Brief

Intel® Data-Centric Technologies
Cloud Services

Amazon Elastic Compute Cloud (Amazon EC2) Cloud Computing Instances Powered by Intel

Cloud computing is currently an inherent piece of people's everyday lives. Daily activities such as banking, e-mail, media streaming, and e-commerce all rely on the cloud. Businesses are experiencing intense change, and they need to adopt new working models as economic uncertainties push cloud transformation to the forefront of the enterprise agenda. Navigating this changing environment requires optimized, agile, trusted, and scalable solutions to help ensure business continuity and long-term success. It's not surprising that these challenges are driving the adoption and integration of Amazon Web Services (AWS).

Collaborative solutions for today and tomorrow

Business needs are evolving fast, and enterprise infrastructure must be able to respond at equal speed. AWS and Intel have a 14+-year relationship dedicated to developing, building, and supporting cloud services that are designed to manage cost and complexity, accelerate business outcomes, and scale to meet current and future computing requirements. Intel processors provide the foundation of many cloud computing services deployed by AWS. Amazon Elastic Compute Cloud (Amazon EC2) instances powered by Intel® Xeon® processors have the largest breadth, global reach, and availability of compute instances across AWS geographies (24 regions and 76 availability regions at the time of this writing), and they are tailored to meet all your diverse computing needs. Whether deploying a hybrid cloud with VMware Cloud on AWS or AWS Outposts, SAP HANA in Amazon EC2 High Memory instances, or a multi-node high-performance computing (HPC) cluster with Amazon EC2 C5n instances, Intel architecture is at the heart of Amazon EC2 cloud computing instances, providing strong data protection, fast processing of large data volumes, and service flexibility without a hit to performance. Intel processors feature:

- **Intel Advanced Vector Extension 512 (Intel AVX-512)**, which offers accelerated application performance up to 2x better than previous-generation technologies, enabling significant improvements in workload speeds and data applications.[1]
- **Intel Deep Learning Boost (Intel DL Boost)** offers built-in AI acceleration, with up to 14x better inference performance on image classification in 2nd Generation Intel Xeon Scalable processors, compared to competing processors.[2]
- **Intel Turbo Boost Technology** accelerates processor and graphics performance for peak loads, automatically allowing processor cores to run faster than the rated operating frequency if they're operating below power, current, and temperature specification limits.
- **Intel AES New Instructions (Intel AES-NI)** improves upon the original Advanced Encryption Standard (AES) algorithm to provide faster data protection and greater security. All current-generation Amazon EC2 instances support this processor feature.

Figure 1. The AWS and Intel relationship extends across a wide range of cloud and Internet of Things (IoT) services

Delivering optimized performance and costs

The latest Amazon EC2 instances, featuring 2nd Generation Intel Xeon Scalable processors, deliver an optimal mix of price/performance across compute and memory compared to similar instances based on other processors. Because many workloads utilize only a fraction of the processor's maximum performance, these instances offer a better fit for those workloads (see Figure 1). Research performed by TSO Logic, now part of AWS, concluded that customers who upgrade their instances to current-generation Intel Xeon Scalable processors can reduce per-instance infrastructure costs by as much as 57 percent.[3] Intel works extensively with ISVs to test, optimize, and certify leading enterprise applications. And for enterprise workloads demanding massive scalability, AWS and Intel have collaborated to tailor design instances for specific use cases.

HPC

Further examples of Intel and AWS collaboration to deliver breakthrough performance innovations include the C5n and z1d instances, specifically tuned for HPC:

- **C5n instance:** In late 2019, Intel MPI Library—a multi-fabric message-passing library based upon an open industry initiative—added support for Amazon Elastic Fabric Adapter (EFA) to enhance the speed and efficiency of inter-node communications critical for scaling HPC workloads. The combination of Intel MPI and Amazon EFA enables you to achieve the application performance of an on-premises HPC cluster, with the scalability, flexibility, and elasticity provided by the AWS cloud. For more information, see intel.com/content/www/us/en/develop/articles/intel-mpi-library-support-for-amazon-web-services-elastic-fabric-adapter.html.
- **z1d instance:** Amazon EC2 z1d instances based on a customized Intel Xeon Scalable processor push the envelope of application performance and deliver high single-thread performance while boosting sustained all-core frequency of up to 4.0 GHz. z1d is the fastest of any Amazon EC2 cloud instance, providing both high compute performance and high memory, which is ideal for electronic design automation (EDA), gaming, optical proximity correction (computational lithography), HPC, financial usage, actuarial usage, data analytics, and certain relational database workloads with high per-core licensing costs. z1d is also ideal for applications requiring high single-threaded performance and high memory usage.

