon Scalable processors with Intel AMX for new AI deployments is an efficient and cost-effective approach to accelerating AI workloads.[4]

The case for built-in accelerators

AI deployments powered by 3rd Gen Intel Xeon Scalable processors with Intel® Deep Learning Boost (Intel® DL Boost) allow IT teams to meet customer service-level agreements (SLAs) today. But 4th Gen Intel Xeon Scalable processors with Intel AMX change the game.

Figure 2 illustrates how Intel AMX delivers up to 5.7x–10x higher generation-to-generation PyTorch real-time inference performance gains, and Figure 3 illustrates how Intel AMX delivers up to 3.5x–10x higher generation-to-generation PyTorch training gains.[5] With improved performance, Intel AMX can help turn satisfied customers into delighted customers. The built-in Intel AMX accelerator simplifies the choice of CPU for AI applications by packaging significant performance gains in a solution with which organizations are already familiar.

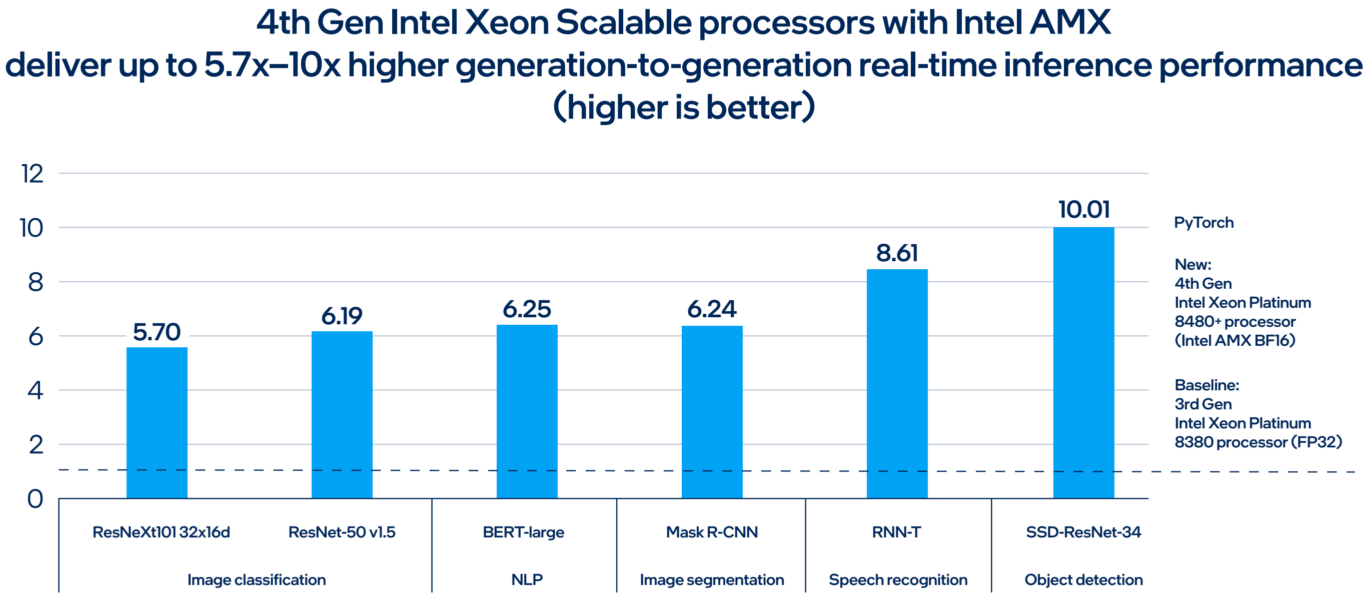


Figure 2. 4th Gen Intel Xeon Scalable processors with Intel AMX accelerate PyTorch real-time inference performance[5]

