December 16, 2022

New Intel oneAPI 2023 Tools Maximize Value of Upcoming Intel Hardware

New capabilities in Intel tools expand cross-platform productivity.

SANTA CLARA, Calif.--(BUSINESS WIRE)-- **What's New:** Today, Intel announced the 2023 release of the Intel® oneAPI tools – available in the Intel® Developer Cloud and rolling out through regular distribution channels. The new oneAPI 2023 tools support the upcoming 4th Gen Intel® Xeon® Scalable processors, Intel® Xeon® CPU Max Series and Intel® Data Center GPUs, including Flex Series and the new Max Series. The tools deliver performance and productivity enhancements, and also add support for new Codeplay[1] plug-ins that make it easier than ever for developers to write SYCL code for non-Intel GPU architectures. These standards-based tools deliver choice in hardware and ease in developing high-performance applications that run on multiarchitecture systems.

This press release features multimedia. View the full release here: https://www.businesswire.com/news/home/20221216005068/en/

The oneAPI multiarchitecture open accelerated computing. (Credit: Intel Corporation)

> *"We're seeing encouraging early application performance results on our development systems using Intel Max Series GPU accelerators – applications built with Intel's oneAPI compilers and libraries. For leadership-class computational science, we value the benefits of code portability from multivendor, multiarchitecture programming standards such as SYCL and Python AI frameworks such as PyTorch, accelerated by Intel libraries. We look forward to the first exascale scientific discoveries from these technologies on the Aurora system next year."*
> –Timothy Williams, deputy director, Argonne Computational Science Division

What oneAPI Tools Deliver: Intel's 2023 developer tools include a comprehensive set of the latest compilers and libraries, analysis and porting tools, and optimized artificial intelligence (AI) and machine learning frameworks to build high-performance, multiarchitecture applications for CPUs, GPUs and FPGAs, powered by oneAPI. The tools enable developers to quickly meet performance objectives and save time by using a single codebase, allowing more time for innovation.

This new oneAPI tools release helps developers take advantage of the advanced capabilities of Intel hardware:

- 4th Gen Intel Xeon Scalable and Xeon CPU Max Series processors with Intel® Advanced Matrix Extensions (Intel® AMX), Intel® Quick Assist Technology (Intel® QAT), Intel® AVX-512, bfloat16 and more.
- Intel® Data Center GPUs, including Flex Series with hardware-based AV1 encoder,

and Max Series GPUs with data type flexibility, Intel® X^e Matrix Extensions (Intel® XMX), vector engine, Intel® X^e Link and other features.

Example benchmarks:

- MLPerf™ DeepCAM deep learning inference and training performance with Xeon Max CPU showed a 3.6x performance gain over Nvidia at 2.4 and AMD as the baseline 1.0 using Intel® AMX enabled by the Intel® oneAPI Deep Neural Network Library (oneDNN).[2]
- LAMMPS (large-scale atomic/molecular massively parallel simulator) workloads running on Xeon Max CPU with kernels offloaded to six Max Series GPUs and optimized by oneAPI tools resulted in an up to 16x performance gain over 3rd Gen Intel Xeon or AMD Milan alone.[3]

Advanced software performance:

- Intel® Fortran Compiler provides full Fortran language standards support up through Fortran 2018 and expands OpenMP GPU offload support, speeding development of standards-compliant applications.
- Intel® oneAPI Math Kernel Library (oneMKL) with extended OpenMP offload capability improves portability.
- Intel® oneAPI Deep Neural Network Library (oneDNN) enables 4th Gen Intel Xeon and Max Series CPU processors' advanced deep learning features including Intel® AMX, Intel AVX-512, VNNI and bfloat16.

To boost developer productivity, enriched SYCL support and robust code migration and analysis tools make it easier to develop code for multiarchitecture systems.

