

HOW TO DELIVER HPC AND AI WORKLOADS AT LOW TCO

Discover the advantages of choosing Amazon EC2* instances featuring the Intel® Xeon® Scalable processor

Intel® Xeon® Scalable processors offer:

Up to



higher performance/\$ for HPC workloads¹

Save on software licenses by upgrading from older Intel[®] processors to newer Intel processors. **Same performance: up to**



fewer cores²

Achieve incredible performance per dollar on HPC and AI workloads with AWS* and Intel

The cloud puts powerful infrastructure at your disposal, so you can tackle the most challenging artificial intelligence (AI) and high-performance computing (HPC) tasks. You need to ensure, though, that you're using cloud instances that deliver the value you need, and support your workloads with built-in optimizations.

Choose the Intel® Xeon® Scalable processor to get the low total cost of ownership (TCO) you want. According to test results from two popular benchmarks (LAMMPS* and Highperformance Linpack*), Amazon EC2 instances featuring the Intel Xeon Scalable processor offer up to 4.15x higher performance per dollar¹.

If you're using a compute optimized instance, which AWS recommends for science and engineering applications, you can **cut the core count for a workload by up to 40**² if you switch from older c4.8xlarge instances to c5.4xlarge. If you're licensing software per core, there can be **huge cost savings**.

The Intel Xeon Scalable and 2nd Generation Intel® Xeon® Scalable processors have built-in features that help to accelerate AI and HPC workloads.

- **INT8** enables machine learning programs to be accelerated by reducing the precision of the calculations, discarding unnecessary detail.
- Intel[®] Advanced Vector Extensions 512 (Intel[®] AVX-512) offers 512-bit vector instructions to help accelerate applications that intensively use floating point calculations, including scientific simulations. The 2nd generation Intel Xeon Scalable processor



introduced a new instruction (Intel® Deep Learning Boost) that can speed up typical machine learning inference operations for applications including image classification, speech recognition, language translation, and object detection.

• Intel[®] Turbo Boost Technology enables you to run cores faster than the base operating frequency, so you can accelerate performance when you need it most.

Achieve incredible performance per dollar for your AI and HPC workloads by choosing an Amazon EC2 instance featuring the Intel Xeon Scalable processor. Find out more about building value on your critical workloads by reading this Intel eGuide.

¹ Results calculated by Intel using AWS pricing (\$/hour, standard 1-year term, no up-front) as of 12th January, 2019.

Performance per dollar testing done on Amazon EC2 M5 and M5a instances (<u>https://aws.amazon.com/ec2/instance-types/</u>), comparing 96 vCPU Intel[®] Xeon[®] Scalable processor performance per dollar to Competitor processor performance per dollar.

AWS pricing as of 12th January 2019

Standard 1-Year term Reserved Instance Pricing (<u>https://aws.amazon.com/ec2/pricing/reserved-instances/pricing/</u>) On Demand Linux/Unix Usage Pricing per hour (<u>https://aws.amazon.com/ec2/pricing/on-demand/</u>)

Workload: LAMMPS*

Results: Competitor processor performance per dollar = baseline of 1; Intel[®] Xeon[®] Scalable processor performance per dollar = 2.19X (higher is better) HPC Materials Science – LAMMPS (higher is better):

Amazon EC2 m5.24xlarge (Intel) Instance, LAMMPS version: 2018-08-22 (Code: <u>https://lammps.sandia.gov/download.html</u>), Workload: Water – 512K Particles, Intel ICC 18.0.3.20180410, Intel[®] MPI Library for Linux* OS, Version 2018 Update 3 Build 20180411, 48 MPI Ranks, RedHat* Enterprise Linux 7.5, Kernel 3.10.0-862.el7.x86_64, OMP_NUM_THREADS=2, Score 137.5 timesteps/sec, measured by Intel on 10/31/18.

Amazon EC2 m5a.24xlarge (Competitor processor) Instance, LAMMPS version: 2018-08-22 (Code: https://lammps.sandia.gov/download.html), Workload: Water – 512K Particles, Intel ICC 18.0.3.20180410, Intel® MPI Library for Linux* OS, Version 2018 Update 3 Build 20180411, 48 MPI Ranks, RedHat* Enterprise Linux 7.5, Kernel 3.10.0-862.el7.x86_64, OMP_NUM_THREADS=2, Score 55.8 timesteps/sec, measured by Intel on 11/7/18. Changes for Competitor processor to support AVX2 (Competitor processor only supports AVX2, so these changes were necessary): sed -i 's/-xHost/-xCORE-AVX2/g' Makefile.intel_cpu_intelmpi

sed -i 's/-qopt-zmm-usage=high/-xCORE-AVX2/g' Makefile.intel_cpu_intelmpi

Workload: High-performance Linpack*

Results: Competitor processor performance per dollar = baseline of 1; Intel[®] Xeon[®] Scalable processor performance per dollar = 4.15X (higher is better) HPC Linpack (higher is better):

Amazon EC2 m5.24xlarge (Intel) Instance, HP Linpack Version 2.2 (https://software.intel.com/en-us/articles/intel-mkl-benchmarks-suite Directory: benchmarks_2018.3.222/linux/mkl/benchmarks/mp_linpack/bin_intel/intel64), Intel ICC 18.0.3.20180410 with AVX512, Intel® MPI Library for Linux* OS, Version 2018 Update 3 Build 20180411, RedHat* Enterprise Linux 7.5, Kernel 3.10.0-862.el7.x86_64, OMP_NUM_THREADS=24, 2 MPI processes, Score 3152 GB/s, measured by Intel on 10/31/18.

Amazon EC2 m5a.24xlarge (Competitor processor) Instance, HP Linpack Version 2.2, (HPL Source: http://www.netlib.org/benchmark/hpl/hpl-2.2.tar.gz; Version 2.2; icc (ICC) 18.0.2 20180210 used to compile and link to BLIS library version 0.4.0; https://github.com/flame/blis; Addt'l Compiler flags: -O3 -funroll-loops -W -Wall –qopenmp; make arch=zen OMP_NUM_THREADS=8; 6 MPI processes.), Intel ICC 18.0.3.20180410 with AVX2, Intel® MPI Library for Linux* OS, Version 2018 Update 3 Build 20180411, RedHat* Enterprise Linux 7.5, Kernel 3.10.0-862.el7.x86_64, OMP_NUM_THREADS=8, 6 MPI processes, Score 677.7 GB/s, measured by Intel on 11/7/18.

² Migrating from c4.8xlarge to c5.4xlarge instances reduced the core count by 40 for a workload measured by TSO Logic. Source: TSO Logic / Intel research report: "New Advances by Intel, Amazon Web Services, Drive Major Cloud Savings"

https://www.intel.com/content/dam/www/public/us/en/documents/reports/tso-logic-research.pdf?spredfast-trk-id=sf215253580



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 the date set forth in the configurations and may not reflect all publicly available security updates. See configuration disclosure for details. No product or component can be absolutely secure.

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.

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. Check with your system manufacturer or retailer or learn more at intel.com

Intel, the Intel logo, and Xeon are trademarks of Intel Corporation or its subsidiaries in the U.S. and/or other countries.

*Other names and brands may be claimed as the property of others. © Intel Corporation 071

0719/AK/CAT/PDF