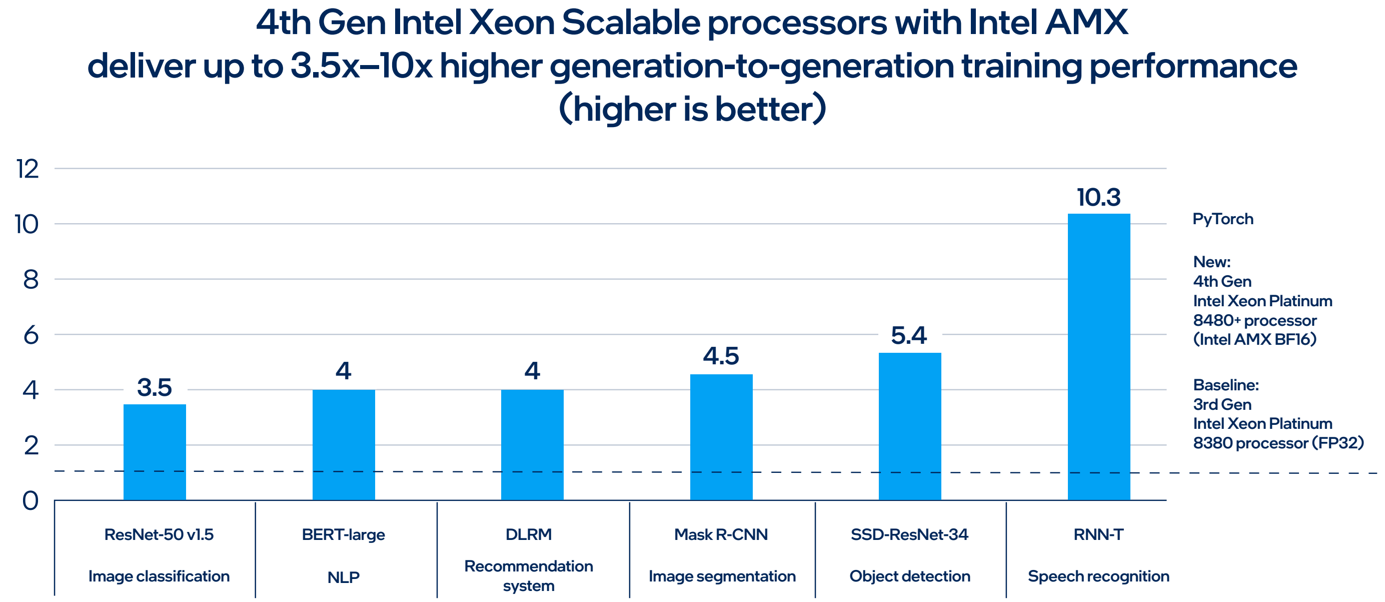


Figure 3. 4th Gen Intel Xeon Scalable processors with Intel AMX accelerate PyTorch training performance[5]

Figure 4 shows how Intel AMX delivers performance proportionally greater than the incremental core count for each generation, starting with 1st Gen Intel Xeon Scalable processors.

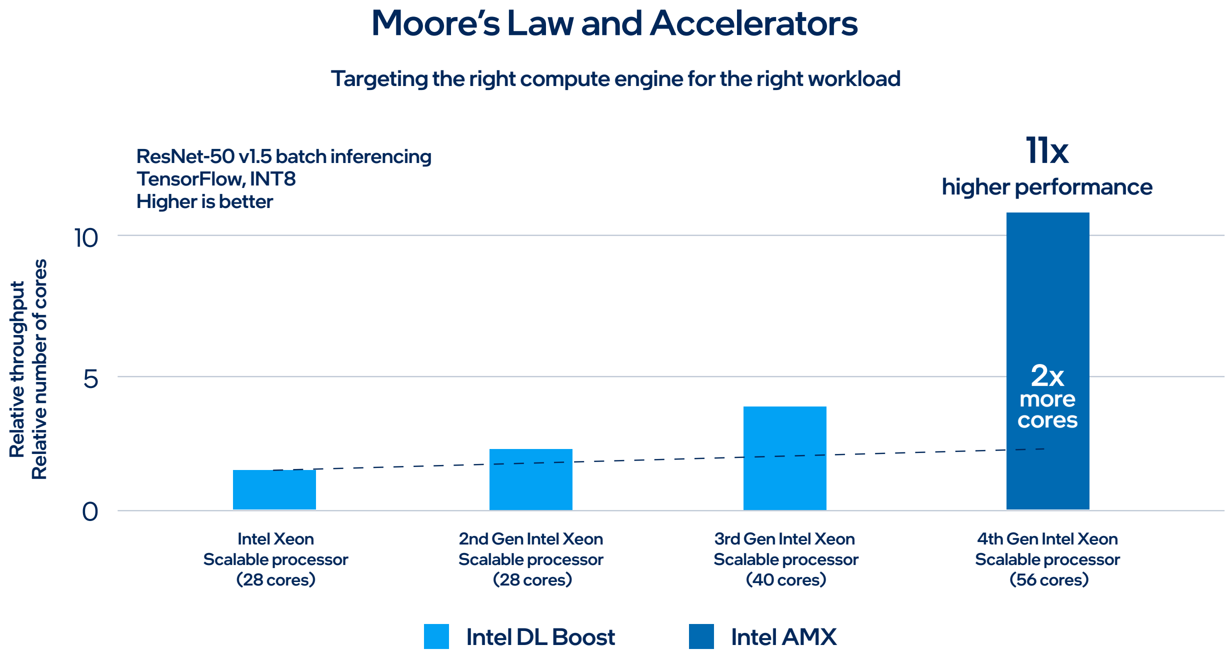


Figure 4. Using the 1st Gen Intel Xeon Scalable processor as a baseline, Intel AMX delivers a non-linear performance improvement compared to previous generations[6]

What is Intel AMX?