- The Intel® oneAPI DPC++/C++ Compiler adds support for new plug-ins from Codeplay Software for Nvidia and AMD GPUs to simplify writing SYCL code and extend code portability across these processor architectures. This provides a unified build environment with integrated tools for cross-platform productivity. As part of this solution, Intel and Codeplay will offer commercial priority support starting with the oneAPI plug-in for Nvidia GPUs.
- CUDA-to-SYCL code migration is now easier with more than 100 CUDA APIs added to the Intel® DPC++ Compatibility Tool, which is based on open source SYCLomatic.
- Users can identify MPI imbalances at scale with the Intel® VTune™ Profiler.
- Intel® Advisor adds automated roofline analysis for Intel Data Center GPU Max Series to identify and prioritize memory, cache or compute bottlenecks and causes, with actionable insights for optimizing data-transfer reuse costs of CPU-to-GPU offloading.

Why It Matters: With 48% of developers targeting heterogeneous systems that use more than one kind of processor[4], more efficient multiarchitecture programming is required to address the increasing scope and scale of real-world workloads. Using oneAPI's open, unified programming model with Intel's standards-based multiarchitecture tools provides freedom of choice in hardware, performance, productivity and code portability for CPUs and accelerators. Code written for proprietary programming models, like CUDA, lacks portability to other hardware, creating a siloed development practice that locks organizations into a closed ecosystem.

About oneAPI Ecosystem Adoption: Continued ecosystem adoption of oneAPI is ongoing with new Centers of Excellence being established. One, the Open Zettascale Lab at the University of Cambridge, is focused on porting significant exascale candidate codes to oneAPI, including CASTEP, FEniCS and AREPO. The center offers courses and workshops with experts teaching oneAPI methodologies and tools for compiling and porting code and optimizing performance. In total, 30 oneAPI Centers of Excellence have been established.

More Context: Codeplay Announces oneAPI for NVIDIA and AMD GPU Hardware | Intel oneAPI 2023 Tools Feature Details | Introducing Intel oneAPI 2023 Tools (Blog)| Intel oneAPI 2023 Release Preview (Webinar Recording) | oneAPI and Codeplay

About Intel

Intel (Nasdaq: INTC) is an industry leader, creating world-changing technology that enables global progress and enriches lives. Inspired by Moore's Law, we continuously work to advance the design and manufacturing of semiconductors to help address our customers' greatest challenges. By embedding intelligence in the cloud, network, edge and every kind of computing device, we unleash the potential of data to transform business and society for the better. To learn more about Intel's innovations, go to newsroom.intel.com and intel.com.

[1] Codeplay is an Intel company.

[2] MLPerf™ DeepCAM:

- Unverified performance Gains on MLPerf™ HPC-AI v 0.7 DeepCAM Training benchmark using optimized PyTorch 1.11. Result not verified by MLCommons Association. Unverified results have not been through an MLPerf™ review and may use measurement methodologies and/or workload implementations that are inconsistent with the MLPerf™ specification for verified results. The MLPerf™ name and logo are trademarks of MLCommons Association in the United States and other countries. All rights reserved. Unauthorized use strictly prohibited. See www.mlcommons.org for more information.
- AMD EPYC 7763: Test by Intel as of 04/07/2022. 1-node, 2x AMD EPYC 7763, HT On, Turbo Off, Total Memory 512 GB (16 slots/ 32 GB/ 3200 MHz, DDR4), BIOS AMI 1.1b, ucode 0xa001144, OS Red Hat Enterprise Linux 8.5 (Ootpa), kernel 4.18.0-348.7.1.el8_5.x86_64, compiler gcc (GCC) 8.5.0 20210514 (Red Hat 8.5.0-4), https://github.com/mlcommons/hpc/tree/main/deepcam, torch1.11.0a0+git13cdb98, torch-1.11.0a0+git13cdb98-cp38-cp38-linux_x86_64.whl, torch_ccl-1.2.0+44e473a-cp38-cp38-linux_x86_64.whl, intel_extension_for_pytorch-1.10.0+cpu-cp38-cp38-linux_x86_64.whl, Intel MPI 2021.5, Python3.8.
- Intel® Xeon® Scalable Processor 8380: Test by Intel as of 04/07/2022. 1-node, 2x Intel® Xeon® 8380 processor, HT On, Turbo Off, Total Memory 512 GB (16 slots/ 32 GB/ 3200 MHz, DDR4), BIOS SE5C6200.86B.0022.D64.2105220049, ucode 0xd0002b1, OS Red Hat Enterprise Linux 8.5 (Ootpa), kernel 4.18.0-348.7.1.el8_5.x86_64, compiler gcc (GCC) 8.5.0 20210514 (Red Hat 8.5.0-4), https://github.com/mlcommons/hpc/tree/main/deepcam, torch1.11.0a0+git13cdb98, torch-1.11.0a0+git13cdb98-cp38-cp38-linux_x86_64.whl, torch_ccl-1.2.0+44e473a-cp38-cp38-linux_x86_64.whl, intel_extension_for_pytorch-1.10.0+cpu-cp38-cp38-linux_x86_64.whl (AVX-512), Intel MPI 2021.5, Python3.8.

