June 22, 2023

Aurora Supercomputer Blade Installation Complete

The installation of Aurora's 10,624 blades containing the Intel Max Series product family marks a major milestone for exascale computing.

SANTA CLARA, Calif.--(BUSINESS WIRE)-- Intel (Nasdaq: INTC):

This press release features multimedia. View the full release here: https://www.businesswire.com/news/home/20230622628670/en/

On June 22, 2023, Argonne National Laboratory, Intel and HPE announced that the installation progress of the Aurora Supercomputer is complete. In this photo, members of the installation team walk through the many racks at Argonne National Laboratory. (Credit: Argonne National Laboratory)

What's New: The Aurora supercomputer at Argonne National Laboratory is now fully equipped with all 10,624 compute blades, boasting 63,744 Intel® Data Center GPU Max Series and 21,248 Intel® Xeon® CPU Max Series processors.

> *"Aurora is the first deployment of Intel's Max Series GPU, the biggest Xeon Max CPU-based system, and the largest GPU cluster in the world. We're proud to be part of this historic system and excited for the groundbreaking AI, science and engineering Aurora will enable."*
>
> —Jeff McVeigh, Intel corporate vice president and general manager of the Super Compute Group

What Aurora Is: A collaboration of Intel, Hewlett Packard Enterprise (HPE) and the Department of Energy (DOE), the Aurora supercomputer is designed to unlock the potential of the three pillars of high performance computing (HPC): simulations, data analytics and artificial intelligence (AI) on an extremely large scale. The system incorporates more than 1,024 storage nodes (using DAOS, Intel's distributed asynchronous object storage), providing 220 terabytes (TB) of capacity at 31TBs of total bandwidth, and leverages the HPE Slingshot high-performance fabric. Later this year, Aurora is expected to be the world's first supercomputer to achieve a theoretical peak performance of more than 2 exaflops (an exaflop is 10^{18} or a billion billion operations per second) when it enters the TOP500 list.

Aurora will harness the full power of the Intel Max Series GPU and CPU product family. Designed to meet the demands of dynamic and emerging HPC and AI workloads, early results with the Max Series GPUs demonstrate leading performance on real-world science and engineering workloads, showcasing up to 2 times the performance of AMD MI250X GPUs on OpenMC, and near linear scaling up to hundreds of nodes.[1] The Intel Xeon Max Series CPU drives a 40% performance advantage over the competition in many real-world

HPC workloads, such as earth systems modeling, energy and manufacturing.[2]

Why It Matters: From tackling climate change to finding cures for deadly diseases, researchers face monumental challenges that demand advanced computing technologies at scale. Aurora is poised to address the needs of the HPC and AI communities, providing the necessary tools to push the boundaries of scientific exploration.

“While we work toward acceptance testing, we're going to be using Aurora to train some large-scale open source generative AI models for science," said Rick Stevens, Argonne National Laboratory associate laboratory director. "Aurora, with over 60,000 Intel Max GPUs, a very fast I/O system, and an all-solid-state mass storage system, is the perfect environment to train these models.”

How It Works: At the heart of this state-of-the-art system are Aurora's sleek rectangular blades, housing processors, memory, networking and cooling technologies. Each blade consists of two Intel Xeon Max Series CPUs and six Intel Max Series GPUs. The Xeon Max Series product family is already demonstrating great early performance on Sunspot, the test bed and development system with the same architecture as Aurora. Developers are utilizing oneAPI and AI tools to accelerate HPC and AI workloads and enhance code portability across multiple architectures.

The installation of these blades has been a delicate operation, with each 70-pound blade requiring specialized machinery to be vertically integrated into Aurora's refrigerator-sized racks. The system's 166 racks accommodate 64 blades each and span eight rows, occupying a space equivalent to two professional basketball courts in the Argonne Leadership Computing Facility (ALCF) data center.

Researchers from the ALCF’s Aurora Early Science Program (ESP) and DOE’s Exascale Computing Project will migrate their work from the Sunspot test bed to the fully installed Aurora. This transition will allow them to scale their applications on the full system. Early users will stress test the supercomputer and identify potential bugs that need to be resolved before deployment. This includes efforts to develop generative AI models for science, recently announced at the ISC’23 conference.

More context: A Sneak Peek Inside Intel’s Supercomputing Lab (Video) | Intel Dives into the Future of Cooling | Broad, Open HPC+AI Portfolio Powers Performance, Generative AI for Science | Intel Data Center GPU Max Series (Product Brief) | Intel Xeon CPU Max Series (Product Brief) | Aurora: HPC and AI at Exascale, Blade Video (Intel Vision 2022)

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.

Disclaimers and configuration:

Intel does not control or audit third-party data. You should consult other sources to evaluate accuracy.