Intel AMX is a built-in accelerator that enables 4th Gen Intel Xeon Scalable processors to optimize deep learning (DL) training and inferencing workloads. With Intel AMX, 4th Gen Intel Xeon Scalable processors can quickly pivot between optimizing general computing and AI workloads. Imagine an automobile that could excel at city driving and then quickly change to deliver Formula 1 racing performance. 4th Gen Intel Xeon Scalable processors deliver this type of flexibility. Developers can code AI functionality to take advantage of the Intel AMX instruction set, and they can code non-AI functionality to use the processor instruction set architecture (ISA). Intel has integrated the Intel® oneAPI Deep Neural Network Library (oneDNN), its oneAPI DL engine, into popular open source tools for AI applications, including TensorFlow, PyTorch, PaddlePaddle, and ONNX.

Intel AMX architecture

Intel AMX architecture consists of two components (see Figure 5):

- The first component is tiles. Tiles consist of eight two-dimensional registers, each 1 kilobyte in size. They store large chunks of data.
- The second component is Tile Matrix Multiplication (TMUL); TMUL is an accelerator engine attached to the tiles that performs matrix-multiply computations for AI.

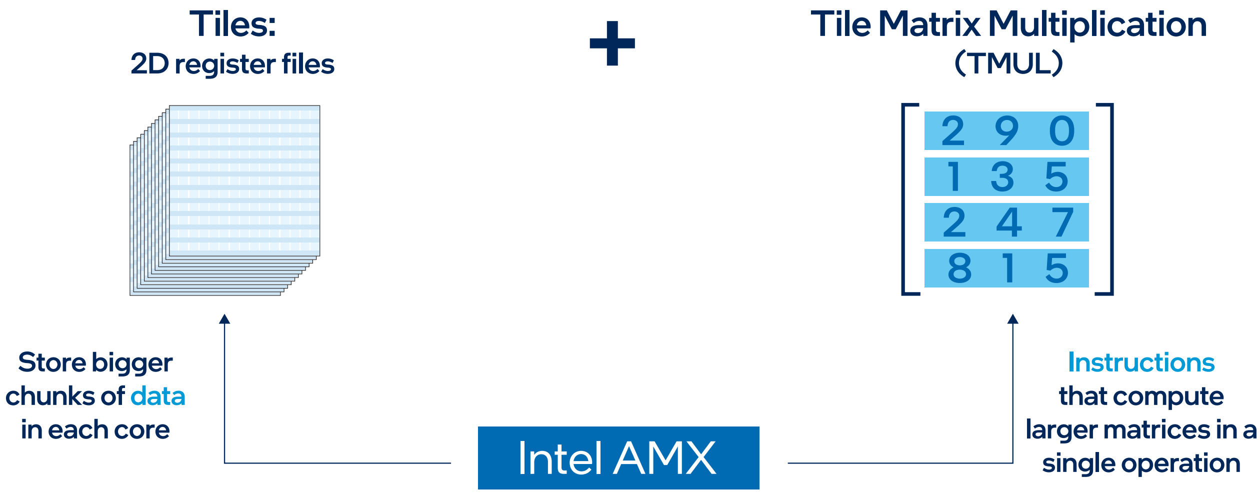

Tiles:
2D register files
+
Tile Matrix Multiplication
(TMUL)
2 9 0
1 3 5
2 4 7
8 1 5
Store bigger chunks of data in each core
Intel AMX
Instructions that compute larger matrices in a single operation

Figure 5. Intel AMX architecture consists of 2D register files (tiles) and TMUL

Intel AMX supports two data types, INT8 and BF16, for the matrix multiplication required for AI workloads:

- INT8 is a data type used for inferencing when the precision of FP32, a single-precision floating-point format often used in AI, isn't needed. Because the INT8 data type is lower precision, more INT8 operations can be processed per compute cycle.
- BF16 is a data type that delivers sufficient accuracy for most training. It can also deliver higher accuracy for inferencing if needed.