- Intel® Xeon® CPU Max Series (cache mode) AVX-512: Test by Intel as of 05/25/2022. 1-node, 2x Intel® Xeon® CPU Max Series, HT On,Turbo Off, Total Memory 128GB HBM and 1TB (16 slots/ 64 GB/ 4800 MHz, DDR5), Cluster Mode: SNC4, BIOS EGSDCRB1.86B.0080.D05.2205081330, ucode 0x8f000320, OS CentOS Stream 8, kernel 5.18.0-0523.intel_next.1.x86_64+server, compiler gcc (GCC) 8.5.0 20210514 (Red Hat 8.5.0-10, https://github.com/mlcommons/hpc/tree/main/deepcam, torch1.11.0a0+git13cdb98, AVX-512, FP32, torch-1.11.0a0+git13cdb98-cp38-cp38-linux_x86_64.whl, torch_ccl-1.2.0+44e473a-cp38-cp38-linux_x86_64.whl, intel_extension_for_pytorch-1.10.0+cpu-cp38-cp38-linux_x86_64.whl (AVX-512), Intel MPI 2021.5, Python3.8.
- NVIDIA A100 80GB: Test by Intel as of 04/13/2022. 1-node, 2x Intel® Xeon® Scalable 8360Y, HT On, Turbo On, Total Memory 256 GB (16 slots/ 16 GB/ 3200 MHz), NVIDIA A100 80GB PCIe, BIOS AMI 1.1b, ucode 0xd000311, OS Red Hat Enterprise Linux 8.4 (Ootpa), kernel 4.18.0-305.el8.x86_64, compiler gcc (GCC) 8.4.1 20200928 (Red Hat 8.4.1-1), https://github.com/mlcommons/hpc/tree/main/deepcam, pytorch1.11.0 py3.7_cuda11.3_cudnn8.2.0_0, cudnn 8.2.1, cuda11.3_0, intel-openmp 2022.0.1 h06a4308_3633, python3.7.
- Intel® Xeon® CPU Max Series (cache mode) BF16/AMX: Test by Intel as of 05/25/2022. 1-node, 2x Intel® Xeon® CPU Max Series , HT On, Turbo Off, Total Memory 128GB HBM and 1TB (16 slots/ 64 GB/ 4800 MHz, DDR5), Cluster Mode: SNC4, BIOS EGSDCRB1.86B.0080.D05.2205081330, ucode 0x8f000320, OS CentOS Stream 8, kernel 5.18.0-0523.intel_next.1.x86_64+server, compiler gcc (GCC) 8.5.0 20210514 (Red Hat 8.5.0-10), https://github.com/mlcommons/hpc/tree/main/deepcam, torch1.11.0a0+git13cdb98, AVX-512 FP32, torch-1.11.0a0+git13cdb98-cp38-cp38-linux_x86_64.whl, torch_ccl-1.2.0+44e473a-cp38-cp38-linux_x86_64.whl, intel_extension_for_pytorch-1.10.0+cpu-cp38-cp38-linux_x86_64.whl (AVX-512, AMX, BFloat16 Enabled), Intel MPI 2021.5, Python3.8.
- 4th Gen Intel® Xeon® Scalable Processors Mulit-Node cluster: Test by Intel as of 04/09/2022. 16-nodes Cluster, 1-node, 2x 4th Gen Intel® Xeon® Scalable Processor, HT On, Turbo On, Total Memory 256 GB (16 slots/ 16 GB/ 4800 MHz, DDR5), BIOS Intel SE5C6301.86B.6712.D23.2111241351, ucode 0x8d000360, OS Red Hat Enterprise Linux 8.4 (Ootpa), kernel 4.18.0-305.el8.x86_64, compiler gcc (GCC) 8.4.1 20200928 (Red Hat 8.4.1-1), https://github.com/mlcommons/hpc/tree/main/deepcam, torch1.11.0a0+git13cdb98 AVX-512, FP32, torch-1.11.0a0+git13cdb98-cp38-cp38-linux_x86_64.whl, torch_ccl-1.2.0+44e473a-cp38-cp38-linux_x86_64.whl, intel_extension_for_pytorch-1.10.0+cpu-cp38-cp38-linux_x86_64.whl (AVX-512), Intel MPI 2021.5, Python3.8.