> “On average, the Intel counterparts provide higher value from anywhere between 1.25x all the way up to 4.1x with HPC.[4]”
>
> — Usman Pirzada
> Senior Editor, Hardware, Wccftech

Why choose Intel-based infrastructure as a service (IaaS)?[4]

1.25x Price/perf
Enterprise applications
SPECint

1.9x Price/perf
NoSQL database
MongoDB

4.1x Price/perf
High-performance computing (HPC)
High Perf. LINPACK (HPL)

Artificial intelligence (AI) and machine learning (ML)

The challenges associated with operating AI workloads at scale have been reduced in the cloud, and it's never been easier or more cost-effective to collect data, store it, and build custom ML and deep learning (DL) models. Amazon EC2 C5 instances, with the computational power of Intel Xeon Scalable processors with Intel AVX-512 and Intel DL Boost,[5] enable customers to create intelligent and innovative new products and experiences, powered by ML. AWS provides users access to Intel-optimized images for TensorFlow, Apache MXNet, and PyTorch, in addition to Intel performance libraries, to enhance application performance. Intel and AWS work together to make AI simple, accessible, and fast.

In-memory databases

Amazon EC2 High Memory instances meet the needs of massive in-memory databases, such as SAP solutions,[6] and they are the first-ever instances built on 8-socket hosts powered by Intel Xeon Scalable processors. These instances are offered as dedicated bare-metal hosts, and they take advantage of the innovative AWS Nitro System, which enhances security while offloading functions such as hypervisors, networking, and management capabilities so that practically all of the instance server resources can be used by your workloads. These instances set a new world record of 480,600 for the SAP Application Performance Standard (SAPS) when launched.[7]

Hybrid cloud

AWS Outposts

AWS Outposts, powered by Intel Xeon Scalable processors, extends Amazon EC2 C5, M5, R5, I3en, and G4dn instances on premises. With AWS Outposts, you can seamlessly extend your Amazon Virtual Private Cloud (Amazon VPC) on premises and run some AWS services locally. You can connect to a broad range of services available in the local AWS region, and you can also use the same AWS APIs, tools, and security controls to run, manage, and help secure your on-premises applications, just as you would in the cloud.

If you're already running your applications on Intel Xeon processor–based servers on premises and benefitting from Intel software optimizations and tuning for enterprise applications, you'll enjoy the same robust performance on AWS Outposts and the AWS cloud.

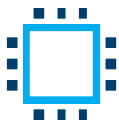
VMware Cloud on AWS

VMware Cloud on AWS is exclusively built on Intel Xeon Scalable processors, and it offers organizations a fast, security-enabled path to the cloud. With VMware Cloud on AWS, you can migrate data centers to the cloud for rapid data center evacuation, disaster recovery, and application modernization. VMware Cloud on AWS delivers on next-generation bare-metal instances, and it is powered by Intel Xeon Scalable processors, providing optimized, high input/output (I/O) instances and featuring low-latency NVM Express (NVMe)-based solid state drives (SSDs). For example, organizations can extend their data centers with AWS virtual desktop infrastructure (VDI) built on VMware Cloud on AWS to enable employees to work more securely from any location. Customers with VMware solutions on premises and VMware Cloud on AWS can take advantage of familiar features such as VMware vSphere vMotion for load balancing, zero downtime upgrades, and migration of workloads across instances between on premises and the cloud.

VMware Cloud on AWS provides different host types—I3, R5, and I3en—for use in your software-defined data center (SDDC). VMware Cloud on AWS has historically run on I3 and R5 metal instances. In July of 2020, the new I3en instance based on 2nd Generation Intel Xeon Scalable processors became available.

Zignal Labs performs next-level sentiment analysis using Amazon SageMaker and Amazon EC2

The new Zignal Labs sentiment pipeline is delivering results that show at least **30 percent improvement** in precision compared to prior methods, helping the company win and retain customers while **reducing development and operations costs by up to 90 percent.**



General-purpose computing

Amazon EC2 M5, M5n, and M5dn instances

M5 instances are general-purpose computing instances powered by 1st or 2nd Generation Intel Xeon Scalable processors with a sustained all-core turbo clock speed of 3.5 GHz and a maximum single-core clock speed of 3.5 GHz with Intel Turbo Boost Technology. 2nd Generation Intel Xeon Scalable processors deliver built-in acceleration for DL workloads with Intel DL Boost

