

# intel

### Łukasz Chrzanowski

Field Application Engineer, Sales & Marketing Group

3rd Gen Intel Xeon Scalable Processors

### Notices and Disclaimers

Performance varies by use, configuration and other factors. Learn more at <a href="www.lntel.com/PerformanceIndex">www.lntel.com/PerformanceIndex</a>.

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.

Intel contributes to the development of benchmarks by participating in, sponsoring, and/or contributing technical support to various benchmarking groups, including the BenchmarkXPRT Development Community administered by Principled Technologies.

Your costs and results may vary.

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

Some results may have been estimated or simulated.

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

All product plans and roadmaps are subject to change without notice.

Statements in this document that refer to future plans or expectations are forward-looking statements. These statements are based on current expectations and involve many risks and uncertainties that could cause actual results to differ materially from those expressed or implied in such statements. For more information on the factors that could cause actual results to differ materially, see our most recent earnings release and SEC filings at www.intc.com.

© Intel Corporation. Intel, the Intel logo, and other Intel marks are trademarks of Intel Corporation or its subsidiaries. Other names and brands may be claimed as the property of others.

# Business Needs and Demands are Dynamically Shifting













### Future of the Data Center

Compute, storage, and memory will be increasingly disaggregated.

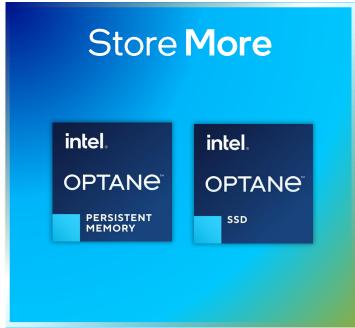
Security architected-in at the chip level

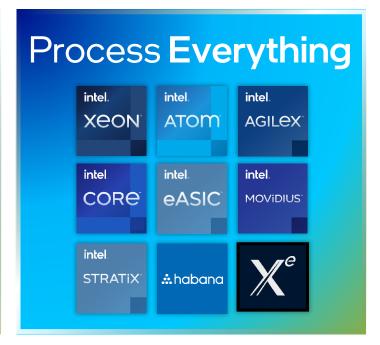
Software deployed as smaller units called Microservices

CPUs and XPUs will work together to solve complex challenges

### Unmatched Portfolio of Hardware, Software and Solutions



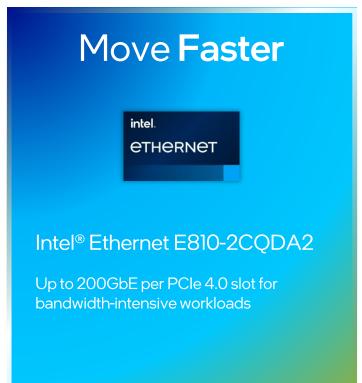




Optimized Software and System-Level Solutions

# Announcing Today – Intel's Latest Data Center Portfolio

Targeting 3rd Gen Intel® Xeon® Scalable processors







**Optimized Solutions** 





>500 Partner Solutions

# Why Customers Choose Intel

### Delivering workload-optimized performance

#### **Platform Features**

>2X more TLS requests processed with new Intel Crypto acceleration for fast, predictable response and lower TCO



#### Comprehensive Portfolio

Realized 2:1 server consolidation while quickly accessing and reformatting images and video using Intel® Xeon® processors, Intel® Optane™ and Intel FPGAs

### Al Capabilities

1.8X performance improvement for Monte Carlo simulation with no accuracy loss using Intel DL Boost and oneAPI-optimized software



# Higher Performance Through Optimized Software

3.78X faster inference with lower power and higher utilization after switching from GPU text recognition to Intel® Xeon® processors with optimized PyTorch

### Partnership and Co-Engineering

Faster TTM and 20X faster boot time using Intel's early access programs and engineering support to create SMB continuity and recovery solution

datto

#### **Solutions**

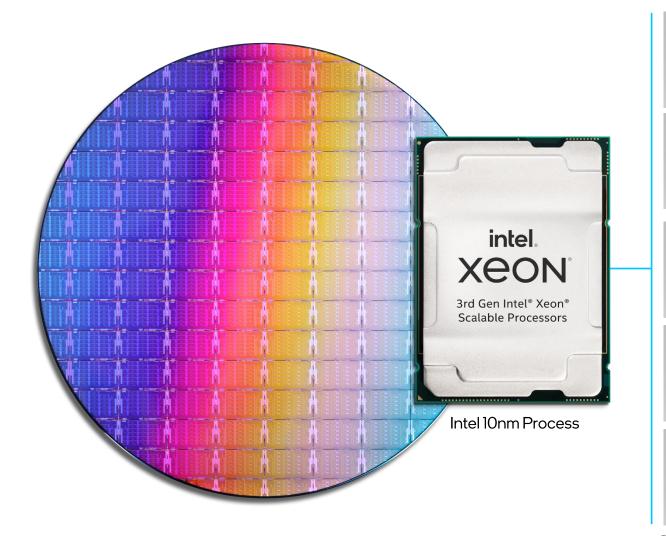
Quickly scaled from 8-node POC to 68-node full production COVID-19 research cluster using Intel Select Solution for Genomics Analytics



### Platform to deliver true flexibility

Performance varies by use, configuration and other factors. Configurations see appendix [45]

### Performance made flexible



### Up to 40 cores

per processor

### 20% IPC improvement

28 core, ISO Freq, ISO compiler

### 1.46x average performance increase

Geomean of Integer, Floating Point, Stream Triad, LINPACK 8380 vs. 8280

#### 1.74x Al inference increase

8380 vs. 8280 BERT

# 2.65x average performance increase vs. 5-year-old system 8380 vs. E5-2699v4

Performance varies by use, configuration and other factors. Configurations see appendix [1,3,5,55]

### Performance made flexible

# Only x86 data center processor with built-in AI & security solutions





**Guard Extensions** 





Intel Crypto Acceleration



Intel Total Memory Encryption



Intel Platform Firmware Resilience



Targeted for 1S-2S systems

#### Scalable, flexible, customizable



Intel Deep Learning Boost



Intel Speed Select Technology



Intel AVX-512



Optimized Software

#### Next-gen Xeon Scalable Platform



System Memory
Capacity
(Per Socket)
DRAM + PMem

8CH

DDR4-3200 2 DPC (Per Socket) 2.6X

Memory Capacity Increase vs. 2nd Gen Xeon 64

Lanes
PCI Express 4
(per Socket)

#### Breakthrough Data Performance



Intel®
Optane™
persistent
memory 200



Intel® Optane™ SSD P5800 series



Intel® SSD D series

#### Faster, Flexible, Data Scale



Intel® Ethernet 800 series



Intel® Agilex FPGA solutions





# Flexible Performance for Most Demanding Workloads

Outstanding gen-on-gen performance from intelligent edge to cloud











Cloud

1.5x

Improvement in Latency Sensitive Workloads

5G

1.62x

Improvement in Network and Communications Workloads

IoT

1.56x

Image Classification Inference Improvement **HPC** 

1.57x

Faster Modeling for Critical Vaccine Research Artificial Intelligence

**UPTO** 

1.74x

Language Processing Inference Improvements

Performance varies by use, configuration and other factors. Configurations see appendix [5,7,17,19,52]

### Broad industry adoption

#### **Cloud Service Providers**

- All top CSPs planning to deploy services
- Over 200 ISVs and partners have deployed Intel SGX





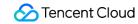












### **OEM/ODMs**

- >250 designs with more than 50 unique OxM partners
- Broad software and ecosystem readiness to speed faster time to value



Hewlett Packard Enterprise









#### **Network Providers**

 >15 major TEMs, OEMs and CoSPs readying POCs or initial deployments









#### **HPC Labs and HPCaaS**

- Over 20 publicly-declared HPC adopters to date
- HPC solutions coming from all major OEMs
- Growing HPCaaS footprint including:









>200k units shipped in Q12021

### Only Data Center CPU with Built-in Al Acceleration

Artificial intelligence made flexible

### Vs. Competition

- Up to 25x performance on image recognition vs.
   AMD EPYC 7763
- Xeon leads on average across a broad mix of 20 popular AI & ML workloads:
  - **Up to 1.5x vs.** AMD EPYC 7763
  - Up to 1.3x vs. Nvidia A100 GPU

### Intel Software Optimizations

- Up to 10x performance improvement with TensorFlow for deep learning on ResNet50 vs. default distro
- Up to 100x improvement with Scikit-Learn for machine learning on SVC/kNN predict vs. the default distro



#### Performance

Medical imaging solution exceeds performance requirements using Intel DL Boost powered by OneAPI



### Productivity

Drive-thru recommendation engine using Analytics Zoo to unify end-to-end Spark data pipeline on a Xeon-based cluster



#### Simplicity

Automating underground pipe inspection using a solution developed with Intel Al Builders partner Wipro

150+ Containers and 200+ Turnkey Solutions Accelerate Development and Deployment

Performance varies by use, configuration and other factors. Configurations see appendix [26, 28, 29, 36, 54]

### Intel's Most Secure, Compliant and Performant Data Center Platform Security made flexible

#### Built-in Intel Software Guard Extensions

- Hundreds of customers currently using Intel SGX to enhance security and enable business transformation through data privacy
- Smallest attack surface of any data center confidential computing technology
- 4000x\* increase in protected enclave size to 1TB

### Introducing New Security Features

- Intel Total Memory Encryption for basic bulk encryption of the entire memory space to protect against physical attacks
- Intel Platform Firmware Resilience for defending and recovering the underlying firmware layer to protect against permanent denial of service



- Among the largest health insurance providers in Germany
- Intel SGX selected as Trusted Execution Environment for federated digital health record management serving over 26 million people



- Globally \$2T is laundered each year and can be used to fund crime or terrorism
- Intel SGX allows banks to share records without revealing private customer information and reduce false positives 83%

\*3rd Gen Intel® Xeon® Scalable processors provides up to 1TB enclave size in a 2S configuration (SKU dependent) vs Xeon E 2100 256MB AOK and Consilient systems initially deployed on Intel® Xeon® E processors

### Built-in Crypto Acceleration for Encryption-Intensive Workloads Encryption made flexible

### Intel Crypto Acceleration

- New instructions and architectural features parallelize execution of encryption functions
- Reduces penalty of implementing pervasive data encryption
- Higher throughput with fast and strong encryption for AVX-512 ISA
- Increases performance of encryption-intensive workloads such as SSL web server, 5G infrastructure and VPN/firewalls

4.2x

#### **Encrypted Web Server**

More TLS encrypted web server connections per second; more content delivered per server



1.94x

#### Vector Packet Processing

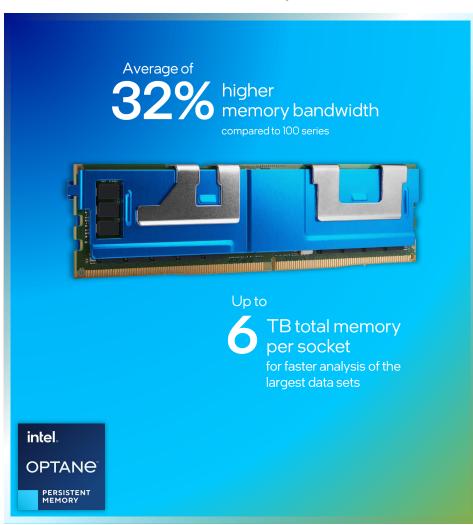
More encrypted packets processed per second; higher network and VPN capacity per node

ipsøc 🖯

Performance varies by use, configuration and other factors. Configurations see appendix [17, 51]

# Intel® Optane™ Persistent Memory 200 Series

### Persistent memory made flexible



### eADR (Enhanced Asynchronous DRAM Refresh)

improves performance of apps that use persistent memory by eliminating "cache flushes" – volatile data including the CPU caches save automatically, even if power fails

Intel Optane PMem 200 series is compatible with existing PMem SW ecosystem & it continues to grow



#### KATANA GRAPH

Computes up to **2X** faster graph analytics algorithms used in search, social networks, and fraud detection

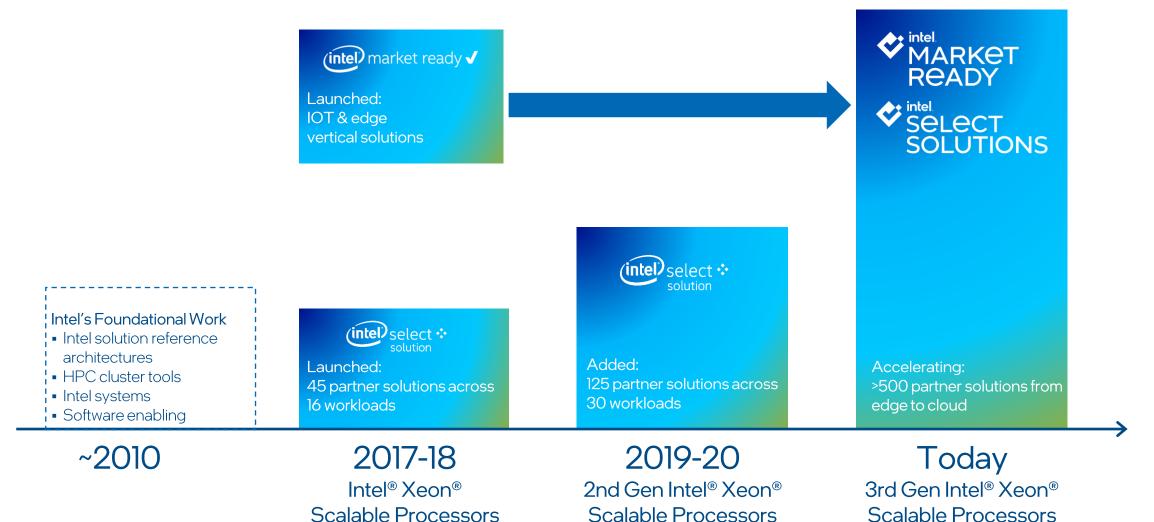
**vm**ware

Lower infrastructure costs by up **25%** per VM while delivering the same performance

Performance varies by use, configuration and other factors. Configurations see appendix [53]

# Decades of Solutions Delivery

### Accelerating customer deployments



### Current Intel® Select Solutions Portfolio



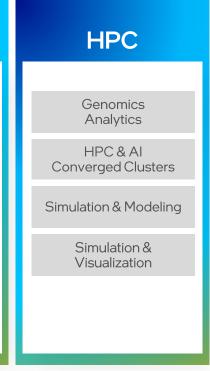


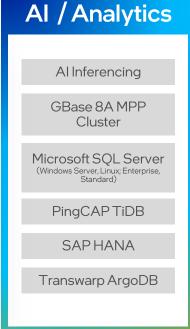


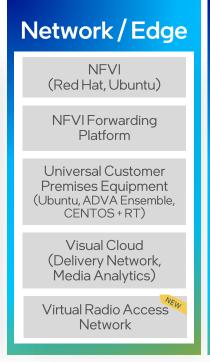




#### HCI/Storage **Cloud Services** Cisco Hyperflex EasyStack Microsoft Azure Stack HCI Nutanix HCI Google Cloud's Anthos Qing3 HCI Sangfor HCI & SDS Open Cloud SmartX HCI VMware Horizon VDI on vSAN VMware vSAN XSKY SDS













#### FOUR & EIGHT SOCKET SCALABLE PERFORMANCE

SKU	CORES	BASE (GHz)	SINGLE CORE TURBO (GHz)	ALL CORE TURBO (GHz)	CACHE (MB)	TDP (Watts)	Support for Intel Optane Persistent Memory 200 Series	Recommended Customer Pricing (RCP) in \$ US Dollars
8380HL	28	2.9	4.3	3.8	38.5	250	Yes	\$13012
8380H	28	2.9	4.3	3.8	38.5	250	Yes	\$10009
8376HL	28	2.6	4.3	3.5	38.5	205	Yes	\$11772
8376H	28	2.6	4.3	3.5	38.5	205	Yes	\$8719
8360HL	24	3.0	4.2	3.8	33	225	Yes	\$7203
8360H	24	3.0	4.2	3.8	33	225	Yes	\$4200
8356H	8	3.9	4.4	4.3	35.75	190	Yes	\$3400
8354H	18	3.1	4.3	4.0	24.75	205	Yes	\$3500
8353H	18	2.5	3.8	3.3	24.75	150	Yes	\$3003
6348H	24	2.3	4.2	3.1	33	165	Yes	\$2700
6330H	24	2.0	3.7	2.8	33	150	Yes	\$1894
6328HL	16	2.8	4.3	3.7	22	165	Yes	\$4779
6328H	16	2.8	4.3	3.7	22	165	Yes	\$1776
5320H	20	2.4	4.2	3.3	27.5	150	Yes	\$1555
5318H	18	2.5	3.8	3.3	24.75	150	Yes	\$1273

H and HL SKUs are only supported on a unique 4 or 8-socket platform. Please contact your hardware provider for a list of system availability supporting your specific SKU configuration.