[1] Sunspot, Intel® Data Center GPU Max 1550: Testing as of 5/12/2023 by Argonne National Laboratory. Each Node: 2x 52C Intel® Xeon® Max CPU, 6x Intel® Data Center GPU Max Polaris, NVIDIA A100: Testing as of 5/12/2023 by Argonne National Laboratory. Each Node: 1x AMD EPYC Milan, 4x NVIDIA A100 40G PCIe Crusher, AMD Instinct MI250X: Testing as of 5/12/2023 by Argonne National Laboratory. Each Node: 1x Optimized 3rd Gen AMD EPYC, 4x AMD Instinct MI250X Intel does not control or audit third-party data. You should consult other sources to evaluate accuracy.

[2] Geomean of 5 workloads [HPCG, NEMO-GYRE, Anelastic Wave Propagation, BlackScholes, OpenFOAM]

- HPCG
    - Intel Configuration: 1-node, 2x Intel® Xeon® Max 9480, HT On, Turbo On, SNC4, Total Memory 128 GB (8x16GB HBM2 3200MT/s), 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, HPCG from MKL_v2022.1.0. Test by Intel as of 9/2/2022. O AMD Configuration: 1-node, 2x AMD EPYC 9654, HT On, Turbo On, CTDP=360W, NPS=4, 1536GB DDR5-4800, BIOS 1.2, microcode 0xa101111, Red Hat Enterprise Linux 8.7, Kernel 4.18, AMD official binary. Test by Intel as of 03/27/23.
- NEMO-GYRE
    - Intel Configuration: 1-node, 2x Intel® Xeon® Max 9480, 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, NEMO v4.2 build with Intel® Fortran Compiler Classic and Intel® MPI from 2022.3 Intel® oneAPI HPC Toolkit with compiler flags “-i4 -r8 -O3 -fno-alias -march=core-avx2 - fp-model fast=2 -no-prec-div -no-prec-sqrt -align array64byte -fimf-usesvml=true”. Test by Intel as of 10/12/2022. o AMD Configuration: 1-node, 2x AMD EPYC 9654, HT On, Turbo On, CTDP=360W, NPS=4, 1536GB DDR5-4800, BIOS 1.2, microcode 0xa101111, Red Hat Enterprise Linux 8.7, Kernel 4.18, NEMO version 4.2 compiled with Intel® Fortran Compiler Classic and Intel® MPI from 2022.3 Intel® oneAPI HPC Toolkit with compiler flags “-O3 -xCORE-AVX2”. Test by Intel as of 03/27/23.
- Anelastic Wave Propagation
    - Intel Configuration: 1-node, 2x Intel® Xeon® Max 9480, HT On, Turbo On, SNC4, 128 GB HBM2e, BIOS Version SE5C7411.86B.8424.D03.2208100444, ucode Performance Footnotes revision=0x2c000020, CentOS Stream 8, Linux version 5.19, YASK v3.05.07 . Test by Intel as of 9/2/2022. o AMD Configuration: 1-node, 2x AMD EPYC 9654, HT On, Turbo On, CTDP=360W, NPS=4, 1536GB DDR5-4800, BIOS 1.2, microcode 0xa101111, Red Hat Enterprise Linux 8.7, Kernel 4.18, YASK v3.05.07. Test by Intel as of 03/27/23.
- BlackScholes
    - Intel Configuration: 1-node, 2x Intel® Xeon® Max 9480,HT On, Turbo On, SNC4, Total Memory 128 GB (8x16GB HBM2 3200MT/s), 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, Black Scholes v1.4. Test by Intel as of 9/2/2022. o AMD Configuration: 1-node, 2x AMD EPYC 9654, HT On, Turbo On, CTDP=360W, NPS=4, 1536GB DDR5-4800, BIOS 1.2, microcode 0xa101111, Red Hat Enterprise Linux 8.7, Kernel 4.18, Black Scholes v1.4. Test by Intel as of 03/27/23.

- OpenFOAM
  - This offering is not approved or endorsed by OpenCFD Limited, producer and distributor of the OpenFOAM software via www.openfoam.com, and owner of the OPENFOAM® and OpenCFD® trademark
  - Intel Configuration: 1-node, 2x Intel® Xeon® Max 9480, HT On, Turbo On, SNC4, Total Memory 128 GB (8x16GB HBM2 3200MT/s), 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, OpenFOAM 8, Motorbike 20M @ 250 iterations, Motorbike 42M @ 250 iterations. Test by Intel as of 9/2/2022. o AMD Configuration: 1-node, 2x AMD EPYC 9654, HT On, Turbo On, CTDP=360W, NPS=4, 1536GB DDR5-4800, BIOS 1.2, microcode 0xa101111, Red Hat Enterprise Linux 8.7, Kernel 4.18, OpenFOAM 8, Motorbike 20M @ 250 iterations, Motorbike 42M @ 250 iterations. Test by Intel as of 03/27/23.

View source version on businesswire.com:
https://www.businesswire.com/news/home/20230622628670/en/

Bats Jafferji
1-603-809-5145
bats.jafferji@intel.com

Source: Intel