Benefits of Intel-based AWS EC2 instances[8]

General purpose

T3
Burstable CPU usage
SKX—up to 8 vCPUs

M5
Non-burstable CPU usage
SKX—up to 96 vCPUs

M5D
M5 with local host-attached NVMe SSDs
SKX—up to 96 vCPUs

Compute optimized

C5
High-performance low price/compute ratio
SKX—up to 36 vCPUs
CLX—up to 96 vCPUs

C5D
C5 with local host-attached NVMe SSDs
SKX—up to 36 vCPUs
CLX—up to 96 vCPUs

C5n
C5 with up to 100 Gbps network bandwidth
SKX—up to 36 vCPUs

Memory optimized

R5
Up to 768 GiB RAM
SKX or CLX
Up to 96 vCPUs

X1, X1e
Lowest price/GiB RAM
Up to 3.9 TB RAM
HSX—up to 128 vCPUs

High Memory
Most memory—up to 24 TB
HSX—up to 448 vCPUs

Z1D
High compute capacity and high memory
Up to 384 GiB RAM
SKX—up to 48 vCPUs

Storage optimized

I3en
Up to 60 TB NVMe SSD bare-metal option
SKX or CLX
Up to 96 vCPUs

H1
Compute and memory balanced, up to 16 TB HDD
BDX—up to 64 vCPUs

D2
Up to 48 TB HDD storage, lowest price/disk throughput
HSX—up to 36 vCPUs

Vector Neural Network Instructions (VNNI). M5n and M5dn instance variants are ideal for applications that can take advantage of improved network throughput and packet-rate performance. This family provides a balance of compute, memory, and network resources, and it is a good choice for many applications.

Use cases

M5, M5n, and M5dn are ideal for a broad range of workloads, including business-critical applications, web and application servers, small and mid-sized databases, data processing tasks that require additional memory, caching fleets, and for running back-end servers for SAP, Microsoft SharePoint, cluster computing, and other enterprise applications.

Amazon EC2 T3 instances

Amazon EC2 T3 instances are a low-cost Amazon EC2 instance type designed to provide a baseline level of CPU performance with the ability to burst above the baseline based on CPU credits. T3 instances are powered by Intel Xeon Scalable processors with sustained all-core turbo CPU clock speed of 2.5 GHz and Intel AVX-512, delivering a burstable general-purpose instance type that provides a baseline level of CPU performance with the ability to burst CPU usage at any time for as long as required. T3 instances accumulate CPU credits when a workload is operating below a baseline threshold. Each earned CPU credit provides the T3 instance the opportunity to burst with the performance of a full CPU core for one minute when needed. T3 instances start in unlimited mode by default, giving users the ability to sustain high CPU performance over any desired timeframe while keeping costs as low as possible.

Use cases

T3 instances offer a balance of compute, memory, and network resources, and they are designed for applications with moderate CPU usage that experience temporary spikes in use including microservices, low-latency interactive applications, small and medium databases, virtual desktops, development environments, code repositories, and business-critical applications.

Compute-optimized computing

Amazon EC2 C5 and C5n instances

Compute-optimized instances are ideal for compute-bound applications that benefit from high-performance Intel Xeon Scalable processors and 2nd Generation Intel Xeon Scalable processors. C5 instances deliver 2x the floating point operations per second (FLOPS) per core per cycle compared to the previous-generation C4 instances. C5n instances offer significantly higher network bandwidth and 33 percent higher memory footprint compared to C5 instances.

Use cases

C5 and C5n instances are ideal for the most compute-intensive workloads, including HPC, batch processing, ad serving, video encoding, gaming, scientific modelling, distributed analytics, and CPU-based ML inference. C5n instances are ideal for applications that can take advantage of improved network throughput and packet-rate performance. C5 and C5n instances deliver cost-effective high performance at a low price-per-compute ratio.

Memory-optimized computing

Amazon EC2 R5, R5n, and R5dn instances

R5 instances deliver 5 percent additional memory per vCPU compared to R4 instances, and the largest size provides 768 GiB of memory. In addition, R5 instances deliver a 10 percent price-per-GiB improvement and about a 20 percent increase in CPU performance compared to R4 instances. R5 instances are ideal for memory-bound workloads including high-performance databases, distributed web-scale in-memory caches, mid-sized in-memory databases, real-time big data analytics, and other enterprise applications. The higher-bandwidth R5n and R5dn instance variants are ideal for applications that can take advantage of improved network throughput and packet-rate performance.