H SKUs are configured to support up to 1.12 TB of system memory, per processor. HL SKUs are configured to support up to 4.5 TB of system memory, per processor. H and HL SKUs are validated for up to 256 GB capacity DRAM memory modules, as of March 2021.

H and HL SKUs support Intel Optane persistent memory 200 series in a 4-socket platform only. H SKUs are validated for up to 768 GB of Intel Optane persistent memory 200 series, per processor. HL SKUs are validated for up to 3 TB of Intel Optane persistent memory 200 series, per processor.

6330H, 6328H, 6328HL & 5320H processors include Intel Speed Select technology (Intel SST) supporting Intel SST Core Power (SST-CP) and Intel SST Turbo Frequency (SST-TF) capabilities.

#### 3rd Gen Intel Xeon Scalable Processors intel.com/xeonscalable







Intel may make changes to specifications and product descriptions at any time, without notice. Please visit intel.com/xeon or contact your Intel representative to obtain the latest product specifications.

Intel processor numbers are not a measure of performance. Processor numbers differentiate features within each processor family, not across different processor families. All processors support Intel Virtualization Technology (Intel VT-x). No product or component can be absolutely secure.

Intel, the Intel logo, Xeon, and Optane are trademarks of Intel Corporation or its subsidiaries. Prepared on 5 April 2021. © Copyright 2021. Intel Corporation.

#### OPTIMIZED FOR HIGHEST DED CODE COAL A DI E DEDECOMANICE

PER-C	DKE.	SCAL	ABLE	PERF		AINCE			
SKU	CORES	BASE (GHz)	SINGLE CORE TURBO (GHz)	ALL CORE TURBO (GHz)	CACHE (MB)	TDP (Watts)	Support for Intel Optane Persistent Memory 200 Series	Intel SGX Enclave Capacity (Per Processor)	Recommended Customer Pricing (RCP) in \$ US Dollars
8380	40	2.3	3.4	3.0	60	270	Yes	512 GB	\$8099
8368	38	2.4	3.4	3.2	57	270	Yes	512 GB	\$6302
8362	32	2.8	3.6	3.5	48	265	Yes	64 GB	\$5448
8360Y	36	2.4	3.5	3.1	54	250	Yes	64 GB	\$4702
8358	32	2.6	3.4	3.3	48	250	Yes	64 GB	\$3950
6348	28	2.6	3.5	3.4	42	235	Yes	64 GB	\$3072
6342	24	2.8	3.5	3.3	36	230	Yes	64 GB	\$2529
6354	18	3.0	3.6	3.6	39	205	Yes	64 GB	\$2445
6346	16	3.1	3.6	3.6	36	205	Yes	64 GB	\$2300
6334	8	3.6	3.7	3.6	18	165	Yes	64 GB	\$2214
6326	16	2.9	3.5	3.3	24	185	Yes	64 GB	\$1300
5317	12	3.0	3.6	3.4	18	150	Yes	64 GB	\$950
5315Y	8	3.2	3.6	3.5	12	140	Yes	64 GB	\$895

SCAL.	ABLE	PERF	ORM,	ANCE					
8352Y	32	2.2	3.4	2.8	48	205	Yes	64 GB	\$3450
6338	32	2.0	3.2	2.6	48	205	Yes	64 GB	\$2612
6336Y	24	2.4	3.6	3.0	36	185	Yes	64 GB	\$1977
6330	28	2.0	3.1	2.6	42	205	Yes	64 GB	\$1894
5320	26	2.2	3.4	2.8	39	185	Yes	64 GB	\$1555
5318Y	24	2.1	3.4	2.6	36	165	Yes	64 GB	\$1273
4316	20	2.3	3.4	2.8	30	150		8 GB	\$1002
4314	16	2.4	3.4	2.9	24	135	Yes	8 GB	\$694
4310	12	2.1	3.3	2.7	18	120		8 GB	\$501
4309Y	8	2.8	3.6	3.4	12	105		8 GB	\$501

Y Supports Intel Speed Select Technology – Performance Profile 2.0 (Intel SST-PP)

All 8300, 6300, 5300 and 4300 processors, Non-H/HL SKUs, are supported on a unique 1 or 2 socket platform. Please contact your hardware provider for a list of system availability supporting your specific SKU configuration.

All 8300, 6300, 5300 and 4300 processors, Non-H/HL SKUs, are configured to support up to 6 TB of system memory, per processor. Intel has validated for up to 4 TB of Intel Optane persistent memory 200 series, per processor, Intel has validated for up to 256 GB capacity DRAM memory modules, as of March 2021,

Unless noted, all 8300, 6300 and 5300 processors, Non-H/HL SKUs, include support for Intel Speed Select technology (Intel SST) featuring Intel SST Base Frequency (SST-BF), Intel SST Core Power (SST-CP) and Intel SST Turbo Frequency (SST-TF) capabilities.

M, P, Q, V SKUs and 8362 do not include support Intel Speed Select Technology Base Frequency (SST-BF).

### SKUS SUPPORTING MAXIMUM INTEL SGX ENCLAVE CAPACITY

SKU	CORES	BASE (GHz)	SINGLE CORE TURBO (GHz)	ALL CORE TURBO (GHz)	CACHE (MB)	TDP (Watts)	Support for Intel Optane Persistent Memory 200 Series	Intel SGX Enclave Capacity (Per Processor)	Recommended Customer Pricing (RCP) in \$ US Dollars
8380	40	2.3	3.4	3.0	60	270	Yes	512 GB	\$8099
8368Q	38	2.6	3.7	3.3	57	270	Yes	512 GB	\$6743
8368	38	2.4	3.4	3.2	57	270	Yes	512 GB	\$6302
8352S	32	2.2	3.4	2.8	48	205	Yes	512 GB	\$4046
5318S	24	2.1	3.4	2.6	36	165	Yes	512 GB	\$1667
8352S and	5318S sup	port Intel	Speed Se	lect Tech	nology – F	erformano	ce Profile 2.0 (	Intel SST-PP)	

8 GB

8 GB

64 GB

\$3950

\$3450

\$6743

\$3864

#### CLOUD OPTIMIZED FOR VM UTILIZATION

00001	02	2.0	0. 1	0.2	-10	2-10	103	
8352V	36	2.1	3.5	2.5	54	195	Yes	
P JaaS Clou	d Special	ized Proce	essor V S	aaS Cloud	d Speciali:	zed Proces	sor	

8352V supports Intel Speed Select Technology – Performance Profile 2.0 (Intel SST-PP)

#### LIOUID COOLED

512 GB 2.6 Yes

8368O supports up to 512 GB Intel Software Guard Extensions (Intel SGX) enclave capacity

NETW	ORKII	NG/NF	-V OP	I IMIZE	-D				
8351N	36	2.4	3.5	3.1	54	225	Yes	64 GB	\$3027
6338N	32	2.2	3.5	2.7	48	185	Yes	64 GB	\$2795
6330N	28	2.2	3.4	2.6	42	165	Yes	64 GB	\$2029
5318N	24	2.1	3.4	2.7	36	150	Yes	64 GB	\$1375

8351N is supported in a one-socket configuration only

5318N supports Intel Speed Select Technology – Performance Profile 2.0 (Intel SST-PP)

MEDIA	N PRO	CESS	ING OI	PTIMIZ	ZED			
8352M	32	2.3	3.5	2.8	48	185	Yes	

Optimized for processing Al and media workloads and services.

#### LONG HER LICE AND NEDGET HEDMAL EDIEN

LONG	-LIFE	USE P	NIND IN	EB2-1	HEKIY	IAL FRI	IEINDLY		
6338T	24	2.1	3.4	2.7	36	165	Yes	64 GB	\$2742
5320T	20	2.3	3.5	2.9	30	150	Yes	64 GB	\$1727
4310T	10	23	34	29	15	105		8 GB	\$555

Support for up to 10-year reliability, higher Tcase

#### CINICLE COCKET ODTIMIZED

SINGL	SINGLE-SOCKET OPTIMIZED									
8351N	36	2.4	3.5	3.1	54	225	Yes	64 GB	\$3027	
6314U	32	2.3	3.4	2.9	48	205	Yes	64 GB	\$2600	
6312U	24	2.4	3.6	3.1	36	185	Yes	64 GB	\$1450	

Supported in one-socket configurations only



### One and two-socket solutions

3rd Gen Intel Xeon Scalable Processors Product 30-3-30

#### Performance made flexible for 1 and 2-socket solutions

#### Flexible Performance

Up to 40 cores per processor, up to 80 cores in a 2S configuration Up to 60 MB of L3 Cache (1.5 MB/core), per processor Support for 85W\* to 270W TDP

> Up to 3.7 GHz single-core turbo frequency Up to 3.6 GHz all-core turbo frequency Up to 3.6 GHz highest base frequency

Intel Turbo Boost 2.0 Technology, Intel Hyper-Threading Technology (Intel HT), Intel Speed Shift Technology, Enhanced Intel SpeedStep Technology

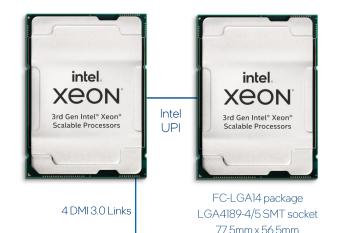
Up to 3 Intel Ultra Path Interconnect (Intel UPI) with up to 11.2 GT/s per processor

#### Next-gen Xeon Scalable Platform

8 channels, 2 DPC, DDR4, Up to 3200 MT/s Support for up to 6 TB of system memory per processor, up to 12 TB in a 2S configuration

Support for Intel Optane persistent memory 200 series Support for up to 4 TB of PMem per processor, up to 8 TB in a 2S configuration

64 lanes of PCI Express 4.0 per processor, 128 lanes in a 2S configuration



#### Built-In Acceleration and Advanced Security Solutions

Intel Deep Learning Boost (Intel DL Boost) with VNNI (int8) Intel Advanced Vector Extension 512 (Intel AVX-512) with 2 FMA Intel Direct Data I/O Technology (Intel DDIO)

Intel Software Guard Extensions (Intel SGX)

- up to 512 GB of secure enclave capacity per processor, up to 1 TB in a 2S configuration Intel Crypto Acceleration
- Vector AES-NI, Vector CLMUL, Intel Secure Hash Algorithm Extensions, VPMADD52 instructions, and RSA/DH encryption protocols

Intel Total Memory Encryption (Intel TME) with AES-XTS encryption
Intel Platform Firmware Resilience (Intel PFR)

#### Scalable, Flexible and Customizable

Intel Speed Select Technologies

- Performance Profile 2.0, on select SKUs
- Base Frequency, Turbo Frequency, and Core Power, available on most SKUs

Intel Infrastructure Management Technologies (Intel IMT)

- Intel Resource Director Technology (Intel RDT)
- Intel Virtualization Technology (Intel VT-x, VT-d)

PCI Express 3.0 (Up to 20 lanes)
USB 3.0 (Up to 10 ports)
USB 2.0 (Up to 14 ports)
SATA Gen 3 (Up to 14 ports)



Platform and Storage Manageability

Enhanced Firmware Security

4189 pins

Advanced Server Reliability, Serviceability and Availability (RAS) Feature Set Available with optional Intel QuickAssist technology (Intel QAT)

Processors, chipset and diagram provided for illustration purposes only and is not to scale

Two socket example shown

No product or component can be absolutely secure

Please visit intel.com/xeon to obtain the latest product specifications

Software information for one and two-socket solutions

Software Type / Supplier	Supported
Linux	Red Hat Enterprise Linux 7.9 and later 7.x branches (basic support); Red Hat Enterprise Linux 8.2 and later 8.x branches SUSE Enterprise Linux SLE 15 SP2 and later Ubuntu 20.04 LTS and later Clear Linux Latest
VMware	VMWare ESXi
VMM	Linux KVM Microsoft Azure Hyper-V: Windows Server 2016 LTSC Hyper-V: Windows Server 2019 LTSC
Microsoft	Windows Server 2016 LTSC Windows Server 2019 LTSC Windows Server 19H1 Windows Server 19H2 Windows Server 20H1 Windows Server 20H2

#### IMPORTANT NOTE (PLEASE READ):

- The information provided above is subject to change without notice.
- The listing of software in the table indicates only that it is being used by Intel for platform validation.
- The listing of a software in the table does not imply Intel has completed platform validation with it, or that the platform has full support from the software supplier.
- Please contact your software supplier or check their web site on latest and future software support for the Intel Xeon Scalable platform.
- Intel does not validate or certify software on the Intel Xeon Scalable platform for system manufacturers (e.g. OEMs or ODMs). These customers are responsible for certification of software for their platforms and system. They may validate additional operating systems and/or platform features that are not validated by Intel and claim support on their platforms.



## Four and eight-socket solutions

3rd Gen Intel Xeon Scalable Processors Product 30-3-30

Built for today's Al-infused, data-intensive services used in 4 and 8-socket solutions

1.93X

Higher Al performance with intel DL boost<sup>1</sup>

For image classification vs. Intel Xeon Platinum 8280 processors

### Breakthrough memory

Intel Optane persistent memory 200 series

#### Built-in Al acceleration

Intel Deep Learning Boost with bfloat16\*



Targeted for 4S-8S systems

1.93X

Higher performance<sup>2</sup> on cloud data analytics usage models vs. 5-year-old system

Flexibility

Enhanced

Intel Speed Select Technology

#### Business resilience with hardware-enhanced security



Intel Platform firmware resilience



Intel Security Essentials



Intel Security Libraries for Data Center



Intel QuickAssist Technology

Agile service delivery with enhanced efficiency



Intel Speed Select Ir Technology



Intel Optane persistent memory 200 series



Application Device
Oueues

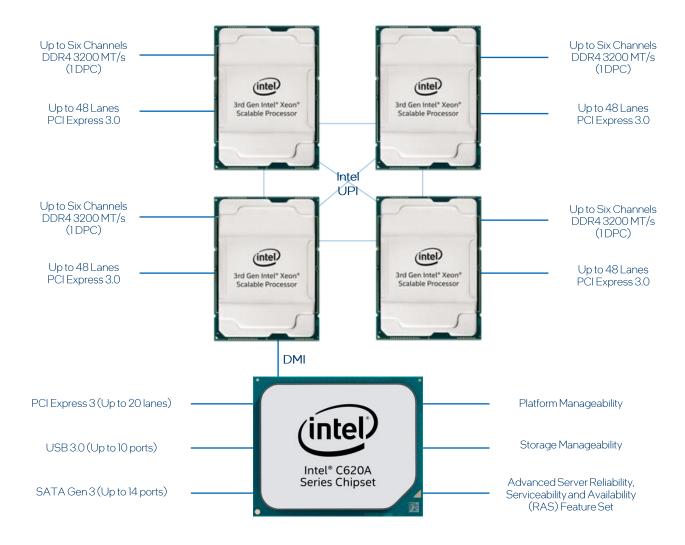


Intel Infrastructure Management Technologies

1, 2 For more complete information about performance and benchmark results, visit <u>www.intel.com/3rd-gen-xeon-configs</u>

<sup>\*</sup>Available on select 3<sup>rd</sup> Gen Intel Xeon Scalable processors

Built for today's Al-infused, data-intensive services used in 4 and 8-socket solutions Intel Xeon Platinum 8300 H/HL, Gold 6300 H/HL, and Gold 5300 H processors



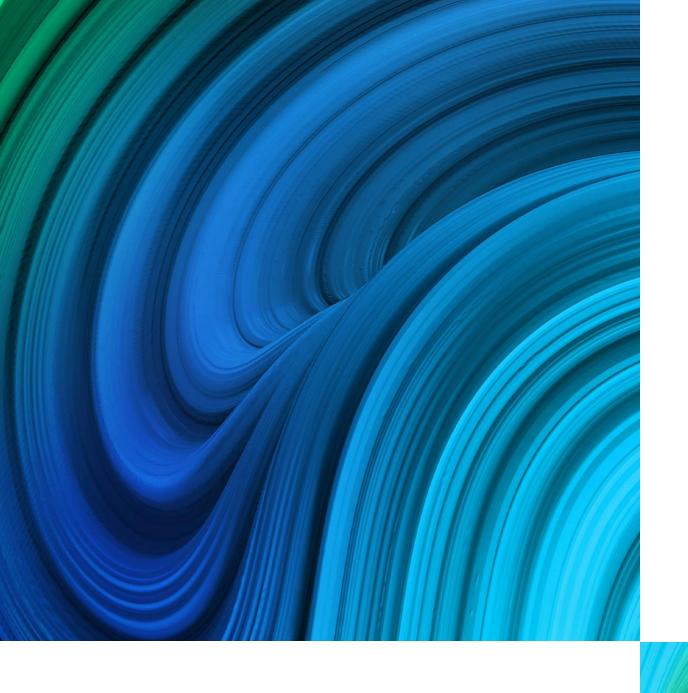
Maximum Core Count Supported	Up to 28
Maximum Base Frequency Supported	Up to 3.9 GHz
Maximum Single-Core Intel Turbo Boost Technology 2.0 Frequency Supported	Up to 4.4 GHz
Processor Cache Memory Support	Up to 38.5 MB of L3 Cache (1.375 MB/core)
Intel Ultra Path Interconnect (Intel UPI)	Six Intel UPI supported per processor (Up to 10.4 GT/s)
Processor Performance Support	Intel Turbo Boost 2.0 Technology, Intel Hyper-Threading Technology (Intel HT), Intel Speed Shift Technology, Enhanced Intel SpeedStep Technology
Intel Speed Select Technology (Intel SST)	Intel SST Core Power (SST-CP) and Intel SST Turbo Frequency (SST-TF) features are supported on select SKUs
Intel Deep Learning Boost	VNNI and bfloat16 numeric format supported
Intel Advanced Vector Extension 512 (Intel AVX-512) Support	Intel AVX-512 with up to 2 FMA support
Number of Processor Sockets Supported	4 or 8 sockets
Thermal Design Point (TDP)	Supporting a range from 150W to 250W depending on customer needs and implementation
Socket Type	FCLGA4189 P5, 77.5x56.5 mm, 4189 pins
System Memory Support	6 channels, Up to DDR4 3200 MT/s (1 DPC) with ECC support 12 DIMMs/socket, Support for 16Gb density DIMMs
Intel Optane Persistent Memory Support	Intel Optane persistent memory 200 series supported on 4- socket systems only
Maximum DRAM System Memory Supported	Up to 4.5 TB per socket
Supported Chipsets	Intel C620A series chipsets, available with integrated Intel QuickAssist Technology
PCHI/O	PCI Express 3 – Up to 20 lanes (8 GT/s) USB 3.0 – Up to 10 ports SATA 3.0 – Up to 14 ports DMI – Up to 4 lanes, Gen 3
Platform Manageability	Intel Server Platform Services (Intel SPS) and Intel Resource Director Technology (Intel RDT)

