Brief

Intel® Data-Centric Technologies
Cloud Services

Advance Your Business with Confidence

Drive business value and innovation with support from Microsoft Azure and Intel.

Microsoft Azure

Streamline and scale your technology investments, unlock new opportunities with hybrid cloud and artificial intelligence (AI), and innovate with confidence with industry-leading security. As you look toward shifting apps to the cloud, you'll find that you can do so seamlessly and consistently on the Microsoft Azure platform and a purpose-built Intel infrastructure. As a result, you can deploy any app anywhere and operate diverse workloads like a single environment. You can uncover new insights and business models with powerful analytics and AI tools, whether in a hybrid or Internet of Things (IoT) environment.

Intel and Microsoft share a common goal for delivering state-of-the-art business solutions to customers around the world. Together, they continue to co-engineer around a shared vision to build consistent cloud solutions and services for all. Intel® Xeon® processors are the foundation of many Azure services and infrastructure-as-a-service (IaaS) offerings, from the Azure cloud extending into edge platforms.

Microsoft Azure Virtual Machines powered by Intel

Azure virtual machines (VMs) powered by Intel Xeon processors have the largest breadth, global reach, and availability of compute instances across 58 regions and 124 countries tailored to meet your diverse computing needs. From deploying basic web applications to running SAP HANA in the cloud, you can trust that the underlying Intel foundation is optimized to provide strong data protection, fast processing of large data volumes, and service flexibility without a hit to performance. Migrating your applications to Intel Xeon processor–powered Azure VMs can improve your performance and productivity compared to using legacy on-premises infrastructure.

Intel recommends always selecting the latest generation of Azure VMs featuring Intel Xeon processors for high performance, scalability, and security capabilities.

Appendix A shows the Azure VM families powered by the latest generation of Intel Xeon processors.

Intel and Microsoft partner on specialized workloads for Azure with dedicated co-engineering and deep partner collaboration to scale and deliver solutions to help ensure the success of mutual customers.

Azure confidential computing is an emerging paradigm in today's data-centric era that helps maintain privacy and confidentiality regardless of where data is processed. Intel is working to make it possible for customers to maintain control of their data, even when it's in the cloud. Azure DCsv2-series VMs feature Intel SGX, which offers hardware-based memory encryption to isolate specific application code and data in memory, allowing user-level code to allocate private regions of memory, called enclaves, which are designed to be protected from processes

running at higher privilege levels. Confidential computing expands opportunities for sharing your data confidently and forging trusted relationships with third parties to solve the biggest problems or create innovation breakthroughs.

SAP on Azure offers a high-performance, scalable infrastructure for SAP HANA, with VMs offering up to 12 TB of memory per VM, in addition to purpose-built, bare-metal instances with up to 24 TB of memory, referred to as SAP HANA on Azure (Large Instances). SAP on Azure infrastructure, with VMs or Large Instances, is powered by purpose-built Intel Xeon processors, and Azure Large Instances feature optional Intel® Optane™ persistent memory (PMem). Intel Optane PMem, combined with 2nd Generation Intel Xeon Scalable processors, brings the persistence of solid state drives (SSDs), with access times at DRAM-like speeds, to deliver high performance, low total cost of ownership (TCO), and consistent uptime.

Azure VMware Solution allows customers to bring their existing VMware applications seamlessly into Azure, and it enables them to continue to use the familiar VMware tools while benefiting from the cloud. Azure VMware Solution runs on dedicated hardware, providing customers fully managed, single-tenant, bare-metal Azure infrastructure powered by Intel Xeon processors and Intel SSDs.

High-performance computing (HPC) on Azure is available for customers looking for short-term bursts or full migration of big compute workloads into Azure. Azure HC-series VMs, powered by Intel Xeon Scalable processors and Intel AVX–tuned applications, enable HPC that's optimized for large-scale, computationally intensive applications.

Deep learning on Azure: Intel collaborated with Azure to build the Intel Optimized Data Science Virtual Machine to take advantage of the Intel AVX-512 instruction set for faster vector operations. This Intel-optimized VM provides an average of 7.7x speedup over non-optimized environments.[1]

Intel Xeon 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 and data application speeds.[2]
- **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.[3]
- **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.

A true edge-to-cloud platform

Microsoft Azure and Intel are committed to offering customers the ability to run their workloads wherever makes the most sense for their businesses. Customers can extend Azure services and capabilities on premises and build, deploy, and operate hybrid cloud and edge applications consistently with the Azure Stack family and Azure IoT.