Use cases

R5 instances are well suited for memory-intensive applications such as high-performance databases, distributed web-scale in-memory caches, mid-size in-memory databases, real-time big data analytics, and other enterprise applications.

ENGIE seeks to digitally transform and streamline financial processes using SAP S/4HANA on AWS

"Deciding to migrate our SAP footprint to AWS while developing our new business framework and analytics system proved to be an excellent choice for us," says Thierry Langer [chief information officer of the finance division at ENGIE]. "We were able to reduce the footprint of our database substantially: on premises, our SAP database was 4 TB on the primary database and 4 TB on the secondary database. By moving to AWS, we right-sized our database. We reduced the size of the primary database from 4 TB to 2 TB and reduced the secondary database from 4 TB to around 200 GB." ENGIE has not only saved costs by reducing the size of its SAP HANA databases, the company also achieved cost savings by avoiding the need to procure on-premises hardware on an ongoing basis.

Amazon EC2 X1e, High Memory, and z1d instances

X1e instances offer the highest memory per vCPU among Amazon EC2 instance types and one of the lowest prices-per-GiB of memory. X1e instances are powered by four Intel Xeon processor E7-8880 v3 processors that feature high memory bandwidth and large L3 caches to boost the performance of in-memory applications.

High Memory instances offer the most memory of any Amazon EC2 instance, with up to 24 TB of memory paired with up to 8 sockets, delivering 448 vCPUs. These instances are purpose-built to run large in-memory databases, including production deployments of SAP HANA, in the cloud.

Amazon EC2 z1d instances offer both high compute capacity and a high memory footprint. High-frequency z1d instances deliver a sustained all-core frequency of up to 4.0 GHz, the fastest of any cloud instance and have been custom designed by AWS and Intel for HPC workloads.

Storage-optimized computing

Amazon EC2 I3 and I3en instances

This instance family provides NVMe Storage-on-Demand (SoD)-backed instances optimized for low latency, high random input/output (I/O) performance, high sequential read throughput, and high I/O operations per second (IOPS) at a low cost. I3 also offers bare-metal instances (i3.metal), powered by the AWS Nitro System for non-virtualized workloads, workloads that benefit from access to physical resources, or workloads that might have license restrictions. I3 instance types make use of a 2.3 GHz clock speed Intel Xeon processor E5-2686 v4. The I3en instance benefits from Intel Xeon Scalable processors with the Intel AVX-512 instruction, up to 3.1 GHz clock speed, and with 4x the raw storage capacity at roughly half the cost-per-GB of storage per host of current I3 offerings. In addition, it comes with low-latency NVMe SSD capacity for applications that require high random I/O access to large amounts of data such as relational databases. I3en hosts have up to 96 logical cores from 2nd Generation Intel Xeon Scalable processors, 768 GiB RAM, and approximately 45.84 TiB raw storage capacity per host. I3 and I3en support bare-metal instance sizes for workloads that benefit from direct access to physical process or memory.

Use cases

I3 instances are well suited for NoSQL databases, in-memory databases, scale-out transactional databases, data warehousing, Elasticsearch, and analytics workloads. I3en is designed for storage-dense workloads with high-performance requirements and delivers superior economics at scale for data center migration and disaster-recovery transformation projects.

AWS INSTANCE TYPE	COMPUTE OPTIMIZED C5 (12.XLARGE, 24.XLARGE, BARE METAL)	GENERAL PURPOSE M5N, M5DN	MEMORY OPTIMIZED R5N, R5DN	HIGH MEMORY (18 TIB, 24 TIB)
TARGET WORKLOAD	HPC	HPC, big data, AI, ML, Internet of Things (IoT)	SAP, VMware, enterprise resource planning (ERP), in-memory database (IMDB)	SAP, IMDB
USE CASE	Compute, performance	Mainstream	Virtualization, databases	Enterprise-scale IMDB
PROCESSOR	Intel Xeon Platinum 8275CL processor *custom*	Intel Xeon Platinum 8259CL processor *custom*	Intel Xeon Platinum 8259CL processor *custom*	Intel Xeon Platinum 8280L processor
MAX ALL-CORE TURBO CLOCK SPEED	3.6 GHz	3.1 GHz	3.1 GHz	4.0 GHz
BASELINE CLOCK SPEED	3.0 GHz	2.5 GHz	2.5 GHz	2.1 GHz
INTEL AVX	Intel AVX-512 (Intel DL Boost capable)	Intel AVX-512 (Intel DL Boost capable)	Intel AVX-512 (Intel DL Boost capable)	Intel AVX-512 (Intel DL Boost capable)
INTEL AES-NI	Yes	Yes	Yes	Yes
INTEL TURBO BOOST TECHNOLOGY	Yes	Yes	Yes	Yes
MEMORY SPEED (MEGATRANSFERS PER SECOND [MT/S])	2,933	2,666	2,666	2,666