Software information for use in four and eight-socket solutions

Software Type / Supplier	Supported
Linux	<ul> <li>Red Hat Enterprise Linux 7.8 and later 7.x branches</li> <li>Red Hat Enterprise Linux 8.2 and later 8.x branches</li> <li>SUSE Enterprise Linux SLE 15 SP2 and later</li> <li>Ubuntu 20.04 LTS and later</li> <li>Clear Linux Latest</li> </ul>
VMware	VMWare ESXi
VMM	<ul> <li>Linux KVM</li> <li>Microsoft Azure</li> <li>Hyper-V: Windows Server 2016 LTSC</li> <li>Hyper-V: Windows Server 2019 LTSC</li> </ul>
Microsoft	<ul> <li>Windows Server 2016 LTSC</li> <li>Windows Server 2019 LTSC</li> <li>Windows Server 19H1</li> <li>Windows Server 19H2</li> <li>Windows Server 20H1</li> </ul>

#### IMPORTANT NOTE (PLEASE READ):

- The information provided above is subject to change without notice.
- The listing of software in the table indicates only that it is being used by Intel for platform validation.
- The listing of a software in the table does not imply Intel has completed platform validation with it, or that the platform has full support from the software supplier.
- Please contact your software supplier or check their web site on latest and future software support for the Intel Xeon Scalable platform.
- Intel does not validate or certify software on the Intel Xeon Scalable platform for system manufacturers (e.g. OEMs or ODMs). These customers are responsible for certification of software for their platforms and system. They may validate additional operating systems and/or platform features that are not validated by Intel and claim support on their platforms.

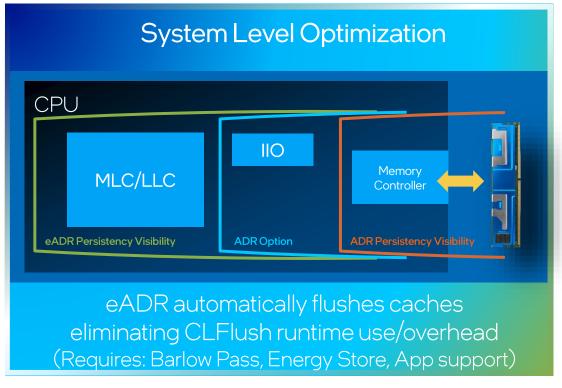


# Intel Portfolio

3rd Gen Intel Xeon Scalable Processors Product 30-3-30

# Intel® Optane™ Persistent Memory 200 Series (Barlow Pass)





### Intel® Optane™ Persistent Memory 200 Series

- Consistent developer target: capacity, latency...
- Increased bandwidth, better power efficiency
- Cross system innovation for increased application performance



# Intel® Optane™ SSD P5800X

### The world's fastest data center SSD



intel. **OPTANE** 

Optane SSDs accelerate slower bulk storage to increase responsiveness of HCl, VDl, databases and more

(-) Alibaba Cloud

Optimized their cloud storage solution with a minimal number of P5800X's per server to, so they can focus on streamlining bulk storage



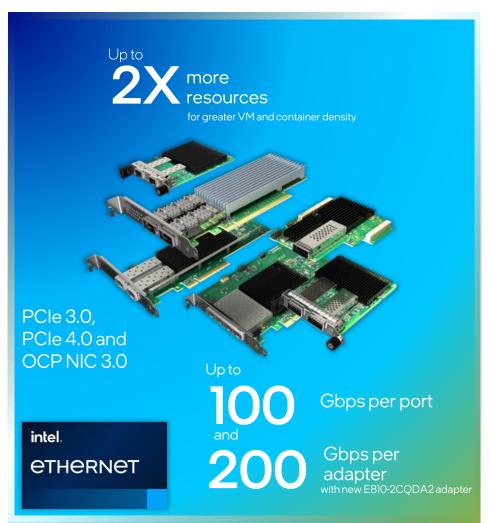
V A S T 20% faster writes on P5800X vs previous gen

Excelero Achieves a 2.5x bandwidth improvement when moving data to QLC capacity storage

Performance varies by use, configuration and other factors. Configurations see appendix [46]

# Intel® Ethernet 800 Series Network Adapters

### Connectivity made flexible



Prioritizes application traffic to help deliver the performance required for high-priority, network intensive workloads

Fully programmable pipeline to enable frame classification for advanced and proprietary protocols

Supports RDMA over iWARP or RoCEv2 protocols and NVMe over TCP with ADQ for high throughput, low latency storage, cloud and HPC clusters

New E810-2CQDA2 adapter targets high-performance workloads such as vRAN, NFV forwarding plane, storage, HPC, cloud and content delivery networks

Source: 2x more resources from 100Gbps to 200Gpbs

## Intel® Ethernet Network Adapter

E810-2CQDA2

#### Hardware Features

Up to 200Gbps of bandwidth in a single PCle 4.0 x16 slot

Two QSFP28 ports enable active/active configuration with up to 100Gpbs on each port

Uses PCle slot bifurcation to enable the functionality of two 100Gbps adapters in one slot

### Key Use Cases

- Bandwidth-intensive Comms workloads, including 5G UPF, vRAN, and CDN
- Cloud workloads, including edge services, database applications, caching servers
- High-bandwidth Storage targets
- HPC/Al fabrics

# Intel® Ethernet 800 Series Features

- Flexible port configurations: 2x100Gb, 2x50Gb
- Application Device Queues: improved application response time predictability
- Dynamic Device Personalization: enhanced packet classification capabilities improve throughput
- RDMA iWARP and RoCEv2: highspeed, low latency connectivity

Up to 200Gbps in a single PCle slot for bandwidth-intensive workloads

# Intel® Agilex™ FPGA + Quartus Prime 20.4 Software FPGA performance made flexible



Industry-leading FPGA offering customers the adaptability and agility to innovate for diverse workload transitions in 5G, Network, Cloud, and Edge

~2X better fabric performance per watt compared to competing 7nm FPGAs enabling flexible, energy efficient designs for the data center and beyond

50% faster Video IP performance vs. competing 7nm FPGAs for video processing applications

Up to 49% faster fabric performance for high-speed 5G fronthaul gateway applications critical to enable high speed fiber-based connections

Performance varies by use, configuration and other factors. Configurations see appendix [48]



## Technologies & Capabilities

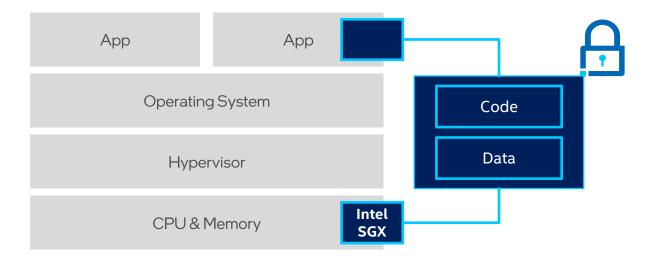
3rd Gen Intel Xeon Scalable Processors Product 30-3-30



# Intel Software Guard Extensions (Intel SGX)

Enables privacy assurances for sensitive data segments without compromising performance

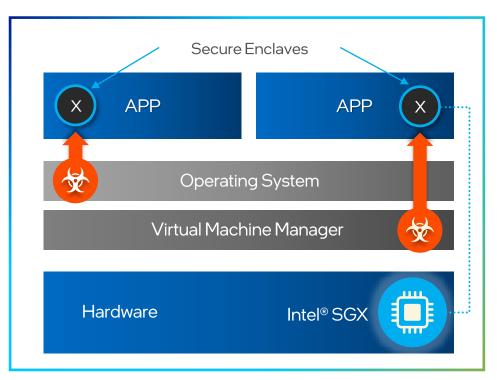
Huge enclaves now support demands of mainstream workloads (up to 1TB memory spaces)



### Trusted Execution

### Intel Software Guard Extensions (SGX)

Helps provide enhanced security protections for application data independent of operating system or hardware configuration



- Helps protect against SW attacks even if OS/drivers/BIOS/VMM/SMM are compromised
- Helps increase protections for secrets (data/keys/et al) even when attacker has full control of platform
- Helps prevent attacks such as memory bus snooping, memory tampering, and "cold boot" attacks against memory contents in RAM
- Provides an option for hardwarebased attestation capabilities to measure and verify valid code and data signatures

Minimally-sized Trusted Compute Base (TCB)

Other technologies allow some privileged SW in their trust boundary

Helps enhance protections for hard-to-protect spaces

Helps increase transparency and accountability

## Intel AVX-512

For demanding computational workloads like 3D modeling, audio and video processing and financial analytics

Financial Services Higher Monte Carlo FSI Performance

50%

2 Socket Intel® Xeon® Platinum 8380 (40C) vs. 2S AMD EPYC 7763 (64C)

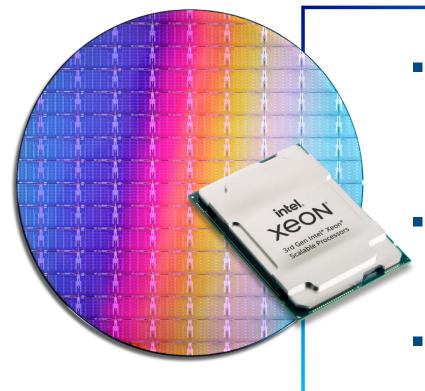
Molecular Dynamics Higher NAMD Performance

27%

2 Socket Intel® Xeon® Platinum 8380 (40C) vs. 2S AMD EPYC 7763 (64C)

For workloads and configurations see [36,37] at https://www.intel.com/3gen-xeon-config. Results may vary.

# Summary



 Intel's highest performing data center processor with built-in security and Al and crypto acceleration

 Unmatched portfolio of hardware and software solutions to move, store and process data

 Broadest ecosystem and decades of experience to ease customer deployments



3rd Gen Intel Xeon Scalable Processors Product 30-3-30

#### Notices and Disclaimers

Performance varies by use, configuration and other factors. Learn more at <a href="https://www.lntel.com/PerformanceIndex">www.lntel.com/PerformanceIndex</a>.

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.

Intel contributes to the development of benchmarks by participating in, sponsoring, and/or contributing technical support to various benchmarking groups, including the BenchmarkXPRT Development Community administered by Principled Technologies.

Your costs and results may vary.

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

Some results may have been estimated or simulated.

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

All product plans and roadmaps are subject to change without notice.

Statements in this document that refer to future plans or expectations are forward-looking statements. These statements are based on current expectations and involve many risks and uncertainties that could cause actual results to differ materially from those expressed or implied in such statements. For more information on the factors that could cause actual results to differ materially, see our most recent earnings release and SEC filings at www.intc.com.

© Intel Corporation. Intel, the Intel logo, and other Intel marks are trademarks of Intel Corporation or its subsidiaries. Other names and brands may be claimed as the property of others.