With this new tiled architecture, Intel AMX generation-on-generation performance gains are significant. Compared to 3rd Gen Intel Xeon Scalable processors running Intel® Advanced Vector Extensions 512 Neural Network Instructions (Intel® AVX-512 VNNI), 4th Gen Intel Xeon Scalable processors running Intel AMX can perform 2,048 INT8 operations per cycle, rather than 256 INT8 operations per cycle. They can also perform 1,024 BF16 operations per cycle, as compared to 64 FP32 operations per cycle, as shown in Figure 6.[7]

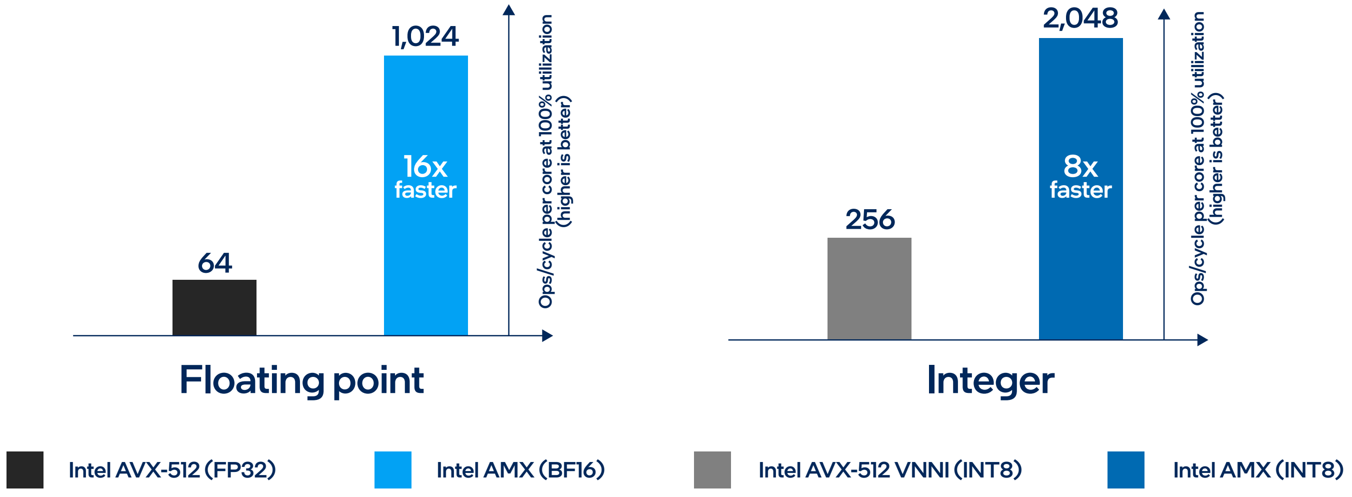


Figure 6. Intel AMX offers better performance than Intel AVX-512 VNNI for INT8 and BF16 data types[7]

AI use cases

4th Gen Intel Xeon Scalable processors with Intel AMX can be deployed in a wide range of DL use cases.

Recommender systems
Deliver a customized end-user experience, whether recommending movies and books or showing targeted ads. Create a DL-based recommender system that accounts for real-time user behavior signals and context features such as time and location in near real time.

Natural language processing (NLP)
With a global market projected to reach 80.68 billion USD by 2026,[8] NLP applications, including language inferencing and machine learning (ML), are critical for businesses to support and scale a variety of functions including sentiment analysis, chatbots, and machine translation.

Retail e-commerce software solutions
Grow revenue and deliver an exceptional customer experience by minimizing transaction time and keeping up with peak demands using DL inference and training, in addition to AI-optimized frameworks like PyTorch and TensorFlow.

Get started with Intel AMX

Near zero effort is required to improve performance with Intel AMX. This is because default frameworks are optimized with Intel oneDNN. Windows and Linux operating systems, kernel-based virtual machines (KVM), and popular hypervisors expose the Intel AMX instruction set. INT8 and BF16 operations are automatically optimized in open source frameworks like TensorFlow and PyTorch. The Intel® Distribution of OpenVINO™ toolkit allows developers to automate, optimize, tune, and run AI inferencing with little or no coding knowledge. The only thing developers need to do is to quantize training models to the INT8 data type using the Intel® Neural Compressor.

Accelerate AI with Intel Xeon Scalable processors

Harness the untapped potential of AI for business by moving to 4th Gen Intel Xeon Scalable processors with Intel AMX. Experience exceptional AI training and inference performance with all-new accelerated matrix-multiply operations while building on the broad foundation of Intel Xeon Scalable processors already in the data center.

Learn more about Intel AI and Intel AMX: intel.com/ai

[1] See [A16, A17] at intel.com/processorclaims: 4th Gen Intel Xeon Scalable processors. Results may vary.