AWS INSTANCE TYPE	COMPUTE OPTIMIZED **C5**	GENERAL PURPOSE **M5**	MEMORY OPTIMIZED **R5**	BURSTABLE **T3**	I/O OPTIMIZED **I3EN**	HIGH MEMORY (UP TO 24 TIB)	MEMORY OPTIMIZED **Z1D**	MEMORY OPTIMIZED **X1E**
TARGET WORKLOAD	HPC, big data, AI, ML, IoT	HPC, big data, AI, ML, IoT	SAP, VMware, ERP, IMDB	Burstable general-purpose	VMware, databases	SAP, IMDB	HPC, electronic design automation (EDA), gaming, databases	SAP, big data, HPC
USE CASE	Compute, performance	Mainstream	Virtualization, databases	Burstable, mainstream	SSD storage-optimized performance	Enterprise-scale IMDB	Single-threaded frequency performance	High-memory, performance database
PROCESSOR	Intel Xeon Platinum 8124M processor *custom*	Intel Xeon Platinum 8175M processor *custom*	Intel Xeon Platinum 8175M processor *custom*	Intel Xeon Scalable processor *custom*	Intel Xeon Scalable processor *custom*	Intel Xeon Platinum 8176M processor	Intel Xeon Platinum 8151 processor *custom*	Intel Xeon E7-8880 processor
MAX ALL-CORE TURBO CLOCK SPEED	3.4 GHz	3.1 GHz	3.1 GHz	-	3.1 GHz	3.8 GHz	4.0 GHz	2.7 GHz
BASELINE CLOCK SPEED	3.0 GHz	2.5 GHz	2.5 GHz	-	2.5 GHz	2.1 GHz	3.4 GHz	2.3 GHz
INTEL AVX	Intel AVX-512	Intel AVX-512	Intel AVX-512	Intel AVX-512	Intel AVX-512	Intel AVX-512	Intel AVX-512	Intel AVX2
INTEL AES-NI	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
INTEL TURBO BOOST TECHNOLOGY	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
MEMORY SPEED (MEGATRANSFERS PER SECOND [MT/S])	2,666	2,666	2,666	2,666	2,666	2,666	2,666	1,866

Learn More

"Today's top clouds are powered by Intel" white paper: **intel.com/content/www/us/en/cloud-computing/top-clouds-powered-by-intel-paper.html**

Intel cloud computing: **intel.com/cloud**

Intel and Amazon Web Services solutions: **intel.com/aws**

AWS with Intel overview: **https://aws.amazon.com/intel**

[1] FSI kernels baseline: **Intel Xeon Platinum 8268 processor configuration**: Intel "Wolf Pass" platform with 2-socket Intel Xeon Platinum 8268 processors (2.9 GHz, 24 cores), 12 x 16 GB DDR4-2,933, 1 SSD, BIOS: SE5C620.86B.02.01.0008.031920191559; microcode: 0x500001c, Red Hat Enterprise Linux (RHEL) 7.7, kernel 3.10.0-1062.1.1. FSI kernels v2.0: geomean (three workloads: Binomial Options, Black Scholes, Monte Carlo), AVX2_256 build, Intel Compiler 2019u5, Intel Math Kernel Library (Intel MKL) 2019u5, BIOS: Binomial (Intel Hyper-Threading Technology [Intel HT Technology] on, Intel Turbo Boost Technology on, SNC off, 2 threads/core), Black Scholes (Intel HT Technology off, Intel Turbo Boost Technology on, SNC off, 1 threads/core), Monte Carlo (Intel HT Technology on, Intel Turbo Boost Technology on, SNC off, 2 threads/core). Tested by Intel as of November 1, 2019. **With Intel AVX-512: Intel Xeon Platinum 8268 processor configuration**: Intel "Wolf Pass" platform with 2-socket Intel Xeon Platinum 8268 processors (2.9 GHz, 24 cores), 12 x 16 GB DDR4-2,933, 1 SSD, BIOS: SE5C620.86B.02.01.0008.031920191559; microcode: 0x500001c, RHEL 7.7, kernel 3.10.0-1062.1.1. FSI kernels v2.0: geomean (three workloads: Binomial Options, Black Scholes, Monte Carlo), Intel AVX-512 build, Intel Compiler 2019u5, Intel Math Kernel Library (Intel MKL) 2019u5, BIOS: BIOS: Binomial (Intel HT Technology on, Intel Turbo Boost Technology on, SNC off, 2 threads/core), Black Scholes (Intel HT Technology off, Intel Turbo Boost Technology on, SNC off, 1 threads/core), Monte Carlo (Intel HT Technology on, Intel Turbo Boost Technology on, SNC off, 2 threads/core). Tested by Intel as of November 1, 2019.