> "Intel and Microsoft are co-engineering solutions to empower enterprise customers to realize significant business value from their hybrid cloud strategies. Specifically, we're optimizing the Azure Stack family with the latest Intel Xeon processors, Intel Optane technologies for storage and memory, SSDs, and FPGAs to give customers the flexibility to embrace the cloud on their terms, wherever they deploy them, with security, scalability, and predictable performance."
>
> — William Giard, Intel CTO, Cloud and Enterprise Solutions Group

Azure Stack Edge is an AI-enabled edge computing device with network data-transfer capabilities allowing customers to extend their intelligent, pre-trained models into the data center. Powered by Intel Xeon Scalable processors and fast Intel SSDs, customers have the choice of Intel® Arria® 10 FPGAs or NVIDIA graphics processing units (GPUs) for real-time, local inferencing. You don't need to buy Azure Stack Edge—you can order the appliance through the Azure portal and pay for it as you go, just like any other Azure services. Then you can manage your appliance through the Azure portal and make use of common management and identity policies across the cloud and the edge.

Azure Stack HCI provides a simplified, low-cost, hyper-converged infrastructure (HCI) tailored to the different compute, memory, and storage needs in the data center. With Azure Stack HCI, you can improve VM performance with millions of input/output operations per second (IOPS) and consistent sub-millisecond latency.
Azure Stack HCI can be purchased from major OEMs, and it runs on the latest Intel Xeon processors, Intel SSDs, and Intel networking options, tailored to each customer's needs.

Azure Stack Hub is an integrated system of software and validated Intel hardware. Azure Stack Hub is best suited for customers needing to run their own autonomous clouds, completely or partially disconnected from the Internet and public cloud.

Appendix A: Intel Xeon processors in Azure instances

AZURE IAAS	TARGETED WORKLOADS OR APPLICATIONS	INTEL XEON PROCESSOR	VCPU RANGE	MEMORY (GIB)	INTEL TURBO BOOST TECHNOLOGY 2.0	INTEL HYPER-THREADING TECHNOLOGY (INTEL HT TECHNOLOGY)	INTEL SOFTWARE GUARD EXTENSIONS (INTEL SGX)	INTEL ADVANCED VECTOR EXTENSIONS 512 (INTEL AVX-512)	INTEL DEEP LEARNING BOOST (INTEL DL BOOST)	INTEL OPTANE PMEM
Ddv4 (with disk)	Enterprise-grade apps Relational databases, web app servers	2nd Generation Intel Xeon Platinum 8272CL processor	2–64	8-256	√	√		√	√	
DCsv2-series	Confidential computing	Intel Xeon processor E-2288G	1–8	4–32	√	√	√			
Fsv2-series	Data science, machine learning (ML)/AI inference, gaming	Intel Xeon Platinum 8168 processor	2–72	4–144	√	√		√		
HC-series	HPC	Intel Xeon Platinum 8168 processor	44	352	Off	√		√		
Edv4-series	In-memory apps, in-memory analytics, relational database servers	2nd Generation Intel Xeon Platinum 8272CL processor	2–64	16–504	√	√		√	√	
Mv2-series	SAP HANA, SQL Hekaton, in-memory critical	Intel Xeon Platinum 8160M processor	208–416	11,400	√	√		√		
Bare metal: **SAP HANA on Azure S224**	SAP HANA	2nd Generation Intel Xeon Platinum 8276 processor	112–448	4.5–36 TB	√	√		√	√	Optional

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 and Microsoft partnership: **intel.com/microsoftazure**

Intel and Azure: **azure.com/intel**

[1] Performance results are based on Intel testing as of January 15, 2019. Configuration: Azure instance size: F72s_v2. Architecture: x86_64. CPU op-mode(s): 32-bit, 64-bit. Byte order: Little Endian. CPU(s): 72. On-line CPU(s) list: 0–71. Thread(s) per core: 2. Core(s) per socket: 18. Socket(s): 2. NUMA node(s): 2. Vendor ID: GenuineIntel. CPU family: 6. Model: 85. Model name: Intel Xeon Platinum 8168 processor at 2.70 GHz. Stepping: 4. CPU MHz: 2693.855. BogoMIPS: 5,387.73. Virtualization: Intel Virtualization Technology (Intel VT-x). Hypervisor vendor: Microsoft. Virtualization type: full. L1d cache: 32K. L1i cache: 32K. L2 cache: 1,024K. L3 cache: 33,792K. NUMA node0 CPU(s): 0–35. NUMA node1 CPU(s): 36–71.