[2] Forbes. "Top Artificial Intelligence (AI) Predictions For 2020 From IDC and Forrester." November 2019. forbes.com/sites/gilpress/2019/11/22/top-artificial-intelligence-ai-predictions-for-2020-from-idc-and-forrester/#4fef9821315a.

[3] The Next Platform. "With AMX, Intel Adds AI/ML Sparkle to Sapphire Rapids." August 2021. nextplatform.com/2021/08/19/with-amx-intel-adds-ai-ml-sparkle-to-sapphire-rapids/.

[4] Based on Intel market modeling of the worldwide installed base of data center servers running AI inference workloads as of December 2021.

[5] PyTorch model performance configurations. **PT-NLP BERT-large: 8480**: 1-node, pre-production platform with 2 x Intel Xeon Platinum 8480 processors with 1,024 GB (16 slots/64 GB/DDR5-4800) total memory, ucode 0x2b0000a1, Intel® Hyper-Threading Technology (Intel® HT Technology) on, Intel® Turbo Boost Technology on, CentOS Stream 8, 5.15.0, 1 x 1 TB Intel SSDSC2KW256G8 (PT)/Samsung SSD 860 EVO (TF), BERT-large, Inf: SQuAD1.1 (seq len=384), bs=1 (4 cores/instance), bs=n (1 socket/instance), bs: FP32=1,56, Intel AMX BF16=1,16, Intel AMX INT8=1,56, Trg: Wikipedia 2020/01/01 (seq len=512), bs: FP32=28, Intel AMX BF16=56 (1 instance, 1 socket), framework: https://github.com/intel-innersource/frameworks.ai.pytorch.private-cpu/tree/d7607bdd983093396a70713344828a989b766a66; ModelZoo: https://github.com/IntelAI/models/tree/spr-launch-public, PT:1.13, IPEX: 1.13, oneDNN: v2.7, tested by Intel on 10/24/2022. **8380**: 1-node, 2 x Intel Xeon Platinum 8380 processor with 1,024 GB (16 slots/64 GB/DDR4-3200) total memory, ucode 0xd000375, Intel HT Technology on, Intel Turbo Boost Technology on, Ubuntu 22.04 LTS, 5.15.0-27-generic, 1 x Intel SSDSC2KG960G8, BERT-large, Inf: SQuAD1.1 (seq len=384), bs=1 (4 cores/instance), bs=n (1 socket/instance), bs: FP32=1,56, INT8=1,56, Trg: Wikipedia 2020/01/01 (seq len=512), bs: FP32=28, Intel AMX BF16=56 (1 instance, 1 socket), framework: https://github.com/intel-innersource/frameworks.ai.pytorch.private-cpu/tree/d7607bdd983093396a70713344828a989b766a66; ModelZoo: https://github.com/IntelAI/models/tree/spr-launch-public, PT:1.13, IPEX: 1.13, oneDNN: v2.7, tested by Intel on 10/24/2022. **PT-DLRM: 8480**: 1-node, pre-production platform with 2 x Intel Xeon Platinum 8480 processor with 1,024 GB (16 slots/64 GB/DDR5-4800) total memory, ucode 0x2b0000a1, Intel HT Technology on, Intel Turbo Boost Technology on, CentOS Stream 8, 5.15.0, 1 x 1 TB Intel SSDSC2KW256G8 (PT)/Samsung SSD 860 EVO (TF), DLRM, inference: bs=n (1 socket/instance), bs: FP32=128, Intel AMX BF16=128, Intel AMX INT8=128, training bs:fp32/Intel AMX BF16=32K (1 instance, 1 socket), Criteo terabyte dataset, framework: https://github.com/intel-innersource/frameworks.ai.pytorch.private-cpu/tree/d7607bdd983093396a70713344828a989b766a66; ModelZoo: https://github.com/IntelAI/models/tree/spr-launch-public, PT:1.13, IPEX: 1.13, oneDNN: v2.7, tested by Intel on 10/24/2022. **8380**: 1-node, 2 x Intel Xeon Platinum 8380 processor with 1,024 GB (16 slots/64 GB/DDR4-3200) total memory, ucode 0xd000375, Intel HT Technology on, Intel Turbo Boost Technology on, Ubuntu 22.04 LTS, 5.15.0-27-generic, 1 x Intel SSDSC2KG960G8, DLRM, inference: bs=n (1 socket/instance), bs: FP32=128, INT8=128, training bs: FP32=32K (1 instance, 1 socket), Criteo terabyte dataset, framework: https://github.com/intel-innersource/frameworks.ai.pytorch.private-cpu/tree/d7607bdd983093396a70713344828a989b766a66; ModelZoo: https://github.com/IntelAI/models/tree/spr-launch-public, PT:1.13, IPEX: 1.13, oneDNN: v2.7, tested by Intel on 10/24/2022. **PT-ResNet-34: 8480**: 1-node, pre-production platform with 2 x Intel