[2] Up to 14x AI performance improvement with Intel DL Boost compared to Intel Xeon Platinum 8180 processor (July 2017). Tested by Intel as of February 20, 2019. 2-socket Intel Xeon Platinum 8280 processor, 28 cores, Intel HT Technology on, Intel Turbo Boost Technology on, total memory 384 GB (12 slots/32 GB/2,933 MHz), BIOS: SE5C620.86B.0D.01.0271.120720180605 (ucode: 0x200004d), Ubuntu 18.04.1 LTS, kernel4.15.0-45-generic, 1 x 745.2 GB sda Intel SSDSC2BA80 SSD, 3.7 TB nvme1n1 Intel SSDPE2KX040T7 SSD, deep learning framework: Intel Optimization for Caffe version: 1.1.3 (commit hash: 7010334f159da247db3fe3a9d96a3116ca06b09a), ICC version 18.0.1, Intel Math Kernel Library for Deep Neural Networks (Intel MKL-DNN) version: v0.17 (commit hash: 830a10059a018cd2634d94195140cf2d8790a75a, model: https://github.com/intel/caffe/blob/master/models/intel_optimized_models/int8/resnet50_int8_full_conv.prototxt, BS=64, DummyData, 4 instance/2 socket, datatype: INT8. Compared to system tested by Intel as of July 11, 2017: 2-socket Intel Xeon Platinum 8180 processor at 2.50 GHz (28 cores), Intel HT Technology disabled, Intel Turbo Boost Technology disabled, scaling governor set to "performance" via intel_pstate driver, 384 GB DDR4-2,666 ECC RAM. CentOS Linux release 7.3.1611 (Core), Linux kernel 3.10.0-514.10.2.el7.x86_64. SSD: Intel SSD DC S3700 (800GB, 2.5-in Serial ATA [SATA] 6 gigabits per second [Gb/s], 25-nm, MLC). Performance measured with environment variables: KMP_AFFINITY='granularity=fine, compact', OMP_NUM_THREADS=56, CPU frequency set with cpupower frequency-set -d 2.5G -u 3.8G -g performance. Caffe: (http://github.com/intel/caffe/), revision f96b759f71b2281835f690af267158b82b150b5c. Inference measured with "caffe time --forward_ only" command, training measured with "caffe time" command. For "ConvNet" topologies, dummy dataset was used. For other topologies, data was stored on local storage and cached in memory before training. Topology specs from https://github.com/intel/caffe/tree/master/models/intel_optimized_models (ResNet-50), Intel C++ compiler ver. 17.0.2 20170213, Intel MKL small libraries version 2018.0.20170425. Caffe run with "numactl –l."

[3] Intel and TSO Logic. "New Advances by Intel and Amazon Web Services, Drive Major Cloud Savings." intel.com/content/dam/www/public/us/en/documents/reports/tso-logic-research.pdf.

[4] Wccftech. "First Look: Intel vs AMD EPYC AWS Cloud (IaaS) Benchmarks." January 2019. https://wccftech.com/first-look-intel-vs-amd-epyc-aws-cloud-iaas-benchmarks/.

[5] Intel DL Boost is only available with 2nd Generation Intel Xeon Scalable processors.

[6] SAP. "Find Certified IaaS Platforms." sap.com/dmc/exp/2014-09-02-hana-hardware/enEN/iaas.html#categories=Amazon%20Web%20Services.

[7] AWS. "New SAP Certifications for AWS Instances and World Record Benchmark Results." November 2017. https://aws.amazon.com/blogs/awsforsap/new-sap-certifications-for-aws-instances-and-world-record-benchmark-results/.

[8] Note that this is not a full list of Intel-based instances provided. For a comprehensive list of Intel-based instances, visit the AWS Management Console: https://console.aws.amazon.com/ec2/.

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

Intel technologies may require enabled hardware, software or service activation.

Printed in USA 1020/SM/PRW/PDF Please Recycle 344065-001US