[2] Configuration details updated on March 30, 2020, to reflect the latest SPEC guidelines. 36 percent more estimated performance and 42 percent more estimated performance/dollar: geomean of SPECrate2017_int_base(est), SPECrate2017_fp_base(est), STREAM triad, and Intel Distribution for LINPACK across 10 new 2-socket 2nd Generation Intel Xeon Gold processors versus the previous generation. 2nd Generation Intel Xeon Gold R processors: 1-node, 2 x 2nd Generation Intel Xeon Gold processor (62xxR/$$) on Intel reference platform with 384 GB (12 slots/32 GB/62xx at 2,933, 52xx at 2,666) total memory, ucode 0x500002c, Intel HT Technology on for all except off for STREAM (GB/s), LINPACK (GFLOPS/s), Intel Turbo Boost Technology on, with Ubuntu 19.10, 5.3.0-24-generic, Intel Xeon Gold 6258R processor/$3,950: SPECrate2017_int_base(est)=323, SPECrate2017_fp_base(est)=262, STREAM=224, LINPACK =3305; Intel Xeon Gold 6248R processor/$2,700: SPECrate2017_int_base(est)=299, SPECrate2017_fp_base(est)=248, STREAM=224, LINPACK =3010; Intel Xeon Gold 6246R processor/$3,286: SPECrate2017_int_base(est)=238, SPECrate2017_fp_base(est)=217, STREAM=225, LINPACK=2,394; Intel Xeon Gold 6242R processor/$2,529: SPECrate2017_int_base(est)=265, SPECrate2017_fp_base(est)=231, STREAM=227, LINPACK=2,698; Intel Xeon Gold 6240R processor/$2,200: SPECrate2017_int_base(est)=268, SPECrate2017_fp_base(est)=228, STREAM=223, LINPACK=2,438; Intel Xeon Gold 6238R processor/$2,612: SPECrate2017_int_base(est)=287, SPECrate2017_fp_base(est)=240, STREAM=222, LINPACK =2545; Intel Xeon Gold 6230R processor/$1,894: SPECrate2017_int_base(est)=266, SPECrate2017_fp_base(est)=227, STREAM=222, LINPACK =2219; Intel Xeon Gold 6226R processor/$1,300: SPECrate2017_int_base(est)=208, SPECrate2017_fp_base(est)=192, STREAM=200, LINPACK=2,073; Intel Xeon Gold 5220R processor/$1,555: SPECrate2017_int_base(est)=257, SPECrate2017_fp_base(est)=220, STREAM=210, LINPACK=1,610; Intel Xeon Gold 5218R processor/$1,273: SPECrate2017_int_base(est)=210, SPECrate2017_fp_base(est)=188, STREAM=199, LINPACK=1,290, test by Intel on 12/25/2019. Previous-generation Intel Xeon Gold processor: 1-node, 2 x Intel Xeon Gold processor (61xx/$$) on Intel reference platform with 384 GB (12 slots/32 GB/61xx at 2,666, 51xx at 2,400) total memory, ucode 0x500002c, Intel HT Technology on for all except off for STREAM (GB/s), LINPACK (GFLOPS/s), Intel Turbo Boost Technology on, with Ubuntu 19.10, 5.3.0-24-generic, Intel Xeon Gold 6152 processor/$3,655: SPECrate2017_int_base(est)=224, SPECrate2017_fp_base(est)=198, STREAM=200, LINPACK=1,988; Intel Xeon Gold 6148 processor/$3,072: SPECrate2017_int_base(est)=225, SPECrate2017_fp_base(est)=198, STREAM=197, LINPACK=2,162; Intel Xeon Gold 6146 processor/$3,286: SPECrate2017_int_base(est)=161, SPECrate2017_fp_base(est)=175, STREAM=185, LINPACK=1,896; Intel Xeon Gold 6142 processor/$2,946: SPECrate2017_int_base(est)=193, SPECrate2017_fp_base(est)=176, STREAM=185, LINPACK=1,895; Intel Xeon Gold 6140 processor/$2,445: SPECrate2017_int_base(est)=202, SPECrate2017_fp_base(est)=183, STREAM=188, LINPACK=1,877; Intel Xeon Gold 6138 processor/$2,612: SPECrate2017_int_base(est)=189, SPECrate2017_fp_base(est)=195, STREAM=189, LINPACK=1,976; Intel Xeon Gold 6130 processor/$1,894: SPECrate2017_int_base(est)=172, SPECrate2017_fp_base(est)=165, STREAM=185, LINPACK=1,645; Intel Xeon Gold 6126 processor (proj)/$1,776: SPECrate2017_int_base(est)=141, SPECrate2017_fp_base(est)=157, STREAM=170, LINPACK=1,605; Intel Xeon Gold 5120 processor (proj)/$1,555: SPECrate2017_int_base(est)=148, SPECrate2017_fp_base(est)=148, STREAM=159, LINPACK=924, Intel Xeon Gold 5118 processor/$1,273: SPECrate2017_int_base(est)=134, SPECrate2017_fp_base(est)=132, STREAM=149, LINPACK=818, test by Intel on 2/18/2020. Your costs and results may vary.

[3] 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, 384 GB total memory (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, SSD 1 x sda INTEL SSDSC2BA80 745.2 GB, nvme1n1 INTEL SSDPE2KX040T7 3.7 TB, deep learning framework: Intel Optimization for Caffe version: 1.1.3 (commit hash: 7010334f159da247db3fe3a9d96a3116ca06b09a), ICC version 18.0.1, 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 versus 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 DDR4-2666 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 (800 GB, 2.5-in SATA 6 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".

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 344063-001US