Xeon Platinum 8480 processor with 1,024 GB (16 slots/64 GB/DDR5-4800) total memory, ucode 0x2b0000a1, Intel HT Technology on, Intel Turbo Boost Technology on, CentOS Stream 8, 5.15.0, 1 x 1 TB Intel SSDSC2KW256G8 (PT)/Samsung SSD 860 EVO (TF), SSD-ResNet-34, inference: bs=1 (4 cores/instance), bs=n (1 socket/instance), bs: FP32=1,112, Intel AMX BF16=1,112, Intel AMX INT8=1,112, training bs: FP32/Intel AMX BF16=224 (1 instance, 1 socket), Coco 2017, framework: https://github.com/intel-innersource/frameworks.ai.pytorch.private-cpu/tree/d7607bdd983093396a70713344828a989b766a66; ModelZoo: https://github.com/IntelAI/models/tree/spr-launch-public, PT:1.13, IPEX: 1.13, oneDNN: v2.7, tested by Intel on 10/24/2022. **8380**: 1-node, 2 x Intel Xeon Platinum 8380 processor with 1,024 GB (16 slots/64 GB/DDR4-3200) total memory, ucode 0xd000375, Intel HT Technology on, Intel Turbo Boost Technology on, Ubuntu 22.04 LTS, 5.15.0-27-generic, 1 x Intel SSDSC2KG960G8, SSD-ResNet-34, inference: bs=1 (4 cores/instance), bs=n (1 socket/instance), bs: FP32=1,112, INT8=1,112, training bs: FP32=224 (1 instance, 1 socket), Coco 2017, framework: https://github.com/intel-innersource/frameworks.ai.pytorch.private-cpu/tree/d7607bdd983093396a70713344828a989b766a66; ModelZoo: https://github.com/IntelAI/models/tree/spr-launch-public, PT:1.13, IPEX: 1.13, oneDNN: v2.7, tested by Intel on 10/24/2022. **PT-ResNet-50: 8480**: 1-node, pre-production platform with 2 x Intel Xeon Platinum 8480 processor with 1,024 GB (16 slots/64 GB/DDR5-4800) total memory, ucode 0x2b0000a1, Intel HT Technology on, Intel Turbo Boost Technology on, CentOS Stream 8, 5.15.0, 1 x 1 TB Intel SSDSC2KW256G8 (PT)/Samsung SSD 860 EVO (TF), ResNet-50 v1.5, inference: bs=1 (4 cores/instance), bs=n (1 socket/instance), bs: FP32=1,64, Intel AMX BF16=1,64, Intel AMX INT8=1,116, training bs: FP32, Intel AMX BF16=128 (1 instance, 1 socket), ImageNet (224 x224), framework: https://github.com/intel-innersource/frameworks.ai.pytorch.private-cpu/tree/d7607bdd983093396a70713344828a989b766a66; ModelZoo: https://github.com/IntelAI/models/tree/spr-launch-public, PT:1.13, IPEX: 1.13, oneDNN: v2.7, tested by Intel on 10/24/2022. **8380**: 1-node, 2 x Intel Xeon Platinum 8380 processor with 1,024 GB (16 slots/64 GB/DDR4-3200) total memory, ucode 0xd000375, Intel HT Technology on, Intel Turbo Boost Technology on, Ubuntu 22.04 LTS, 5.15.0-27-generic, 1 x Intel SSDSC2KG960G8, ResNet-50 v1.5, inference: bs=1 (4 cores/instance), bs=n (1 socket/instance), bs: FP32=1,64, INT8=1,116, training bs: FP32=128 (1 instance, 1 socket), ImageNet (224 x224), framework: https://github.com/intel-innersource/frameworks.ai.pytorch.private-cpu/tree/d7607bdd983093396a70713344828a989b766a66; ModelZoo: https://github.com/IntelAI/models/tree/spr-launch-public, PT:1.13, IPEX: 1.13, oneDNN: v2.7, tested by Intel on 10/24/2022. **PT-RNN-T: 8480**: 1-node, pre-production platform with 2 x Intel Xeon Platinum 8480 processor with 1,024 GB (16 slots/64 GB/DDR5-4800) total memory, ucode 0x2b0000a1, Intel HT Technology on, Intel Turbo Boost Technology on, CentOS Stream 8, 5.15.0, 1 x 1 TB Intel SSDSC2KW256G8 (PT)/Samsung SSD 860 EVO (TF), Resnext101 32x16d, inference: bs=1 (4 cores/instance), bs=n (1 socket/instance), bs: FP32=1,64, Intel AMX BF16=1,64, Intel AMX INT8=1,116, ImageNet, framework: https://github.com/intel-innersource/frameworks.ai.pytorch.private-cpu/tree/d7607bdd983093396a70713344828a989b766a66; ModelZoo: https://github.com/IntelAI/models/tree/spr-launch-public, PT:1.13, IPEX: 1.13, oneDNN: v2.7, tested by Intel on 10/24/2022. **8380**: 1-node, 2 x Intel Xeon Platinum 8380 processors with 1,024 GB (16 slots/64 GB/DDR4-3200) total memory, ucode 0xd000375, Intel HT Technology on, Intel Turbo Boost Technology on, Ubuntu 22.04 LTS, 5.15.0-27-generic, 1 x Intel SSDSC2KG960G8, Resnext101 32x16d, inference: bs=1 (4 cores/instance), bs=n (1 socket/instance), bs: FP32=1,64, INT8=1,116, ImageNet, framework: https://github.com/intel-innersource/frameworks.ai.pytorch.private-cpu/tree/d7607bdd983093396a70713344828a989b766a66; ModelZoo: https://github.com/IntelAI/models/tree/spr-launch-public, PT:1.13, IPEX: 1.13, oneDNN: v2.7, tested by Intel on 10/24/2022. **PT-ResNext101: 8480**: 1-node, pre-production platform with 2 x Intel Xeon Platinum 8480 processor with 1,024 GB (16 slots/64 GB/DDR5-4800) total memory, ucode 0x2b0000a1, Intel HT Technology on, Intel Turbo Boost Technology on, CentOS Stream 8, 5.15.0, 1 x 1 TB Intel SSDSC2KW256G8 (PT)/Samsung SSD 860 EVO (TF), Resnext101 32x16d, bs=n (1 socket/instance), inference: bs: FP32=1,64, Intel AMX BF16=1,64, Intel AMX INT8=1,116, framework: https://github.com/intel-innersource/frameworks.ai.pytorch.private-cpu/tree/d7607bdd983093396a70713344828a989b766a66; ModelZoo: https://github.com/IntelAI/models/tree/spr-launch-public, PT:1.13, IPEX: 1.13, oneDNN: v2.7, tested by Intel on 10/24/2022. **8380**: 1-node, 2 x Intel Xeon Platinum 8380 processor with 1,024 GB (16 slots/64 GB/DDR4-3200) total memory, ucode 0xd000375, Intel HT Technology on, Intel Turbo Boost Technology on, Ubuntu 22.04 LTS, 5.15.0-27-generic, 1 x Intel SSDSC2KG960G8, Resnext101 32x16d, bs=n (1 socket/instance), inference: bs: FP32=1,64, INT8=1,116, framework: https://github.com/intel-innersource/frameworks.ai.pytorch.private-cpu/tree/d7607bdd983093396a70713344828a989b766a66; ModelZoo: https://github.com/IntelAI/models/tree/spr-launch-public, PT:1.13, IPEX: 1.13, oneDNN: v2.7, tested by Intel on 10/24/2022. **PT-MaskRCNN: 8480**: 1-node, pre-production platform with 2 x Intel Xeon Platinum 8480 processor with 1,024 GB (16 slots/64 GB/DDR5-4800) total memory, ucode 0x2b0000a1, Intel HT Technology on, Intel Turbo Boost Technology on, CentOS Stream 8, 5.15.0, 1 x Intel SSDSC2KW256G8 (PT)/Samsung SSD 860 EVO (TF), MaskRCNN, inference: bs=1 (4 cores/instance), bs=n (1 socket/instance), bs: FP32=1,112, Intel AMX BF16=1,112, training bs: FP32/Intel AMX BF16=112 (1 instance, 1 socket), Coco 2017, framework: https://github.com/intel-innersource/frameworks.ai.pytorch.private-cpu/tree/d7607bdd983093396a70713344828a989b766a66; ModelZoo: https://github.com/IntelAI/models/tree/spr-launch-public, PT:1.13, IPEX: 1.13, oneDNN: v2.7, tested by Intel on 10/24/2022. **8380**: 1-node, 2 x Intel Xeon Platinum 8380 processor with 1,024 GB (16 slots/64 GB/DDR4-3200) total memory, ucode 0xd000375, Intel HT Technology on, Intel Turbo Boost Technology on, Ubuntu 22.04 LTS, 5.15.0-27-generic, 1 x Intel SSDSC2KG960G8, MaskRCNN, inference: bs=1 (4 cores/instance), bs=n (1 socket/instance), bs: FP32=1,112, training bs: FP32=112 (1 instance, 1 socket), Coco 2017, framework: https://github.com/intel-innersource/frameworks.ai.pytorch.private-cpu/tree/d7607bdd983093396a70713344828a989b766a66; ModelZoo: https://github.com/IntelAI/models/tree/spr-launch-public, PT:1.13, IPEX: 1.13, oneDNN: v2.7, tested by Intel on 10/24/2022. Inference: ResNet-50 v1.5: ImageNet (224 x224), SSD ResNet-34: Coco 2017 (1200 x1200), BERT-large: SQuAD1.1 (seq len=384), Resnext101: ImageNet, Mask RCNN: COCO 2017, DLRM: Criteo terabyte dataset, RNNT: LibriSpeech. Training: ResNet-50 v1.5: ImageNet (224 x224), SSD ResNet-34: COCO 2017, BERT-large: Wikipedia 2020/01/01 (seq len =512), DLRM: Criteo terabyte dataset, RNNT: LibriSpeech, Mask RCNN: COCO 2017.