- 1. 1.46x average performance gain- Ice Lake vs Cascade Lake: Geomean of 1.5x SPECrate2017\_int\_base (est), 1.52x SPECrate2017\_fp\_base (est), 1.47x Stream Triad, 1.38x Intel distribution of LINPACK. Platinum 8380: 1-node, 2x Intel® Xeon® Platinum 8380 processor on Coyote Pass with 512 GB (16 slots/32GB/3200) total DDR4 memory, ucode 0x261, HT on (SPECcpu2017), off (others), Turbo on, Ubuntu 20.04, 5.4.0-66-generic, 1x S4610 SSD 960G, SPECcpu2017 v1.1.0, Stream Triad, Linpack, ic19.1u2, MPI: Version 2019u9; MKL:2020.4.17, test by Intel on 3/15/2021. Platinum 8280: 1-node, 2x Intel® Xeon® Platinum 8280 processor on Wolf Pass with 384 GB (12 slots/32GB/2933) total DDR4 memory, ucode 0x5003003, HT on (SPECcpu2017), off (others), Turbo on, Ubuntu 20.04, 5.4.0-62-generic, 1x S3520 SSD 480G, SPECcpu2017 v1.1.0, Stream Triad, Intel distribution of LINPACK, ic19.1u2, MPI: Version 2019u9; MKL:2020.4.17, test by Intel on 2/4/2021.
- 2. 1.54x average performance gain Ice Lake vs Skylake: Geomean of 1.6x SPECrate2017\_int\_base (est), 1.62x SPECrate2017\_fp\_base (est), 1.52x Stream Triad, 1.44x Intel distribution of LINPACK. 3rd Gen Intel® Xeon® Platinum 8380: 1-node, 2x Intel® Xeon® Platinum 8380 processor on Coyote Pass with 512 GB (16 slots/32GB/3200) total DDR4 memory, ucode 0x261, HT on (SPECcpu2017), off (others), Turbo on, Ubuntu 20.04, 5.4.0-66-generic, 1x S4610 SSD 960G, SPECcpu2017 v1.1.0, Stream Triad, Linpack, ici9.1u2, MPI: Version 2019u9; MKL:2020.4.17, test by Intel on 3/15/2021. Intel® Xeon® Platinum 8180: 1-node, 2x Intel® Xeon® Platinum 8180 processor on Wolf Pass with 192 GB (12 slots/16GB/2933[2666]) total DDR4 memory, ucode 0x2006a08, HT on (SPECcpu2017), off (others), Turbo on, Ubuntu 20.04, 5.4.0-62-generic, SPECcpu2017 v1.1.0, Stream Triad, Intel distribution of LINPACK, ici9.1u2, MPI: Version 2019 Update 9 Build 20200923; MKL: psxe runtime 2020.4.17, test by Intel on 1/27/21.
- 3. 2.65x average performance gain Ice Lake vs Broadwell: Geomean of 2.34x SPECrate2017\_int\_base (est), 2.5x SPECrate2017\_fp\_base (est), 2.55x Stream Triad, 3.18x Intel distribution of LINPACK. 3d Gen Intel® Xeon® Platinum 8380: I-node, 2x Intel® Xeon® Platinum 8380 processor on Coyote Pass with 512 GB (16 slots/32GB/3200) total DDR4 memory, ucode 0x261, HT on (SPECcpu2017), off (others), Turbo on, Ubuntu 20.04, 5.4.0-66-generic, 1x S4610 SSD 960G, SPECcpu2017 v1.1.0, Stream Triad, Linpack, ici9.1u2, MPI: Version 2019u9; MKL:2020.4.17, test by Intel on 3/15/2021. Intel® Xeon® E5-2699v4: I-node, 2x Intel® Xeon® processor E5-2699v4 on Wildcat Pass with 256 GB (8 slots/32GB/2400) total DDR4 memory, ucode 0x038, HT on (SPECcpu2017), off (others), Turbo on, Ubuntu 20.04, 5.4.0-62-generic, 1x S3700 400GB SSD, SPECcpu2017 v1.1.0, Stream Triad, Intel distribution of LINPACK, ici9.1u2, MPI: Version 2019 Update 9 Build 20200923; MKL: psxe runtime 2020.4.17, test by Intel on 1/17/21.
- 4. 3.14x average performance gain Ice Lake vs Haswell: Geomean of 2.85x SPECrate2017\_int\_base (est), 3.08x SPECrate2017\_fp\_base (est), 2.8x Stream Triad, 3.97x Intel distribution of LINPACK. 3rd Gen Intel® Xeon® Platinum 8380: I-node, 2x Intel® Xeon® Platinum 8380 processor on Coyote Pass with 512 GB (16 slots/32GB/3200) total DDR4 memory, ucode 0x261, HT on (SPECcpu2017), off (others), Turbo on, Ubuntu 20.04, 5.4.0-66-generic, 1x S4610 SSD 960G, SPECcpu2017 v1.1.0, Stream Triad, Linpack, ic19.1u2, MPI: Version 2019u9; MKL:2020.4.17, test by Intel on 3/15/2021. Intel® Xeon® E5-2699v3: I-node, 2x Intel® Xeon® processor E5-2699v3 on Wildcat Pass with 256 GB (8 slots/32GB/2666[2133]) total DDR4 memory, ucode 0x44, HT on (SPECcpu2017), off (others), Turbo on, Ubuntu 20.04, 5.4.0-62-generic, 1x S3700 400GB SSD, SPECcpu2017 v1.1.0, Stream Triad, Intel distribution of LINPACK, ic19.1u2, MPI: Version 2019 Update 9 Build 20200923; MKL: psxe runtime 2020.4.17, test by Intel on 2/3/21.
- 5. BERT-Large SQuAD: 1.45x higher INT8 real-time inference throughput & 1.74x higher INT8 batch inference throughput & 1.22x performance/core: Platinum 8380: I-node, 2x Intel® Xeon® Platinum 8380 processor on Coyote Pass with 512 GB (16 slots/32GB/3200) total DDR4 memory, ucode 0x261, HT on, Turbo on, Ubuntu 20.04 LTS, 5.4.0-65-generic, 1x Intel\_SSDSC2KG96, Intel® SSDPE2KX010T8, BERT-Large SQuAD, gcc-9.3.0, oneDNN1.6.4, BS=1,128 INT8, TensorFlow-2.5 (container-intel/intel-optimized-tensorflow:tf-r2.5-icx-b631821f), Model zoo: https://github.com/IntelAl/models/tree/icx-launch-public/quickstart/, test by Intel® No. 1, 12/2021. Platinum 8280: I-node, 2x Intel® Xeon® Platinum 8280 processor on Wolf Pass with 384 GB (12 slots/32GB/2933) total DDR4 memory, ucode 0x5003003, HT on, Turbo on, Ubuntu 20.04 LTS, 5.4.0-48-generic, 1x Samsung\_SSD\_860, Intel® SSDPE2KX040T8, BERT-Large SQuAD, gcc-9.3.0, oneDNN1.6.4, BS=1,128 INT8, TensorFlow-2.5 (container-intel/intel-optimized-tensorflow:tf-r2.5-icx-b631821f), Model zoo: https://github.com/IntelAl/models/tree/icx-launch-public/quickstart/, test by Intel on 2/17/2021.
- 6. MobileNet-v1: 1.59x higher INT8 real-time inference throughput & 1.66x higher INT8 batch inference & 1.16x performance/core throughput: Platinum 8380: 1-node, 2x Intel® Xeon® Platinum 8380 processor on Coyote Pass with 512 GB (16 slots/32GB/3200) total DDR4 memory, ucode 0x261, HT on, Turbo on, Ubuntu 20.04 LTS, 5.4.0-65-generic, 1x Intel\_SSDSC2KG96, Intel® SSDPE2KX010T8, MobileNet-v1, gcc-9.3.0, oneDNN 1.6.4, BS=1,56 INT8, TensorFlow-2.5 (container-intel/intel-optimized-tensorflow:tf-r2.5-icx-b631821f), Model zoo: https://github.com/IntelAl/models/tree/icx-launch-public/quickstart/, test by Intel on 3/12/2021. Platinum 8280: 1-node, 2x Intel® Xeon® Platinum 8280 processor on Wolf Pass with 384 GB (12 slots/32GB/2933) total DDR4 memory, ucode 0x5003003, HT on, Turbo on, Ubuntu 20.04 LTS, 5.4.0-48-generic, 1x Samsung\_SSD\_860, Intel® SSDPE2KX040T8, MobileNet-v1, gcc-9.3.0, oneDNN1.6.4, BS=1,56 INT8, TensorFlow-2.5 (container-intel/intel-optimized-tensorflow:tf-r2.5-icx-b631821f), Model zoo: https://github.com/IntelAl/models/tree/icx-launch-public/quickstart/, test by Intel on 2/17/2021.
- 7. ResNet-50 v1.5: 1.52x higher INT8 real-time inference throughput & 1.56x higher INT8 batch inference throughput on Ice Lake vs. prior generation Cascade Lake Platinum 8380: I-node, 2x Intel® Xeon® Platinum 8380 processor on Coyote Pass with 512 GB (16 slots/32GB/3200) total DDR4 memory, ucode 0x261, HT on, Turbo on, Ubuntu 20.04 LTS, 5.4.0-65-generic, 1x Intel\_SSDSC2KG96, Intel® SSDPE2KX010T8, ResNet-50 v1.5, gcc-9.3.0, oneDNN1.6.4, BS=1,128 INT8, TensorFlow- 2.5 (container-intel/intel-optimized-tensorflow:tf-r2.5-icx-b631821f), Model zoo: https://github.com/IntelAl/models/tree/icx-launch-public/quickstart/, test by Intel on 3/12/2021. Platinum 8280: I-node, 2x Intel® Xeon® Platinum 8280 processor on Wolf Pass with 384 GB (12 slots/32GB/2933) total DDR4 memory, ucode 0x5003003, HT on, Turbo on, Ubuntu 20.04 LTS, 5.4.0-48-generic, 1x Samsung\_SSD\_860, Intel® SSDPE2KX040T8, ResNet-50 v1.5, gcc-9.3.0, oneDNN1.6.4, BS=1,128 INT8, TensorFlow-2.5 (container-intel/intel-optimized-tensorflow:tf-r2.5-icx-b631821f), Model zoo: https://github.com/IntelAl/models/tree/icx-launch-public/quickstart/, test by Intel on 2/17/2021.

- 8. 1.72x higher virtualization performance & 1.2x performance/core: Platinum 8380: I-node, 2x Intel® Xeon® Platinum 8380 processor on Coyote Pass with 2048 GB (32 slots/64GB/3200) total DDR4 memory, ucode 0x261, HT on, Turbo on, RedHat 8.3, 4.18.0-240.el8.x86\_64, 1x S4610 SSD 960G, 4x P5510 3.84TB NVME, 2x Intel E810, Virtualization workload, Qemu-kvm4.2.0-34 (inbox), WebSphere 8.5.5, DB2 v9.7, Nginx 1.14.1, test by Intel on 3/14/2021. Platinum 8280: 1-node, 2x Intel® Xeon® Platinum 8280 processor on Wolf Pass with 1536 GB (24 slots/64GB/2933[2666]) total DDR4 memory, ucode 0x5003005, HT on, Turbo on, RedHat 8.1 (Note: selected higher of RedHat 8.1 and 8.3 scores for baseline), 4.18.0-147.el8.x86\_64, 1x S4510 SSD 240G, 4x P4610 3.2TB NVME, 2x Intel XL710, Virtualization workload, Qemu-kvm 4.2.0-34 (inbox), WebSphere 8.5.5, DB2 v9.7, Nginx 1.14.1, test by Intel on 12/22/2020.
- 9. 1.62x higher throughput under SLA for Server Side Java & 1.13x performance/core: Platinum 8380: 1-node, 2x Intel® Xeon® Platinum 8380 processor on Coyote Pass with 512 GB (16 slots/32GB/3200) total DDR4 memory, ucode 0x261, HT on, Turbo on, Ubuntu 20.04.1 LTS, 5.4.0-64-generic, 1x SSDSC2BA40, Java workload, JDK 1.15.0.1, test by Intel on 3/15/2021. Platinum 8280: 1-node, 2x Intel® Xeon® Platinum 8280 processor on Wolf Pass with 384 GB (12 slots/32GB/2933) total DDR4 memory.ucode 0x5003003. HT on, Turbo on, Ubuntu 20.04.1 LTS, 5.4.0-64-generic, 1x INTEL SSDSC2KG01, Java workload, JDK 1.15.0.1, test by Intel on 2/18/2021.
- 10. 1.64x HammerDBMySQL & 1.15x performance/core: Platinum 8380: I-node, 2x Intel® Xeon® Platinum 8380 processor on Coyote Pass with 512 GB (16 slots/32GB/3200) total DDR4 memory, ucode 0x261, HT on, Turbo on, Redhat 8.3, 4.18.0-240.el8.x86\_64 x86\_64, 1x Intel® SSD 960GB OS Drive, 1x Intel P5800 1.6T, onboard 1G/s, HammerDB 4.0, MySQL 8.0.22, test by Intel on 3/11/2021. Platinum 8280: I-node, 2x Intel® Xeon® Platinum 8280 processor on Wolf Pass with 384 GB (12 slots/32GB/2933) total DDR4 memory, ucode 0x5003003, HT on, Turbo on, Redhat 8.3, 4.18.0-240.el8.x86\_64, 1x Intel 240GB SSD OS Drive, 1x Intel 6.4T P4610, onboard 1G/s, HammerDB 4.0, MySQL 8.0.22, test by Intel on 2/5/2021.
- 11. 1.48x higher responses on WordPress with HTTPS: Platinum 8380: 1-node, 2x Intel® Xeon® Platinum 8380 processor on Coyote Pass with 512 GB (16 slots/32GB/3200) total DDR4 memory, ucode 0x261, HT on, Turbo on, Ubuntu 20.04, 5.4.0-65-generic, 1x Intel 895GB SSDSC2KG96, 1x XL710-Q2, WordPress 4.2 with HTTPS, gcc 9.3.0, GLIBC 2.3Houbuntu9.1, mysqld Ver 10.3.25-MariaDB-Oubuntu0.20.04.1, PHP 7.4.9-dev (fpm-fcgi), Zend Engine v3.4.0, test by Intel on 3/15/2021. Platinum 8280: 1-node, 2x Intel® Xeon® Platinum 8280 processor on Wolf Pass with 384 GB (12 slots/32GB/2933) total DDR4 memory, ucode 0x5003003, HT on, Turbo on, Ubuntu 20.04, 5.4.0-65-generic, 1x Intel 1.8T SSDSC2KG01, 1x Intel X722, test by Intel on 2/5/2021.
- 12. 1.65x higher responses with CloudXPRT Data Analytics & 1.15x performance/core: Platinum 8380: Honde, 2x Intel® Xeon® Platinum 8380 processor on Coyote Pass with 512 GB (16 slots/32GB/3200) total DDR4 memory, ucode 0x261, HT on, Turbo on, Ubuntu 20.04, 5.4.0-65-generic, 1x S4610 SSD 960G, CloudXPRT v1.0, Data Analytics (Analytics per minute @ p.95 <= 90s), test by Intel on 3/12/2021. Platinum 8280: Honde, 2x Intel® Xeon® Platinum 8280 processor on Wolf Pass with 384 GB (12 slots/32GB/2933) total DDR4 memory, ucode 0x5003003, HT on, Turbo on, Ubuntu 20.04, 5.4.0-65-generic, 1x S3520 SSD 480G, CloudXPRT v1.0, test by Intel on 2/4/2021. Intel contributes to the development of benchmarks by participating in, sponsoring, and/or contributing technical support to various benchmarking groups, including the BenchmarkXPRT Development Community administered by Principled Technologies.
- 13. 1.58x higher responses with CloudXPRT Web Microservices: Platinum 8380: 1-node, 2x Intel® Xeon® Platinum 8380 processor on Coyote Pass with 512 GB (16 slots/32GB/3200) total DDR4 memory, ucode 0x261, HT on, Turbo on, Ubuntu 20.04, 5.4.0-65-generic, 1x S4610 SSD 960G, CloudXPRT v1.0, Web Microservices (Requests per minute @ p.95 latency <= 3s), test by Intel on 3/12/2021. Platinum 8280: 1-node, 2x Intel® Xeon® Platinum 8280 processor on Wolf Pass with 384 GB (12 slots/32GB/2933) total DDR4 memory, ucode 0x5003003, HT on, Turbo on, Ubuntu 20.04, 5.4.0-54-generic, 1x S3520 SSD 480G, CloudXPRT v1.0, test by Intel on 2/4/2021. Intel contributes to the development of benchmarks by participating in, sponsoring, and/or contributing technical support to various benchmarking groups, including the BenchmarkXPRT Development Community administered by Principled Technologies.
- 14. 1.3x DDIO on/off with DPDK L3 Packet Forwarding: 1-node, 2(1 socket used)x Intel® Xeon® Gold 6338N on Intel® Whitley with 128 GB (8 slots/16GB/2666) total DDR4 memory, ucode 0x261, HT on, Turbo off, Ubuntu 20.04 LTS (Focal Fossa), 5.4.0-40-generic, 1x INTEL\* 240G SSD, 1x E810-2CQDA2 (Chapman Beach), v20.08.0, Gcc 9.3.0, DPDKL3FWD (4c8t), PCle Writes = Non Allocating for DDIO off case, test by Intel on 3/24/2021
- 15. 1.62x AVX-512 vs AVX2 on Linpack: 1-node, 2x Intel® Xeon® Platinum 8380 on Coyote Pass with 512 GB (12 slots/ 32GB/ 3200) total DDR4 memory, ucode 0x270, HT off, Turbo on, Ubuntu 20.04.2 LTS, 5.4.0-67-generic, 1x INTEL\_SSDSC2BB80, 1x X710, Intel distribution of Linpack with AVX-512, AVX2, test by Intel on 3/24/2021
- 16. 4.3x Intel DL Boost FP32 to INT8: 1-node, 2x Intel® Xeon® Platinum 8380 processor on Coyote Pass with 512 GB (16 slots/32GB/3200) total DDR4 memory, ucode 0x261, HT on, Turbo on, Ubuntu 20.04 LTS, 5.4.0-64-generic, 1x Intel® SSDSC2KG960G7, 1x Intel® SSDSC2KG960G7, 1x Intel® SSDSC2KG960G7, ResNet-50 v1.5, gcc-9.3.0, oneDNN 1.6.4, BS=128, FP32/INT8, TensorFlow-2.5 (container-intel/intel-optimized-tensorflow:tf-r2.5-icx-b631821f), Model zoo: https://github.com/IntelAl/models/tree/icx-launch-public/guickstart/, test by Intel on 3/19/2021.