[3] LAMMPS

- Intel® Xeon® CPU Max Series has 1.9x Higher LAMMPS (Geomean of Atomic Fluid, Copper, DPD, Liquid_crystal, Polyethylene, Protein, Stillinger-Weber, Tersoff, Water) performance vs. AMD EPYC 7773X.
- Intel® Xeon® CPU Max Series has 1.6x Higher LAMMPS (Geomean of Atomic Fluid, Copper, DPD, Liquid_crystal, Polyethylene, Protein, Stillinger-Weber, Tersoff, Water) performance vs. Intel® Xeon® 8380 Intel® Xeon® 8380: Test by Intel as of 10/11/2022. 1-node, 2x Intel® Xeon® 8380 CPU, HT On, Turbo On, NUMA configuration SNC2, Total Memory 256 GB (16x16GB 3200MT/s, Dual-Rank), BIOS Version SE5C620.86B.01.01.0006.2207150335, ucode revision=0xd000375, Rocky

Linux 8.6, Linux version 4.18.0-372.26.1.el8_6.crt1.x86_64, LAMMPS v2021-09-29 cmkl:2022.1.0, icc:2021.6.0, impi:2021.6.0, tbb:2021.6.0; threads/core:; Turbo:on; BuildKnobs:-O3 -ip -xCORE-AVX512 -g -debug inline-debug-info -qopt-zmm-usage=high;

- AMD EPYC 7773X: Test by Intel as of 10/6/2022. 1-node, 2x AMD EPYC , HT On, Turbo On, NUMA configuration NPS=4, Total Memory 256 GB (16x16GB 3200MT/s, Dual-Rank), BIOS Version M10, ucode revision=0xa001224, Rocky Linux 8.6 (Green Obsidian), Linux version 4.18.0-372.26.1.el8_6.crt1.x86_64, LAMMPS v2021-09-29 cmkl:2022.1.0, icc:2021.6.0, impi:2021.6.0, tbb:2021.6.0; threads/core:; Turbo:on; BuildKnobs:-O3 -ip -g -debug inline-debug-info -axCORE-AVX2 -march=core-avx2;
- Intel® Xeon® CPU Max Series: Test by Intel as of 8/31/2022. 1-node, 2x Intel® Xeon® CPU Max Series, HT ON, Turbo ON, NUMA configuration SNC4, Total Memory 128 GB (HBM2e at 3200 MHz), BIOS Version SE5C7411.86B.8424.D03.2208100444, ucode revision=0x2c000020, CentOS Stream 8, Linux version 5.19.0-rc6.0712.intel_next.1.x86_64+server, LAMMPS v2021-09-29 cmkl:2022.1.0, icc:2021.6.0, impi:2021.6.0, tbb:2021.6.0; threads/core:; Turbo:off; BuildKnobs:-O3 -ip -xCORE-AVX512 -g -debug inline-debug-info -qopt-zmm-usage=high.

[4] Evans Data Global Development Survey Report 22.1, June 2022

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

Performance results are based on testing as of dates shown in configurations and may not reflect all publicly available updates.

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.

View source version on businesswire.com:
https://www.businesswire.com/news/home/20221216005068/en/

Rebecca Want
1-650-919-4595
becky.want@ketchum.com

Source: Intel