[6] **Software configuration for INT8 measurements**: TensorFlow ResNet-50 v1.5, inference: BS=116 (INT8), 1 instance/socket. oneDNN v2.7, Intel optimized TensorFlow 2.10. Tested by Intel on 10/24/2022. (3rd and 4th Gen Intel Xeon Scalable processors) and 7/19/2022 (2nd and 1st Gen Intel Xeon Scalable processors). **Hardware configurations: 4th Gen Intel Xeon Scalable processor hardware configuration (measured)**: Pre-production platform with 2S Intel Xeon Platinum 8480 processor (56 cores, 350 W thermal design power [TDP]) with 1 TB (8 channels/64 GB/4,800 MHz) total DDR5 memory, using BKC 01, using Intel AMX/INT8 and BF16, CentOS Stream 8, Intel AMX kernels (5.15), measurements will vary. **3rd Gen Intel Xeon Scalable processor hardware configuration (measured)**: 1 node, 2 x Intel Xeon Platinum 8380 processor (40 cores/2.3 GHz, 270 W TDP) processor with 1 TB (8 slots/64 GB/3,200 MHz) total DDR4 memory, ucode 0xd0002f2, Intel HT Technology on, Intel Turbo Boost Technology on, Ubuntu 20.04.2 LTS (Focal Fossa), 5.4.0-73-generic, 1 x Intel SSDSC2CW480A3 OS drive. **2nd Gen Intel Xeon Scalable processor hardware configuration (measured)**: 1 node, 2-socket Intel Xeon Platinum 8280 processor, 28 cores, Intel HT Technology on, Intel Turbo Boost Technology on, 384 GB total memory (12 slots/32 GB/2,933 MHz), BIOS: SE5C620.86B.02.01.0013.12152020065 (ucode: 0x500320a), CentOS Stream 8, 4.18.0-383.el8.x86_64. **Intel Xeon Scalable processor hardware configuration (measured)**: 1 node, 2-socket Intel Xeon Platinum 8180 processor, 28 cores, Intel HT Technology on, Intel Turbo Boost Technology on, 384 GB total memory (12 slots/32 GB/2,666 MHz), BIOS: SE5C620.86B.0X.01.0117.021220182317 (ucode: 0x2006b06), Ubuntu 20.04.2 LTS, 5.4.0-73-generic.

[7] Based on peak architectural capability of matrix multiply + accumulate operations per cycle per core assuming 100 percent CPU utilization. As of August 2021. For full workloads and configuration details, visit www.intel.com/PerformanceIndex (Architecture Day 2021). Results may vary.

[8] The global NLP market size source: Fortune Business Insights. "Natural Language Processing (NLP) Market Size, Share & COVID-19 Impact Analysis, By Deployment (On-Premises, Cloud, Hybrid), By Enterprise Size (SMEs, and Large Enterprises), By Technology (Interactive Voice Response (IVR), Optical Character Recognition (OCR), Text Analytics, Speech Analytics, Classification and Categorization), By Industry Vertical (Healthcare, Retail, High Tech, and Telecom, BFSI) and Regional Forecast, 2021-2028." June 2021. fortunebusinessinsights.com/industry-reports/natural-language-processing-nlp-market-101933#.

Performance varies by use, configuration and other factors. Learn more at 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.

Printed in USA 1222/SO/PRW/PDF Please Recycle 353107-001US