- 17. 1.62x average network performance gains: geomean of Virtual Broadband Network Gateway, 5G User Plane Function, Virtual Cable Modem Termination System, Vector Packet Processing Forward Information Base 512B, DPDK L3 Forward 512B, CDN-Live, Vector Packet Processing IP Security 512B.
  - a. 1.2x Virtual Broadband Network Gateway: Gold 6338N: 1-node, 2(1 socket used)x Intel® Xeon® Gold 6338N on Intel® Whitley with 256 GB (16 slots/16GB/ 2666) total DDR4 memory, ucode 0x261, HT on, Turbo off, Ubuntu 20.04 LTS (Focal Fossa), 5.4.0-40-generic, 1x INTEL\* 240G SSD, 3x E810-CQDA2 (Tacoma Rapids), vBNG 20.07, Gcc 9.3.0, test by Intel on 3/11/2021. Gold 6252N: 1-node, 2(1 socket used)x Intel® Xeon® Gold 6252N on SuperMicro\* X11DPG-QT with 192 GB (12 slots/16GB/ 2933) total DDR4 memory, ucode 0x5002f01, HT on, Turbo off, Ubuntu 20.04 LTS (Focal Fossa), 5.4.0-40-generic, 1x INTEL\* 240G SSD, 3x E810-CQDA2 (Tacoma Rapids), vBNG 20.07, Gcc 9.3.0, test by Intel on 2/2/2021.
  - b. 1.42x 5G User Plane Function: 1-node, 2(1 socket used)x Intel® Xeon® Gold 6338N on Whitley Coyote Pass 2U with 128 GB (8 slots/16GB/2666) total DDR4 memory, ucode 0x261, HT on, Turbo off, Ubuntu 18.04.5 LTS, 4.15.0-134-generic, 1x Intel 810 (Columbiaville), FlexCore 5G UPF, Jan' 2021 MD5 checksum: c4ad7f8422298ceb69d0le67419ff1c1, GCC 7.5.0, 5G UPF228 Gbps / 294 Gbps, test by Intel on 3/16/2021. 1-node, 2(1 socket used)x Intel® Xeon® Gold 6252N on SuperMicro\* X11DPG-QT with 96 GB (6 slots/16GB/2933) total DDR4 memory, ucode 0x5003003, HT on, Turbo off, Ubuntu 18.04.5 LTS, 4.15.0-132-generic, 1x Intel 810 (Columbiaville), FlexCore 5G UPF, Jan' 2021 MD5 checksum: c4ad7f8422298ceb69d0le67419ff1c1, GCC 7.5.0, 5G UPF161 Gbps / 213 Gbps, test by Intel on 2/12/2021.
  - c. 1.63x CDN Live: 1 node, 2x Intel® Xeon® Gold 6338N Processor, 32 core HT ON Turbo ON, Total DRAM 256GB (16 slots/16GB/2666MT/s), Total Optane Persistent Memory 200 Series 2048GB (16 slots/128GB/2666MT/s), BIOS SE5C6200.86B.2021.D40.2103100308 (ucode: 0x261), 4x Intel® E810, Ubuntu 20.04, kernel 5.4.0-65-generic, gcc 9.3.0 compiler, openssl 1.1.1h, varnish-plus 6.0.7r2. 2 clients, Test by Intel as of 3/11/2021. Gold 6252N: 2x Intel® Xeon® Gold 6252N Processor, 24 core HT ON Turbo ON, Total DRAM 192GB (12 slots/16GB/2666MT/s), Total Optane Persistent Memory 100 Series 1536GB(12 slots/128GB/2666MT/s), 1x Mellanox MCX516A-CCAT, BIOS: SE5C620.86B.02.01.0013.121520200651 (ucode: 0x5003003), Ubuntu 20.04, kernel 5.4.0-65-generic, wrk master 4/17/2019. Test by Intel as of 2/15/2021. Throughput measured with 100% Transport Layer Security (TLS) traffic with 93.3% cache hit ratio and keep alive on, 512 total connections.
  - d. 1.66 Vector Packet Processing Forward Information Base 512B:1-node, 2(1 socket used)x Intel® Xeon® Gold 6338N on Intel® Whitley with 128 GB (8 slots/16GB/2666) total DDR4 memory, ucode 0x261, HT on, Turbo off, Ubuntu 20.04 LTS (Focal Fossa), 5.4.0-40-generic, 1x INTEL\*240G SSD, 1x E810-2CQDA2 (Chapman Beach), v20.05.1-release, Gcc 9.3.0, VPPFIB(24c24t), test by Intel on 3/17/2021. 1-node, 2(1 socket used)x Intel® Xeon® Gold 6252N on SuperMicro\*X11DPG-QT with 96 GB (6 slots/16GB/2933) total DDR4 memory, ucode 0x5002f01, HT off, Turbo off, Ubuntu 20.04 LTS (Focal Fossa), 5.4.0-40-generic, 1x INTEL\*240G SSD, 1x E810-CQDA2 (Tacoma Rapids), v20.05.1-release, Gcc 9.3.0, VPPFIB (18c18t), test by Intel on 2/2/2021.
  - e. 1.72x Virtual Cable Modem Termination System: Gold 6338N: 1-node, 2(1 socket used)x Intel® Xeon® Gold 6338N on Coyote Pass with 256 GB (16 slots/16GB/2666) total DDR4 memory, ucode 0x261, HT on, Turbo off(no SST-BF)/on(SST-BF), Ubuntu 20.04 LTS (Focal Fossa), 5.4.0-40-generic, 1x INTEL\*240G SSD, 3x E810-CQDA2 (Tacoma Rapids), vCMTS 20.10, Gcc 9.3.0, SST-BF (2.4 Ghz,1.9 Ghz frequencies for the priority cores and the other cores respectively), test by Intel on 3/11/2021. Gold 6252N: 1-node, 2(1 socket used)x Intel® Xeon® Gold 6252N on SuperMicro\* X11DPG-QT with 192 GB (12 slots/16GB/2933) total DDR4 memory, ucode 0x5002f01, HT on, Turbo off, Ubuntu 20.04 LTS (Focal Fossa), 5.4.0-40-generic, 1x INTEL\*240G SSD, 2x E810-CQDA2 (Tacoma Rapids), vCMTS 20.10, Gcc 9.3.0, vCMTS 90 (14 instances), test by Intel on 2/2/2021.
  - f. 1.88x DPDK L3 Forward 512B & 1.41x performance/core:1-node, 2(1 socket used)x Intel® Xeon® Gold 6338N on Intel® Whitley with 128 GB (8 slots/16GB/2666) total DDR4 memory, ucode 0x261, HT on, Turbo off, Ubuntu 20.04 LTS (Focal Fossa), 5.4.0-40-generic, 1x INTEL\*240G SSD, 1x E810-2CQDA2 (Chapman Beach), v20.08.0, Gcc 9.3.0, DPDKL3FWD (24c24t), test by Intel on 3/17/2021, 2(1 socket used)x Intel® Xeon® Gold 6252N on SuperMicro\* X11DPG-QT with 96 GB (6 slots/16GB/2933) total DDR4 memory, ucode 0x5002f01, HT off, Turbo off, Ubuntu 20.04 LTS (Focal Fossa), 5.4.0-40-generic, 1x INTEL\*240G SSD, 1x E810-CQDA2 (Tacoma Rapids), v20.08.0, Gcc 9.3.0, DPDKL3FWD (12c12t). test by Intel on 2/2/2021.
  - g. 1.94x Vector Packet Processing- IP Security 1420B: 1-node, 2(1 socket used)x Intel® Xeon® Gold 6338N on Intel® Whitley with 128 GB (8 slots/16GB/2666) total DDR4 memory, ucode 0x261, HT on, Turbo off, Ubuntu 20.04 LTS (Focal Fossa), 5.4.0-40-generic, 1x INTEL\*240G SSD, 1x E810-2CQDA2 (Chapman Beach), v21.01-release, Gcc 9.3.0, VPPIPSEC(24c24t) test by Intel on 3/17/2021.1-node, 2(1 socket used)x Intel® Xeon® Gold 6252N on SuperMicro\* X11DPG-QT with 96 GB (6 slots/16GB/2933) total DDR4 memory, ucode 0x5002f01, HT off, Turbo off, Ubuntu 20.04 LTS (Focal Fossa), 5.4.0-40-generic, 1x INTEL\*240G SSD, 1x E810-CQDA2 (Tacoma Rapids), v21.01-release, Gcc 9.3.0, VPPIPSEC(18c18t) test by Intel on 2/2/2021.
- 18. 2x MIMO Throughput: Results have been estimated or simulated. Based on 2x estimated throughput from 32Tx32R (5Gbps) on 2nd Gen Intel® Xeon® Gold 6212U processor to 64Tx64R (10Gbps) on 3rd Gen Intel® Xeon® Gold 6338N processor at similar power ~185W.

- 19. 1.53x higher HPC performance (geomean HPL, HPCG, Stream Triad, WRF, Binomial Options, Black Scholes, Monte Carlo, OpenFOAM, GROMACS, LAMMPS, NAMD, RELION)
  - a. 1.53x higher FSI Kernel performance (geomean Binomial Options, Black Scholes, Monte Carlo)
  - b. 1.60x higher Life and Material Science performance (geomean GROMACS, LAMMPS, NAMD, RELION)
  - c. Platform setup: 8380: I-node, 2x Intel® Xeon® Platinum 8380 (40C/2.3GHz, 270W TDP) processor on Intel Software Development Platform with 256 GB (16 slots/16GB/3200) total DDR4 memory, ucode 0x055260, HT on, Turbo on, CentOS Linux 8.3.2011, 4.18.0-240.1.1.el8\_3.crt1.x86\_64, lx Intel\_SSDSC2KG96. 8280: I-node, 2x Intel® Xeon® Platinum 8280 (28C/2.7GHz, 205W TDP) processor on Intel Software Development Platform with 192GB (12 slots/16GB/2933) total DDR4 memory, ucode 0x4002f01, HT on, Turbo on, CentOS Linux 8.3.2011, 4.18.0-240.1.1.el8\_3.crt1.x86\_64, lx Intel\_SSDSC2KG48. tested by Intel and results as of March 2021
  - d. 1.41x higher HPCG performance: App Version: 2019u5 MKL; Build notes: Tools: Intel MKL 2020u4, Intel C Compiler 2020u4, Intel MPI 2019u8; threads/core: 1; Turbo: used; Build knobs: -03-ip-xCORE-AVX-512
  - e. 1.38x higher HPL performance App Version: The Intel Distribution for LINPACK Benchmark; Build notes: Turbo: used; BIOS settings: HT=off Turbo=On SNC=Off
  - f. 1.47x higher Stream Triad Performance: App Version: McCalpin\_STREAM\_OMP-version; Build notes: Turbo: used; BIOS settings: HT=off Turbo=On SNC=On
  - g. 1.58x higher WRF performance &1.11x performance/core: (geomean Conus-12km, Conus-2.5km, NWSC-3-NA-3km) App Version: 4.2.2; Build notes: Intel Fortran Compiler 2020u4, Intel MPI 2020u4; threads/core: 1; Turbo: used; Build knobs:-ip-w-O3-xCORE-AVX2-vec-threshold0-ftz-align array64byte-qno-opt-dynamic-align-fno-alias \$(FORMAT\_FREE) \$(BYTESWAPIO)-fp-model fast=2-fimf-use-svml=true-inline-max-size=12000-inline-max-total-size=30000
  - h. 1.28x higher Binomial Options performance: App Version: v1.0; Build notes: Tools: Intel C Compiler 2020u4, Intel Threading Building Blocks; threads/core: 2; Turbo: used; Build knobs: -03-xCORE-AVX-512-qopt-zmm-usage=high-fimf-domain-exclusion=31-fimf-accuracy-bits=11-no-prec-div-no-prec-sqrt
  - i. 1.67x higher Black Scholes performance: App Version: v1.3; Build notes: Tools: Intel MKL, Intel C Compiler 2020u4, Intel Threading Building Blocks 2020u4; threads/core: 1; Turbo: used; Build knobs: -03 -xCORE-AVX-512 qopt-zmm-usage=high-fimf-precision=low-fimf-domain-exclusion=31-no-prec-div-no-prec-sqrt-fimf-domain-exclusion=31
  - j. 1.70x higher Monte Carlo performance &1.19x performance/core: App Version: v1.1; Build notes: Tools: Intel MKL 2020u4, Intel C Compiler 2020u4, Intel Threading Building Blocks 2020u4; threads/core: 1; Turbo: used; Build knobs: -03 -xCORE-AVX-512 -qopt-zmm-usage=high-fimf-precision=low-fimf-domain-exclusion=31-no-prec-div-no-prec-sqrt
  - k. 1.52x higher OpenFOAM performance: (geomean 20M\_cell\_motorbike, 42M\_cell\_motorbike) App Version: v8; Build notes: Tools: Intel FORTRAN Compiler 2020u4, Intel C Compiler 2020u4, Intel MPI 2019u8; threads/core: 1; Turbo: used; Build knobs: -O3-ip -xCORE-AVX-512. OpenFOAM Disclaimer: 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
  - I. 1.64x higher GROMACS performance: (geomean ion\_channel\_pme, lignocellulose\_rf, water\_pme, water\_rf) App Version: v2020.5\_SP; Build notes: Tools: Intel MKL 2020u4, Intel C Compiler 2020u4, Intel MPI 2019u8; threads/core: 2; Turbo: used; Build knobs: -O3 -ip -xCORE-AVX-512
  - m. 1.60x higher LAMMPS performance: (geomean Polyethylene, Stillinger-Weber, Tersoff, Water) App Version: v2020-10-29; Build notes: Tools: Intel MKL 2020u4, Intel C Compiler 2020u4, Intel Threading Building Blocks 2020u4, Intel MPI 2019u8; threads/core: 2; Turbo: used; Build knobs: -O3 -ip-xCORE-AVX-512-qopt-zmm-usage=high
  - n. 1.57x higher NAMD performance: (geomean Apoal, flatpase, STMV) App Version: 2.15-Alphal (includes AVX tiles algorithm); Build notes: Tools: Intel MKL, Intel C Compiler 2020u4, Intel MPI 2019u8, Intel Threading Building Blocks 2020u4; threads/core: 2; Turbo: used; Build knobs: -ip -fp-model fast=2-no-prec-div-qoverride-limits-gopenmp-simd-O3-xCORE-AVX-512-gopt-zmm-usage=high
  - o. 1.61x higher RELION Plasmodium Ribosome performance: App Version: 3\_1\_1; Build notes: Tools: Intel C Compiler 2020u4, Intel MPI 2019u9; threads/core: 2; Turbo: used; Build knobs: -O3-ip-g-debug inline-debug-info-xCOMMON-AVX-512-qopt-report=5-restrict

- 3.34x higher IPSec AES-GCM performance, 3.78x higher IPSec AES-CMAC performance, 3.84x higher IPSec AES-CTR performance; 8380: I-node, 2x Intel(R) Xeon(R) Platinum 8380 CPU on M50CYP2SB2U with 512 GB (16 slots/32GB/3200) total DDR4 memory, ucode 0x8d055260, HT On, Turbo Off, Ubuntu 20.04.2 LTS, 5.4.0-66-generic, 1x Intel 1.8TB SSD OS Drive, intel-ipsec-mb v0.55, gcc 9.3.0, Glibc 2.31, test by Intel on 3/17/2021. 8280M: I-node, 2x Intel(R) Xeon(R) Platinum 8280M CPU on S2600WFT with 384 GB (12 slots/32GB/2933) total DDR4 memory, ucode 0x4003003, HT On, Turbo Off, Ubuntu 20.04.2 LTS, 5.4.0-66-generic, 1x Intel 1.8TB SSD OS Drive, intel-ipsec-mb v0.55, gcc 9.3.0, Glibc 2.31, test by Intel on 3/8/2021.
- 21. 3.5x higher ISA-L AES-XTS performance, 2.30x higher ISA-L CRC performance: ISA-L: 8380: 1-node, 2x Intel® Xeon® Platinum 8380 Processor, 40 cores HT On Turbo OFF Total Memory 512 GB (16 slots/32GB/3200 MHz), Data protection (Reed Solomon EC (10+4)), Data integrity (CRC64), Hashing (Multibuffer MD5), Data encryption (AES-XTS 128 Expanded Key), Data Compression (Level 3 Compression (Calgary Corpus)), BIOS: SE5C6200.86B.3021.D40.2103160200 (ucode: 0x8d05a260), Ubuntu 20.04.2, 5.4.0-67-generic, gcc 9.3.0 compiler, yasm 1.3.0, nasm 2.14.02, isal 2.30, isal\_crypto 2.23, OpenSSL 1.1.1i, zlib 1.2.11, Test by Intel as of 03/19/2021. 8280: 1-node, 2x Intel® Xeon® Platinum 8280 Processor, 28 cores HT On Turbo OFF Total Memory 384 GB (12 slots/32GB/2933 MHz), BIOS: SE5C620.86B.02.01.0013.121520200651 (ucode: 0x4003003), Ubuntu 20.04.2, 5.4.0-67-generic, gcc 9.3.0 compiler, yasm 1.3.0, nasm 2.14.02, isal 2.30, isal\_crypto 2.23, OpenSSL 1.1.1i, zlib 1.2.11 Test by Intel as of 2/9/2021. Performance measured on single core.
- 22. NVMe-over-TCP IOPS Throughput: Platinum 8380: 1-node, 2x Intel® Xeon® Platinum 8380 Processor, 40 cores HT On Turbo ON Total Memory 1024 GB (16 slots/64GB/3200), BIOS:SE5C6200.86B.2021.D40.2103100308 (ucode:0x261), Fedora 30, Linux Kernel 5.7.12, gcc 9.3.1 compiler, fio 3.20, SPDK 21.01, Storage: 16x Intel® SSD D7-P5510 7.68 TB or 16x Intel® Optane™ SSD 400GB P5800X, Network: 2x 100GbE Intel E810-C, Test by Intel as of 3/17/2021. Platinum 8280: 1-node, 2x Intel® Xeon® Platinum 8280 Processor, 28 cores HT On Turbo ON Total Memory 768 GB (24 slots/32GB/2666), BIOS: SE5C620.86B.02.01.0013.121520200651 (ucode:0x4003003), Fedora 30, Linux Kernel 5.7.12, gcc 9.3.1 compiler, fio 3.20, SPDK 21.01, Storage: 16x Intel® SSD DC P4610 1.6TB, Network: 1x 100GbE Intel E810-C, Test by Intel as of 2/10/2021.
- 23. 2.5x higher transactions on Aerospike Database: Platinum 8368: 1-node, 2x Intel® Xeon® Platinum 8368 processor on Coyote Pass with 256 GB (16 slots/16GB/3200) total DDR4 memory, 8192 GB (16 slots/512 GB/3200) total PMem, ucode x261, HT on, Turbo on, CentOS 8.3.2011, 4.18.0-193.el8.x86\_64, 1x Intel 960GB SSD, 7x P5510 3.84TB, 2x Intel E810-C 100Gb/s, Aerospike Enterprise Edition 5.5.0.2; Aerospike C Client 5.1.0 Benchmark Tool; 70R/30W. Dataset size: 1.1TB, 9.3 billion 64B records, PMDK libPMem, Index (PMem)+data (SSD) and Index+data (PMem), test by Intel on 3/16/2021. Platinum 8280: 1-node, 2x Intel® Xeon® Platinum 8280L processor on Wolf Pass with 768 GB (12 slots/64GB/2666) total PMem, ucode 0x5003003, HT on, Turbo on, CentOS 8.3.2011, 4.18.0-193.el8.x86\_64, 7x P4510 1.8TB PCle 3.1, 2x Intel XL710 40Gb/s, Aerospike Enterprise Edition 5.5.0.2; Aerospike C Client 5.1.0 Benchmark Tool; 70R/30W. Dataset size: 1.1TB, 9.3 billion 64B records, PMDK libPMem, Index (PMem)+data (SSD), test by Intel on 3/16/2021.
- 24. 5.63x higher OpenSSL RSA Sign 2048 performance, 1.90x higher OpenSSL ECDSA Sign p256 performance, 4.12x higher OpenSSL ECDHE x25519 performance, 2.73x higher OpenSSL ECDHE p256 performance, 8280M:1-node, 2x Intel(R) Xeon(R) Platinum 8280M CPU on S2600WFT with 384 GB (12 slots/32GB/2933) total DDR4 memory, ucode 0x5003003, HT On, Turbo Off, Ubuntu 20.04.1 LTS, 5.4.0-65-generic, 1x INTEL\_SSDSC2KG01, OpenSSL 1.1.1j, GCC 9.3.0, test by Intel on 3/5/2021. 8380: 1-node, 2x Intel(R) Xeon(R) Platinum 8380 CPU on M50CYP2SB2U with 512 GB (16 slots/32GB/3200) total DDR4 memory, ucode 0xd000270, HT On, Turbo Off, Ubuntu 20.04.1 LTS, 5.4.0-65-generic, 1x INTEL\_SSDSC2KG01, OpenSSL 1.1.1j, GCC 9.3.0, QAT Engine v0.6.4, test by Intel on 3/24/2021. 8380: 1-node, 2x Intel(R) Xeon(R) Platinum 8380 CPU on M50CYP2SB2U with 512 GB (16 slots/32GB/3200) total DDR4 memory, ucode 0xd000270, HT On, Turbo Off, Ubuntu 20.04.1 LTS, 5.4.0-65-generic, 1x INTEL\_SSDSC2KG01, OpenSSL 1.1.1j, GCC 9.3.0, QAT Engine v0.6.5, test by Intel on 3/24/2021.
- 25. 1.44x XGBoost fit, 1.30x XGBoost predict, 1.36x Kmeans fit, 1.44x Kmeans inference, 1.44x Linear Regression fit, 1.60x Linear Regression inference: 8380: 1-node, 2x Intel® Xeon® Platinum 8380 processor on Coyote Pass with 512 GB (16 slots/32GB/3200) total DDR4 memory, ucode 0x261, HT on, Turbo on, Ubuntu 20.04 LTS, 5.4.0-64-generic, 1x Intel® SSDSC2KG960G7, 1x Intel® SSDSC2KG960G7, Python 3.7.9, Sklearn 0.24.1(https://github.com/IntelPython/scikit-learn\_bench), Daal4py 2021.2, XGBoost 1.3.3, test by Intel® ox 5003003, HT on, Turbo on, Ubuntu 20.04 LTS, 5.4.0-65-generic, 1x Intel® SSDSC2BB800G7, Python 3.7.9, Sklearn 0.24.1(https://github.com/IntelPython/scikit-learn\_bench), Daal4py 2021.2, XGBoost 1.3.3, test by Intel® ox 2/5/2021.
- 26. 10x higher batch AI inference performance with Intel-optimized Tensor Flow vs. stock Cascade Lake FP32 configuration8380: 1-node, 2x Intel Xeon Platinum 8380 processor on Coyote Pass with 512 GB (16 slots/32GB/3200) total DDR4 memory, ucode X261, HT on, Turbo on, Ubuntu 20.04 LTS, 5.4.0-65-generic, 1x Intel\_SSDSC2KG96, Intel SSDPE2KX010T8, ResNet-50 v1.5, gcc-9.3.0, oneDNN 1.6.4, BS=128 FP32,INT8, TensorFlow 2.4.1 with Intel optimizations for 3rd Gen Intel Xeon Scalable processor, upstreamed to TensorFlow-2.5 (container-intel/intel-optimized-tensorflow:tf-r2.5-icx-b631821f), Model zoo: https://github.com/IntelAl/models/tree/icx-launch-public/quickstart/, Unoptimized model: TensorFlow-2.4.1, Modelzoo:https://github.com/IntelAl/models-b master, test by Intel on 3/12/2021.8280: 1-node, 2x Intel Xeon Platinum 8280 processor on Wolf Pass with 384 GB (12 slots/32GB/2933) total DDR4 memory, ucode 0x5003003, HT on, Turbo on, Ubuntu 20.04 LTS, 5.4.0-48-generic, 1x Samsung\_SSD\_860, Intel SSDPE2KX040T8, ResNet-50 v1.5, gcc-9.3.0, oneDNN 1.6.4, BS=128 FP32,INT8, Optimized model: TensorFlow 2.4.1 with Intel optimizations for 3rd Gen Intel Xeon Scalable processor, upstreamed to TensorFlow-2.5 (container-intel/intel-optimized-tensorflow:tf-r2.5-icx-b631821f), Model zoo: https://github.com/IntelAl/models/tree/icx-launch-public/quickstart/, Unoptimized model: TensorFlow-2.4.1, Modelzoo:https://github.com/IntelAl/models-b master, test by Intel on 2/17/2021.

- 27. 3.00x higher CloudXPRT Web Microservices with SLA < 1 sec. Ice Lake: 2-socket Intel® Xeon® 8380 (40C/2.3GHz, 270W TDP) on Intel Software Development, HT on, Turbo on, SNC off, 512GB (16x32GB DDR4-3200), ucode x270, Ubuntu 20.04 LTS, 5.8.0-40-generic, CloudXPRT version 1.1, Tested by Intel and results as of February 2021. Milan: 2-socket AMD EPYC 7763 (64C/2.45GHz, 280W cTDP) on GIGABYTE R282-Z92, SMT on, Boost on, Power deterministic mode, NPS=1, 512 GB (16 x32GB DDR4-3200), ucode 0xa001114, Ubuntu 20.04 LTS, 5.8.0-40-generic, CloudXPRT version 1.1. Tested by Intel and results as of March 2021. Intel contributes to the development of benchmarks by participating in, sponsoring, and/or contributing technical support to various benchmarking groups, including the BenchmarkXPRT Development Community administered by Principled Technologies.
- 1.5x higher AI performance with 3rd Gen Intel® Xeon® Scalable processor supporting Intel® DL Boost vs. FP32 AMD EPYC 7763 (64C Milan): (geomean of 20 workloads including logistic regression inference, logistic regression fit, ridge regression inference, ridge regression fit, linear regression inference, linear regression inference, Resnet 50-v1.5 inference, Resnet 50-v1.5 inference, Resnet 50-v1.5 training, BERT Large SOuaD inference, kmeans inference, kmeans fit, brute knninference, SVC inference, SVC fit, dbscan fit, traintestsplit) 8380: 1-node, 2x Intel Xeon Platinum 8380 (40C/2.3GHz, 270W TDP) processor on Intel Software Development Platform with 512 GB (16 slots/32GB/3200) total DDR4 memory, ucode X55260, HT on, Turbo on, Ubuntu 20.04 LTS, 5.4.0-65generic/5.4.0-64-generic, 1x Intel SSDSC2KG96, Intel SSDPE2KX010T8, tested by Intel, and results as of March 2021, 7763: I-node, 2-socket AMD EPYC 7763 (64C/2.45GHz, 280W cTDP) on GIGABYTE R282-Z92 server with 512 GB (16 slots/ 32GB/3200) total DDR4 memory, ucode 0xa001114, SMT on, Boost on, Power deterministic mode, Ubuntu 20.04 LTS, 5.4.0-65-generic, 1x Samsung MZ7LH3T8/INTEL SSDSC2KG019T8, tested by Intel, and results as of March 2021 ResNet50-v1.5 Intel: gcc-9.3.0, oneDNN 1.6.4, BS=128, INT8, TensorFlow 2.4.1 with Intel optimizations for 3rd Gen Intel Xeon Scalable processor, upstreamed to TensorFlow-2.5 (container-intel/intel-optimized-tensorflow:tf-r2.5-icx-b631821f), Model zoo: https://github.com/IntelAI/models/tree/icx-launch-public/quickstart, ResNet50-v1.5 AMD: gcc-9.3.0, oneDNN1.6.4, BS=128, FP32, TensorFlow-2.4.1, Model zoo: https://github.com/IntelAI/models/tree/icx-launch-public/quickstart, ResNet50-v1.5 AMD: gcc-9.3.0, oneDNN1.6.4, BS=128, FP32, TensorFlow-2.4.1, Model zoo: https://github.com/IntelAI/models/tree/icx-launch-public/quickstart, ResNet50-v1.5 AMD: gcc-9.3.0, oneDNN1.6.4, BS=128, FP32, TensorFlow-2.4.1, Model zoo: https://github.com/IntelAI/models/tree/icx-launch-public/quickstart, ResNet50-v1.5 AMD: gcc-9.3.0, oneDNN1.6.4, BS=128, FP32, TensorFlow-2.4.1, Model zoo: https://github.com/IntelAI/models/tree/icx-launch-public/quickstart, ResNet50-v1.5 AMD: gcc-9.3.0, oneDNN1.6.4, BS=128, FP32, TensorFlow-2.4.1, Model zoo: https://github.com/IntelAI/models/tree/icx-launch-public/quickstart, ResNet50-v1.5 AMD: gcc-9.3.0, oneDNN1.6.4, BS=128, FP32, TensorFlow-2.4.1, Model zoo: https://github.com/IntelAI/models/tree/icx-launch-public/quickstart, ResNet50-v1.5 AMD: gcc-9.3.0, oneDNN1.6.4, BS=128, FP32, TensorFlow-2.4.1, Model zoo: https://github.com/IntelAI/models/tree/icx-launch-public/quickstart, ResNet50-v1.5 AMD: gcc-9.3.0, oneDNN1.6.4, BS=128, FP32, TensorFlow-2.4.1, Model zoo: https://github.com/IntelAI/models/tree/icx-launch-public/guickstart, ResNet50-v1.5 AMD: gcc-9.3.0, oneDNN1.6.4, BS=128, FP32, TensorFlow-2.4.1, Model zoo: https://github.com/IntelAI/models/tree/icx-launch-public/guickstart, ResNet50-v1.5 AMD: gcc-9.3.0, oneDNN1.6.4, BS=128, BS=128 public/benchmarks/image recognition/tensorflow/resnet50vl 5 ResNet50-v1.5 Training Intel: gcc-9.3.0, oneDNN1.6.4, BS=256, FP32, TensorFlow 2.4.1 with Intel optimizations for 3rd Gen Intel Xeon Scalable processor, upstreamed to TensorFlow- 2.5 (container-intel/intel-optimized-tensorflow). b631821f), Model zoo: https://github.com/IntelAl/models/tree/icx-launch-public/guickstart, ResNet50-v1.5 Training AMD: gcc-9.3.0, oneDNN 1.6.4, BS=256, FP32, TensorFlow-2.4.1, Model zoo: https://github.com/IntelAl/models/tree/icx-launch-public/guickstart, ResNet50-v1.5 Training AMD: gcc-9.3.0, oneDNN 1.6.4, BS=256, FP32, TensorFlow-2.4.1, Model zoo: https://github.com/IntelAl/models/tree/icx-launch-public/guickstart, ResNet50-v1.5 Training AMD: gcc-9.3.0, oneDNN 1.6.4, BS=256, FP32, TensorFlow-2.4.1, Model zoo: https://github.com/IntelAl/models/tree/icx-launch-public/guickstart, ResNet50-v1.5 Training AMD: gcc-9.3.0, oneDNN 1.6.4, BS=256, FP32, TensorFlow-2.4.1, Model zoo: https://github.com/IntelAl/models/tree/icx-launch-public/guickstart, ResNet50-v1.5 Training AMD: gcc-9.3.0, oneDNN 1.6.4, BS=256, FP32, TensorFlow-2.4.1, Model zoo: https://github.com/IntelAl/models/tree/icx-launch-public/guickstart, ResNet50-v1.5 Training AMD: gcc-9.3.0, oneDNN 1.6.4, BS=256, FP32, TensorFlow-2.4.1, Model zoo: https://github.com/IntelAl/models/tree/icx-launch-public/guickstart, ResNet50-v1.5 Training AMD: gcc-9.3.0, oneDNN 1.6.4, BS=256, FP32, TensorFlow-2.4.1, Model zoo: https://github.com/IntelAl/models/tree/icx-launch-public/guickstart, ResNet50-v1.5 Training AMD: gcc-9.3.0, oneDNN 1.6.4, BS=256, FP32, TensorFlow-2.4.1, Model zoo: https://github.com/IntelAl/models/tree/icx-launch-public/guickstart, ResNet50-v1.5 Training AMD: gcc-9.3.0, oneDNN 1.6.4, BS=256, FP32, TensorFlow-2.4.1, Model zoo: https://github.com/IntelAl/models/tree/icx-launch-public/guickstart, ResNet50-v1.5 Training AMD: gcc-9.3.0, oneDNN 1.6.4, BS=256, FP32, TensorFlow-2.4.1, Model zoo: https://github.com/IntelAl/models/tree/icx-launch-public/guickstart, ResNet50-v1.5 Training AMD: gcc-9.3.0, oneDNN 1.6.4, BS=256, FP32, TensorFlow-2.4.1, Model zoo: https://github.com/IntelAl/models/tree/icx-launch-public/guickstart, ResNet50-v1.5 Training AMD: gcc-9.3.0, oneDNN 1.6.4, BS=256, public/benchmarks/image recognition/tensorflow/resnet50v1 5 SSD-ResNet34 Intel: gcc-9.3.0, oneDNN1.6.4, BS=1, INT8, TensorFlow 2.4.1 with Intel optimizations for 3rd Gen Intel Xeon Scalable processor, upstreamed to TensorFlow-2.5 (container-intel/intel-optimized-tensorflow:tf-r2.5-icx-b63182lf). Model zoo: https://aithub.com/IntelAl/models/tree/icx-launch-public/auickstart/, AMD: SSD-ResNet34, acc-9.3.0, oneDNN 1.6.4, BS=1, FP32, TensorFlow-2.4, Model zoo: https://github.com/IntelAl/models/tree/icx-launch-public/benchmarks/object\_detection/tensorflow/ssd-resnet34BERT-Large SQuAD Intel: gcc-9.3.0, oneDNN1.6.4, BS=1, INT8, TensorFlow 2.4.1 with Intel optimizations for 3rd Gen Intel Xeon Scalable processor, upstreamed to TensorFlow-2.5 (container-intel/intel-optimized-tensorflow:tf-2.5-icx-b63]82[f), Model zoo; https://github.com/IntelA/models/tree/icx-launch-public/quickstart/, AMD; BERT-Large SQuAD, qcc-9.3.0, one DNN 1.6.4, BS=1, FP32, TensorFlow-2.4.1, Model zoo: https://github.com/IntelAl/models/tree/icx-launch-public/benchmarks/language\_modeling/tensorflow/bert\_large\_Python: Python: Py XGBoost 1.3.3: Benchmarks: https://github.com/IntelPython/scikit-learn\_bench
- 1.3x higher AI performance with 3rd Gen Intel® Xeon® Scalable processor supporting Intel® DL Boost vs. NVIDIA A100 GPU: (geomean of 20 workloads including logistic regression inference, logistic regression fit, ridge regression inference, ridge regression fit. linear regression inference, linear regression fit, elastic net inference, XGBoost Fit, XGBoost predict, SSD-ResNet34 inference, Resnet50-v1.5 inference, Resnet50-v1.5 training, BERT Large SOuaDinference, kmeans inference, kmeans brute\_knninference, SVC inference, S ucode X55260, HT on, Turbo on, Ubuntu 20.04 LTS, 5.4.0-65-generic, 1x Intel SSDSC2KG96, Intel SSDPE2KX010T8, tested by Intel, and results as of March 2021. DL Measurements on A100: 1-node, 2-socket AMD EPYC 7742 (64C) with 256GB (8 slots/32GB/3200) total DDR4 memory, ucode 0x8301038, HT on, Turbo on, Ubuntu 20.04 LTS, 5.4.0-42-generic, INTEL SSDSC2KB01, NVIDIA A100-PCIe-40GB, HBM2-40GB, Accelerator per node = 1, tested by Intel, and results as of March 2021. ML Measurements on A100: 1-node, 2-socket AMD EPYC 7742 (64C) with 512 GB (16 slots/32GB/3200) total DDR4 memory, ucode 0x8301034, HT on, Turbo on, Ubuntu 18.04.5 LTS, 5.4.0-42-generic, NVIDIA A100 (DGX-1), 1.92TB M.2 NVMe, 1.92TB M.2 NVMe RAID tested by Intel, and results as of March 2021. ResNet50-v1.5 Intel: acc-9.3.0, oneDNN 1.6.4. BS=1, INT8, TensorFlow 2.4.1 with Intel optimizations for 3rd Gen Intel Xeon Scalable processor, upstreamed to TensorFlow-2.5 (container-intel/intel-optimized-tensorflow:tf-r2.5-icx-b631821f), Model zoo: https://github.com/IntelAI/models/tree/icx-launch-public/guickstart ResNet50-v1.5 NVIDIA: A100 (7 instance/GPU), BS=1,TensorFlow-1.5.5 (NGC: tensorflow:21.02-tfl-py3), https://github.com/NVIDIA/DeepLearningExamples/tree/master/TensorFlow/Classification/ConvNets/resnet50v1.5, TF AMP (FP16+TF32); ResNet50-v1.5 Training Intel: qcc-9.3.0, oneDNN 1.6.4, BS=256, FP32, TensorFlow 2.4.1 with Intel optimizations for 3rd Gen Intel Xeon Scalable processor, upstreamed to TensorFlow-2.5 (container-intel/intel-optimized-tensorflow;tf-r2.5-icx-b631821f). Model zoo: https://aithub.com/IntelAl/models/tree/icx-launch-public/quickstart. ResNet50-v1.5 Training NVIDIA: A100, BS=256, TensorFlow-1.5.5 (NGC: tensorflow:21.02-tfl-py3), https://github.com/NVIDIA/DeepLearningExamples/tree/master/TensorFlow/Classification/ConvNets/resnet50v1.5, TF32; BERT-Large SOuAD Intel: acc-9.3.0, oneDNN1.6.4, BS=1, INT8, TensorFlow 2.4.1 with Intel optimizations for 3rd Gen Intel Xeon Scalable processor, upstreamed to TensorFlow-2.5 (container-intel/intel-optimized-tensorflow:tf-r2.5-icx-b631821f). Model zoo: https://github.com/IntelAl/models/tree/icx-launch-public/guickstart/A100: BERT-LargeSQuAD, BS=1, A100 (7 instance/GPU), TensorFlow - 1.5.5 (NGC: tensorflow: 20.11-tfl-py3), https://github.com/NVIDIA/DeepLearningExamples/tree/master/TensorFlow/LanguageModeling/BERT,TFAMP (FPI6+TF32); SSD-ResNet34 Intel: gcc-9.3.0, oneDNN 1.6.4, BS=1, INT8, TensorFlow 2.4.1 with Intel optimizations for 3rd Gen Intel Xeon Scalable processor, upstreamed to TensorFlow-2.5 (container-intel/intel-optimized-tensorflow:tf-r2.5-icx-b631821f), Model zoo: https://github.com/IntelAI/models/tree/icx-launch-public/guickstart/, SSD-ResNet34 NVIDIA: A100 (7 instance/GPU), BS=1,Pytorch - 1,8,0a0 (NGC Container, latest supported); A100; SSD-ResNet34 (NGC; pytorch; 20,11-by3), https://github.com/NVIDIA/DeepLearningExamples/tree/master/PyTorch/Detection/SSD, AMP (FP16+TF32); Python: Intel: Python 3.7.9, Scikit-Learn: Sklearn 0.24.1, One DAL: Daal4py 2021.2, XGBoost: XGBoost: XGBoost: NVIDIA A100: Python 3.7.9, Scikit-Learn: Sklearn 0.24.1, CuML 0.17, XGBoost 1.3.0 dev.rapidsai0.17, Nvidia RAPIDS: RAPIDS 0.17, CUDA Toolkit: CUDA 11.0.221 Benchmarks: https://github.com/IntelPython/scikit-learn\_bench

- 30. LINPACK Platinum 8380: 1-node, 2x Intel® Xeon® Platinum 8380 (40C/2.3GHz, 270W TDP) processor on Intel Software Development Platform with 256 GB (16 slots/16GB/3200) total DDR4 memory, ucode 0x261, HT on, Turbo on, CentOS Linux 8.3.2011, 4.18.0-240.1.1.el8\_3.crt1.x86\_64, 1x Intel\_SSDSC2KG96, App Version: The Intel Distribution for LINPACK Benchmark; Build notes: Tools: Intel MPI 2019u7; threads/core: 1; Turbo: used; Build: build script from Intel Distribution for LINPACK package; 1 rank per NUMA node: 1 rank per socket, tested by Intel and results as of March 2021 EPYC 7763: 1-node, 2-socket AMD EPYC 7763 (64C/2.45GHz, 280W cTDP) on GIGABYTE R282-Z92 server with 512 GB (16 slots/32GB/3200) total DDR4 memory, ucode 0xa001114, SMT on, Boost on, Power deterministic mode, NPS=4, Red Hat Enterprise Linux 8.3, 4.18, 1x Samsung\_MZ7LH3T8, App Version: AMD official HPL 2.3 MT version with BLIS 2.1; Build notes: Tools: hpc-x 2.7.0; threads/core: 1; Turbo: used; Build: pre-built binary (gcc built) from https://developer.amd.com/amd-aocl/blas-library/; 1 rank per L3 cache, 4 threads per rank, tested by Intel and results as of March 2021
- 31. Monte Carlo FSI Kernel. Platinum 8380: 1-node, 2x Intel® Xeon® Platinum 8380 (40C/2.3GHz, 270W TDP) processor on Intel Software Development Platform with 256 GB (16 slots/16GB/3200) total DDR4 memory, ucode 0x261, HT on, Turbo on, CentOS Linux 8.3.2011, 4.18.0-240.1.1.el8\_3.crt1.x86\_64, 1x Intel\_SSDSC2KG96, App Version: v1.1; Build notes: Tools: Intel MKL 2020u4, Intel C Compiler 2020u4, Intel Threading Building Blocks 2020u4; threads/core: 1; Turbo: used; Build knobs: -03-xCORE-AVX-512-qopt-zmm-usage=high-fimf-precision=low-fimf-domain-exclusion=31-no-prec-div-no-prec-sqrt tested by Intel and results as of March 2021 EPYC 7763: 1-node, 2-socket AMD EPYC 7763 (64C/2.45GHz, 280W cTDP) on GIGABYTE R282-Z92 server with 512 GB (16 slots/32GB/3200) total DDR4 memory, ucode 0xa001114, SMT on, Boost on, Power deterministic mode, NPS=4, Red Hat Enterprise Linux 8.3, 4.18, 1x Samsung\_MZ7LH3T8, App Version: v1.1; Build notes: Tools: Intel MKL 2020u4, Intel C Compiler 2020u4, Intel Threading Building Blocks 2020u4; threads/core: 2; Turbo: used; Build knobs: -03-march=core-avx2-fimf-precision=low-fimf-domain-exclusion=31-no-prec-div-no-prec-sqrt tested by Intel and results as of March 2021
- 32. NAMD Geomean of Apoal, STMV. Platinum 8380: I-node, 2x Intel® Xeon® Platinum 8380 (40C/2.3GHz, 270W TDP) processor on Intel Software Development Platform with 256 GB (16 slots/16GB/3200) total DDR4 memory, ucode 0x261, HT on, Turbo on, CentOS Linux 8.3.2011, 4.18.0-240.11.el8\_3.crt1.x86\_64, 1x Intel\_SSDSC2KG96, App Version: 2.15-Alphal (includes AVX tiles algorithm); Build notes: Tools: Intel MKL, Intel C Compiler 2020u4, Intel MPI 2019u8, Intel Threading Building Blocks 2020u4; threads/core: 2; Turbo: used; Build knobs: ip-fp-model fast=2-no-prec-div-qoverride-limits-qopenmp-simd-O3-xCORE-AVX-512-qopt-zmm-usage=high, tested by Intel and results as of March 2021 EPYC 7763: I-node, 2-socket AMD EPYC 7763 (64C/2.45GHz, 280W cTDP) on GIGABYTE R282-Z92 server with 512 GB (16 slots/32GB/3200) total DDR4 memory, ucode 0xa001114, SMT on, Boost on, Power deterministic mode, NPS=4, Red Hat Enterprise Linux 8.3, 4.18, 1x Samsung\_MZ7LH3T8, App Version: 2.15-Alphal (includes AVX tiles algorithm); Build notes: Tools: Intel MKL, AOCC 2.2.0, gcc 9.3.0, Intel MPI 2019u8; threads/core: 2; Turbo: used; Build knobs: -O3-fomit-frame-pointer-march=znverl-ffast-math, tested by Intel and results as of March 2021
- 33. RELION Plasmodium Ribosome Platinum 8380: 1-node, 2x Intel® Xeon® Platinum 8380 (40C/2.3GHz, 270W TDP) processor on Intel Software Development Platform with 256 GB (16 slots/16GB/3200) total DDR4 memory, ucode 0x261, HT on, Turbo on, CentOS Linux 8.3.2011, 4.18.0-240.1.1.el8\_3.crt1.x86\_64, 1x Intel\_SSDSC2KG96, App Version: 3\_1\_1; Build notes: Tools: Intel C Compiler 2020u4, Intel MPI 2019u9; threads/core: 2; Turbo: used; Build knobs: -03 -ip -g -debug inline-debug-info-xCOMMON-AVX-512 -qopt-report=5 -restrict, tested by Intel and results as of March 2021 EPYC 7763: 1-node, 2-socket AMD EPYC 7763 (64C/2.45GHz, 280W cTDP) on GIGABYTE R282-Z92 server with 512 GB (16 slots/32GB/3200) total DDR4 memory, ucode 0xa001114, SMT on, Boost on, Power deterministic mode, NPS=4, Red Hat Enterprise Linux 8.3, 4.18, 1x Samsung\_MZ7LH3T8, App Version: 3\_1\_1; Build notes: Tools: Intel C Compiler 2020u4, Intel MPI 2019u9; threads/core: 2; Turbo: used; Build knobs: -03 -ip -g -debug inline-debug-info-march=core-avx2-gopt-report=5 -restrict tested by Intel and results as of March 2021
- 34. 3.88x higher INT8 real-time inference throughput & 22.09x higher INT8 batch inference throughput on ResNet-50 with 3rd Gen Intel® Xeon® Scalable processor supporting Intel® DL Boost vs. FP32 AMD EPYC Milan 8380: 1-node, 2x Intel Xeon Platinum 8380 (40C/2.3GHz, 270W TDP) processor on Intel Software Development Platform with 512 GB (16 slots/32GB/3200) total DDR4 memory, ucode X55260, HT on, Turbo on, Ubuntu 20.04 LTS, 5.4.0-65-generic, 1x Intel\_SSDSC2KG96, Intel SSDPE2KX010T8, ResNet50-v1.5, gcc-9.3.0, oneDNN 1.6.4, BS=1,128, INT8, TensorFlow 2.4.1 with Intel optimizations for 3rd Gen Intel Xeon Scalable processor, upstreamed to TensorFlow 2.5 (container-intel/intel-optimized-tensorflow:tf-r2.5-icx-b631821f), Model zoo: https://github.com/IntelAl/models/tree/icx-launch-public/quickstart/, tested by Intel, and results as of March 2021.

  7763: 1-node, 2-socket AMD EPYC 7763 (64C/2.45GHz, 280W cTDP) on GIGABYTE R282-Z92 server with 512 GB (16 slots/32GB/3200) total DDR4 memory, ucode 0xa001114, SMT on, Boost on, Power deterministic mode, NPS=1, Ubuntu 20.04 LTS, 5.4.0-65-generic, 1x Samsung\_MZ7LH3T8, ResNet50-v1.5, gcc-9.3.0, oneDNN 1.6.4, BS=1,128, FP32, TensorFlow-2.4.1, Model: https://github.com/IntelAl/models/tree/icx-launch-public/benchmarks/image\_recognition/tensorflow/resnet50v1\_5, tested by Intel, and results as of March 2021.
- 35. 2.79x higher INT8 real-time inference throughput & 12x higher INT8 batch inference throughput on SSD-MobileNet-v1 with 3rd Gen Intel® Xeon® Scalable processor supporting Intel® DL Boost vs. FP32 AMD EPYC Milan 8380: 1-node, 2x Intel Xeon Platinum 8380 (40C/2,3GHz, 270W TDP) processor on Intel Software Development Platform with 512 GB (16 slots/32GB/3200) total DDR4 memory, ucode X55260, HT on, Turbo on, Ubuntu 20.04 LTS, 5.4.0-65-generic, Ix Intel\_SSDSC2KG96, Intel SSDPE2KX010T8, SSD-MobileNet-v1, gcc-9.3.0, oneDNN 1.6.4, BS=1,448, INT8, TensorFlow 2.4.1 with Intel optimizations for 3rd Gen Intel Xeon Scalable processor, upstreamed to TensorFlow 2.5 (container-intel/intel-optimized-tensorflow:tf-r2.5-icx-b631821f), Model zoo: https://github.com/IntelAl/models/tree/icx-launch-public/quickstart/, tested by Intel, and results as of March 2021. 7763: 1-node, 2-socket AMD EPYC 7763 (64C/2.45GHz, 280W cTDP) on GIGABYTE R282-Z92 server with 512 GB (16 slots/32GB/3200) total DDR4 memory, ucode 0xa001114, SMT on, Boost on, Power deterministic mode, NPS=1, Ubuntu 20.04 LTS, 5.4.0-65-generic, Ix Samsung\_MZ7LH3T8, SSD-MobileNet-v1, gcc-9.3.0, oneDNN 1.6.4, BS=1,448, FP32, TensorFlow-2.4.1, Model zoo: https://github.com/IntelAl/models/tree/icx-launch-public/benchmarks/object\_detection/tensorflow/ssd-mobilenet.tested by Intel, and results as of March 2021.

- 36. Upto 25x higher Al performance with 3rd Gen Intel® Xeon® Scalable processor supporting Intel® DL Boost vs. FP32 AMD EPYC 7763 (64C Milan), 4.0lx higher INT8 real-time inference throughput & 25.05x higher INT8 batch inference throughput on MobileNet-v1 with 3rd Gen Intel® Xeon® Scalable processor supporting Intel® DL Boost vs. FP32 AMD EPYC Milan: 1-node, 2x Intel Xeon Platinum 8380 processor on Coyote Pass with 512 GB (16 slots/32GB/3200) total DDR4 memory, ucode X55260, HT on, Turbo on, Ubuntu 20.04 LTS, 5.4.0-65-generic, 1x Intel\_SSDSC2KG96, Intel SSDPE2KX010T8, MobileNet-v1, gcc-9.3.0, oneDNN1.6.4, BS=1,56, INT8, TensorFlow 2.4.1 with Intel optimizations for 3rd Gen Intel Xeon Scalable processor, upstreamed to TensorFlow-2.5 (container-intel/intel-optimized-tensorflow:tf-r2.5-icx-b631821f), Model zoo: https://github.com/IntelAl/models/tree/icx-launch-public/quickstart/, test by Intel on March 2021. 1-node, 2x AMD Epyc 7763 on GigaByte with 512 GB (16 slots/32GB/3200) total DDR4 memory, ucode 0xa001114, HT on, Turbo on, Ubuntu 20.04 LTS, 5.4.0-65-generic, 1x Samsung\_MZ7LH3T8, MobileNet-v1, gcc-9.3.0, oneDNN1.6.4, BS=1,56, FP32, TensorFlow-2.4.1, Model zoo: https://github.com/IntelAl/models/tree/icx-launch-public/benchmarks/image\_recognition/tensorflow/mobilenet\_v1, tested by Intel and results as of March 2021.
- 3.18x higher INT8 real-time inference throughput & 2.17x higher INT8 batch inference throughput on BERT Large SQuAD with 3rd Gen Intel® Xeon® Scalable processor supporting Intel® DL Boost vs. FP32 AMD EPYC Milan 8380:1-node, 2x Intel Xeon Platinum 8380 (40C/2.3GHz, 270W TDP) processor on Intel Software Development Platform with 512 GB (16 slots/32GB/3200) total DDR4 memory, ucode X55260, HT on, Turbo on, Ubuntu 20.04 LTS, 5.4.0-65-generic, 1x Intel\_SSDSC2KG96, Intel SSDPE2KX010T8, BERT Large SQuAD, gcc-9.3.0, oneDNN1.6.4, BS=1,128, INT8, TensorFlow 2.4.1 with Intel optimizations for 3rd Gen Intel Xeon Scalable processor, upstreamed to TensorFlow-2.5 (container-intel/intel-optimized-tensorflow:tf-r2.5-icx-b631821f), Model zoo: https://github.com/IntelAl/models/tree/icx-launch-public/quickstart/, tested by Intel, and results as of March 2021.

  7763: I-node, 2-socket AMD EPYC 7763 (64C/2.45GHz, 280W cTDP) on GIGABYTE R282-Z92 server with 512 GB (16 slots/32GB/3200) total DDR4 memory, ucode 0xa001114, SMT on, Boost on, Power deterministic mode, NPS=1, Ubuntu 20.04 LTS, 5.4.0-65-generic, 1x Samsung\_MZ7LH3T8, BERT Large SQuAD, gcc-9.3.0, oneDNN1.6.4, BS=1,128, FP32, TensorFlow-2.4.1, Model zoo: https://github.com/IntelAl/models/tree/icx-launch-public/benchmarks/language modeling/tensorflow/bert large, tested by Intel, and results as of March 2021.
- 38. 3.20x higher OpenSSL RSA Sign 2048 performance, 2.03x higher OpenSSL ECDHE x25519 performance 8380: 1-node, 2x Intel(R) Xeon(R) Platinum 8380 CPU on M50CYP2SB2U with 512 GB (16 slots/32GB/3200) total DDR4 memory, ucode 0xd0000270, HT On, Turbo Off, Ubuntu 20.04.1 LTS, 5.4.0-65-generic, 1x INTEL\_SSDSC2KG01, OpenSSL 1.1.1j, GCC 9.3.0, QAT Engine v0.6.4, Tested by Intel and results as of March 2021. 7763: 1-node, 2x AMD EPYC 7763:64-Core Processor on R282-Z92-00 with 512 GB (16 slots/32GB/3200) total DDR4 memory, ucode 0xa001114, HT On, Turbo Off, Ubuntu 20.04.1 LTS, 5.4.0-65-generic, 1x SAMSUNG\_MZ7LH3T8, OpenSSL 1.1.1j, GCC 9.3.0, Tested by Intel and results as of March 2021.
- 39. 1) 5G vRAN: Results have been estimated or simulated: Based on 2x throughput from 32Tx32R (5Gbps) on 2nd Gen Intel® Xeon® Gold 6212U processor to 64Tx64R (10Gbps) on 3rd Gen Intel® Xeon® Gold 6338N processor at ~185W
  - 2) VDI: https://www.principledtechnologies.com/VMware/VMware-HCI-Intel-Optane-VDI-0420.pdf
  - 3) Azure stack HCI Ice Lake Configuration:
  - 4 Node, 2x Intel® Xeon® Gold 6330 CPU, Ix Intel® Server Board M50CYP, Total Memory: 256GB (16 x 16 GB 3200MHz DDR4 RDIMM, HyperThreading: Enable, Turbo: Enable, Storage (boot): 1 x Intel® SSD D3-S4510 Series (480GB, 2.5in SATA 6Gb/s, 3D2, TLC), Storage: 4x Intel® SSD DC P4610 Series (3.2TB) (NVMe), Network devices: 1 x 100 GbE Intel(R) Ethernet Network Adapter E810-C-Q2, Network speed: 25 GbE, 1 x 10 GbE Intel(R) Ethernet Converged Network Adapter X550-T2, Network Speed: 10 GbE, OS/Software: Microsoft Azure Stack HCl build 17763, Benchmarks: DiskSpd (QD=8,30w:70r): 1.396M IOPS @4.03ms(r), @6.95ms(w) for 90% requests, Tested by Intel as of 12-Mar-2021.

Azure stack HCI – Cascade Lake Configuration:

4 Node, 2x Intel® Xeon® Gold 6230, 1x Intel® Server Board S2600WFT, Total Memory: 512 GB Intel® Optane™ DC persistent Memory, 4 slots/128 GB/2666 MT/s and 192 GB, 12 slots/16 GB/2666 MT/s, HyperThreading: Enable, Turbo: Enable, Storage (boot): 1x 480 GB Intel® SSD 3520 Series M.2 SATA, Storage (cache): 2x 375 GB Intel® Optane™ DC SSD P4800X, Storage (capacity): 4x 4 TB Intel® SSD DC P4510 PCIe NVMe, Network devices: 1x 25 Gbps Chelsio\* Network Adapter, Network speed: 25GbE, OS/Software: Windows Server\* 2019 Datacenter Edition build 17763, Benchmarks: DiskSpd (QD=8,30w:70r): 588K IOPS @4.99ms(r), @19.54ms(w) for 90% requests, Tested by Intel® of 22-Feb-2019

- 45. Why Customers Choose Intel Delivering workload-optimized performance:
  - a. Kingsoft Cloud: https://www.intel.com/content/www/us/en/customer-spotlight/stories/kingsoft-cloud-cdn-customer-story.html
  - b. VK: https://www.intel.com/content/www/us/en/customer-spotlight/stories/vk-storage-customer-story.html
  - c. CERN: https://www.nextplatform.com/2021/02/01/cern-uses-dlboost-oneapi-to-juice-inference-without-accuracy-loss/
  - d. Naver. https://www.intel.com/content/www/us/en/customer-spotlight/stories/naver-ocr-customer-story.html
  - e. Datto: https://www.intel.com/content/www/us/en/customer-spotlight/stories/datto-customer-story.html
  - f. BIH: https://www.intel.com/content/www/us/en/customer-spotlight/stories/berlin-institute-health-customer-story.html
- 46. P5800X IOPS Test and System Configs & Specifications
  - a. World's Fastest Data Center SSD: Intel. As compared to generally available PCle Gen4 x4 (4 lanes) Enterprise and Data Center industry SSDs. Results may vary.
  - b. Alibaba, Intel does not control or audit third-party data. You should consult other sources to evaluate accuracy. Results may vary.
  - c. Excelero, December 16, 2020. https://www.excelero.com/blog/a-breakthrough-technology-helps-ai-ml-and-database-storage/. Intel does not control or audit third-party data. You should consult other sources to evaluate accuracy. Results may vary.
  - d. P5800X gains vs Intel® SSD D7-P5600 NAND (both Gen4 PCle):Intel. Date tested March 18, 2021. Workload FIO rev 3.5, based on random 512B transfer size with total queue depth of 64 (QD=8, workers/jobs=8) workload, 4KB transfer size with total queue depth of 32 (QD=4, workers/jobs=8) workload, 8KB transfer size with total queue depth of 16 (QD=4, workers/jobs=4) workload in most case, except where specified. LATENCY System configuration: Intel Optane SSD P5800X: 1.6TB: CPU: Intel® Xeon® Platinum 8380 2.30GHz 270W 40 cores per socket, CPU Sockets: 2, BIOS: SE5C6200.86B.3021.D40.2103160200, UCODE: 0X8D05A260, RAM: 32GB @3200 MT/s DDR4, DIMM Slots Populated: 16 slots, PCle Attach: CPU (not PCH lane attach), OS: Ubuntu 20.04.2 LTS, Kernel: 5.4.0-67-generic, FIO version: 3.16; NVMe Driver: Inbox, C-states: Disabled, Hyper Threading: Disabled, CPU Governor (through OS): Performance Mode. Intel® Version: 3/18/2021 vs Intel® SSD D7-P5600: see https://www.intel.com/content/www/us/en/products/docs/memory-storage/solid-state-drives/data-center-ssds/d7-p5600-p5500-series-brief.html
    QoS, IOPS/GB system configuration: Intel Optane SSD P5800X: CPU: Intel® Xeon® Gold 6254 3.10GHz 30MB 160W 18 cores per socket, CPU Sockets: 2, BIOS: SE5C620.86B.02.01.0009.092820190230, RAM Capacity: 32G, RAM Model: DDR4, RAM Stuffing: NA, DIMM Slots Populated: 4 slots, PCIe Attach: CPU (not PCH lane attach), Chipset: Intel C610 chipset, Switch/ReTimer Model/Vendor: Intel G4SAC switch (PCIe Gen4), OS: Cent0S7.5.1804, Kernel: 4.14.74, FIO version: 3.5; NVMe Driver: Inbox, C-states: Disabled, Hyper Threading: Disabled, CPU Governor (through OS): Performance Mode. Measurements are performed on a full Logical Block Address (LBA) span of the drive. Power mode set at PM0. Test by Intel Nov 2020. vs Intel® SSD D7-P5600: see https://www.intel.com/content/www/us/en/products/docs/memory-storage/solid-state-drives/data-center-ssds/d7-p5600-p5500-series-brief.html
  - e. CEPH: Intel® Optane™ SSD DC P4800X + Intel® SSD P4510: Tested by Intel on 2/20/2019, 5-nodes, 2x Intel® Xeon Gold 6252 on WolfPass with 12 x 16GB 2666MHz DDR4 (total 384GB), NIC: 25x2 GbE Mellanox Conect-4 Lx CX4121A, Storage: Intel® SSD DC S3610 1.5TB, Application drive: 1x Intel® Optane™ SSD DC P4800X (375GB) + 6x Intel® SSD DC P4510 (4TB), Bios: SE5C620.86B.0D.01.0250.112320180145, ucode: 0x4000010 (HT=ON, Turbo=ON), OS: RedHat 7.6, Kernel: 3.10.0-957.el7.x86\_64, Benchmark: Ceph 13.2.4 Mimic, QD= 64, Results: 4KB read = 1313300 IOPS & Latency (99.99th Percentile)= 320.13 ms, 4KB write = 291006.67 IOPS & (99.99th Percentile)= 499.68 ms, 4KB read/write (70/30) = 656517.67 IOPS & (99.99th Percentile)= 519.43 ms. Intel® SSD DC P4510 test results captured on 4TB model, while cost calculations are based on 8TB model pricing. Drive capacity is not material to test results in this benchmarking scenario. Baseline: Tested by Intel on 2/20/2019, 5-nodes, 2x Intel® Xeon Gold 6152 on WolfPass with 12 x 16GB 2666MHz DDR4 (total 384GB), NIC: 25x2 GbE Mellanox Conect-4 Lx CX4121A, Storage: Intel® SSD DC 94500 (4TB), Bios: SE5C620.86B.0D.01.0250.112320180145, ucode: 0x4000010 (HT=ON, Turbo=ON), OS: RedHat 7.6, Kernel: 3.10.0-957.el7.x86\_64, Benchmark: Ceph 13.2.4 Mimic, QD= 64, Results: 4KB read = 1149766.67 IOPS & (99.99th Percentile)= 581.58 ms. 4KB write = 230116.67 IOPS & (99.99th Percentile)= 556.35 ms, 4KB read/write (70/30) = 536652.33 IOPS & (99.99th Percentile)= 574.75 ms. Intel. Test Date Feb 20, 2019. For further test configuration and test setup documentation refer to the Ceph Benchmarking Best-Known Methods (BKMs): Installation Guide @ https://www.intel.com/content/www/us/en/patner/cloud-insider/content-library.html?grouping=rdc%20/Content%20/Conten
  - f. Vmware: Intel does not control or audit third-party data. You should consult other sources to evaluate accuracy. Results may vary. <a href="https://www.evaluatorgroup.com/document/lab-insight-latest-intel-technologies-power-new-performance-levels-vmware-vsan-2018-update/">https://www.evaluatorgroup.com/document/lab-insight-latest-intel-technologies-power-new-performance-levels-vmware-vsan-2018-update/</a>
  - g. DELL EMC: Intel. https://www.intel.com/content/dam/www/public/us/en/documents/solution-briefs/dell-emc-powermax-Optane-ssd-brief.pdf. Results may vary.

- 47. Intel® SSD D5-P5316 Massive storage capacity made flexible
  - a. Up to 7GB/s higher sequential read 128K sequential read bandwidth between Intel® SSD D5-P5316 15.36TB (7.0 GB/s)
  - b. Up to 48% better latency performance at 99.999%: Source-Intel product specification. Comparing measured performance for 4KB Random Read, QDI latency performance at 99.999% between Intel® SSD D5-P5316 15.36TB with Intel® SSD D5-P4326 respectively.
  - c. Up to 5x higher endurance gen over gen Comparing endurance (64K random write) between Intel® SSD D5-P5316 30.72TB (22,930 TBW) and Intel® SSD D5-P4326 15.36TB (4,400 TBW).
  - d. Up to 20x reduction of warm storage footprint. With 4TB HDD drive, it takes 10 (2U) of rack space to fill up 1PB or storage. With Intel® SSD D5-P5316 30.72TB E1.L or U.2, it takes 1U of rack space to fill up 1PB of storage.
- 48. Intel® Agilex™ FPGA + Quartus Prime 20.4 Software FPGA performance made flexible
  - a. ~2x Better Fabric Performance per Watt vs. Versal: The Agilex and Versal devices (part number/speed grade) used in the perf/watt comparison are as follows: AGF014-2, Versal: Equivalent density to AGF014-2 in 2M speed grade, ted March 2021 by Intel. Design profile used for the comparison: Base Stratix 10 frequency: 450MHz, Agilex Fmax = 450 \* 1.59 = 716Mhz, Versal Fmax = 450 \* 1.19 = 536Mhz, Resource usage: 60% of AGF014 resource (logic, M20K memory, DSP), power at the respective Fmax; Version of all the tools used for this data: Agilex Quartus 20.4/PTC 21.1 b149, Versal: Vivado 2020.2/XPE: 2020.2
  - a) 50% faster Video IP performance: Derived from a set of five video IP designs comparing Fmax of each design achieved in Xilinx Versal ACAP devices with the Fmax achieved in Intel® Agilex™ devices, using Intel® Quartus® Prime Software (version 20.4) and Xilinx Vivado Software (version 2020.2). On geomean, designs running in the mid speed grade of Intel® Agilex™ FPGAs achieve a 50% higher in Fmax compared to the same designs running in the mid speed grade of Xilinx Versal devices (-2M speed grade), and 42% higher in Fmax compared to the same designs running in the mid speed grade of Xilinx Versal devices (-2H speed grade) and 24% higher in Fmax compared to the same designs running in the mid speed grade of Xilinx Ionm VUP devices (-2 speed grade), tested January 2021.
  - b) Up to 49% faster fabric performance compared to prior generation FPGA for high-speed 5G fronthaul gateway applications—Derived from comparing the Fmax result of Agilex FPGA and Stratix 10 FPGA in a fronthaul gateway reference example using Quartus Prime 20.4 software, tested in February, 2021.
  - c) Software Configurations: Tests were done by running internal builds of Intel® Quartus® Prime Pro Design Software on a wide variety of internal benchmarks. The computer systems used for the evaluations were Intel® Skylake CPU @ 3.3GHz 256G Memory class machines running SUSE Linux Enterprise Server 12 operating system. The performance results represent average improvements across a wide variety of internal benchmarks, and results may vary for each testcase. Performance results are based on testing as of dates shown in configurations and may not reflect all publicly available updates.
- 50. Cache and memory latency comparisons
  - a) Cascade Lake: 2-socket Intel® Xeon® 8280 (2.7GHz, 28C, 205W TDP) on Intel Software Development Platform, HT on, Turbo on, SNC off, prefetchers disabled, 384 GB (12x32GB DDR4-2933), ucode 0x5000024, Ubuntu 19.04, 5.3.0-rc3-custom, kernel patched for security mitigations, Latencies using Intel Memory Latency Checker version 3.7, <a href="https://software.intel.com/en-us/articles/intelr-memory-latency-checker">https://software.intel.com/en-us/articles/intelr-memory-latency-checker</a>. Tested by Intel as of and results as of October 2019.
  - b) Ice Lake: 2-socket Intel® Xeon® 8380 (40C/2.3GHz, 270W TDP) on Intel Software Development, HT on, Turbo on, SNC off, prefetchers disabled, 512GB (16x32GB DDR4-3200), ucode 0xd0001e0, Ubuntu 20.04 LTS, 5.4.0-65-generic, Latencies using Intel Memory Latency Checker version 3.9, <a href="https://software.intel.com/en-us/articles/intel-memory-latency-checker">https://software.intel.com/en-us/articles/intel-memory-latency-checker</a>. Tested by Intel and results as of March 2021.
  - c) Milan: 2-socket AMD EPYC 7763 (64C/2.45GHz, 280WcTDP) on GIGABYTE R282-Z92, SMT on, Boost on, Power deterministic mode, NPS=1, prefetchers disabled, 512 GB (16 x32GB DDR4-3200), ucode 0xa001114, Ubuntu 20.04 LTS, 5.4.0-65-generic, Latencies using Intel Memory Latency Checker version 3.9, <a href="https://software.intel.com/en-us/articles/intelr-memory-latency-checker">https://software.intel.com/en-us/articles/intelr-memory-latency-checker</a>. Tested by Intel and results as of March 2021
- 4.2x NGINX (TLS 1.2 Handshake) web server connections/sec with ECDHE-X25519-RSA2K Multi-buffer: 6338N: I-node, 2x Intel® Xeon® Gold 6338N processor on Coyote Pass with 256 GB (16 slots/16GB/ 2666) total DDR4 memory, ucode x261, HT on, Turbo off, Ubuntu 20.04.1 LTS, 5.4.0-65-generic, x 3 x Quad Ethernet Controller E810-C for SFP 25 GBE, Async NGINX v0.4.3, OpenSSL 1.1.1h, QAT Engine v0.6.4, Crypto MB-ippcp\_2020u3, GCC 9.3.0, GLIBC 2.31, test by Intel on 3/22/2021. 6252N: I-node, 2x Intel® Xeon® Gold 6252N processor on Supermicro X11DPG-QT with 192 GB (12 slots/16GB/ 2933) total DDR4 memory, ucode 0x5003003, HT on, Turbo off, Ubuntu 20.04.1 LTS, 5.4.0-65-generic, x 2 x Quad Ethernet Controller XXV710 for 25GbE SFP28, Async NGINX v0.4.3, OpenSSL 1.1.1h, GCC 9.3.0, GLIBC 2.31, test by Intel on 1/17/2021.
- 52. 50% higher performance on MySQL, Redis, and Nginx: Alibaba 7th Gen ECS Cloud Server claim, https://developer.alivun.com/article/781850?spm=a2c6h.17735062.0.0.5bca49f2vpLXaT