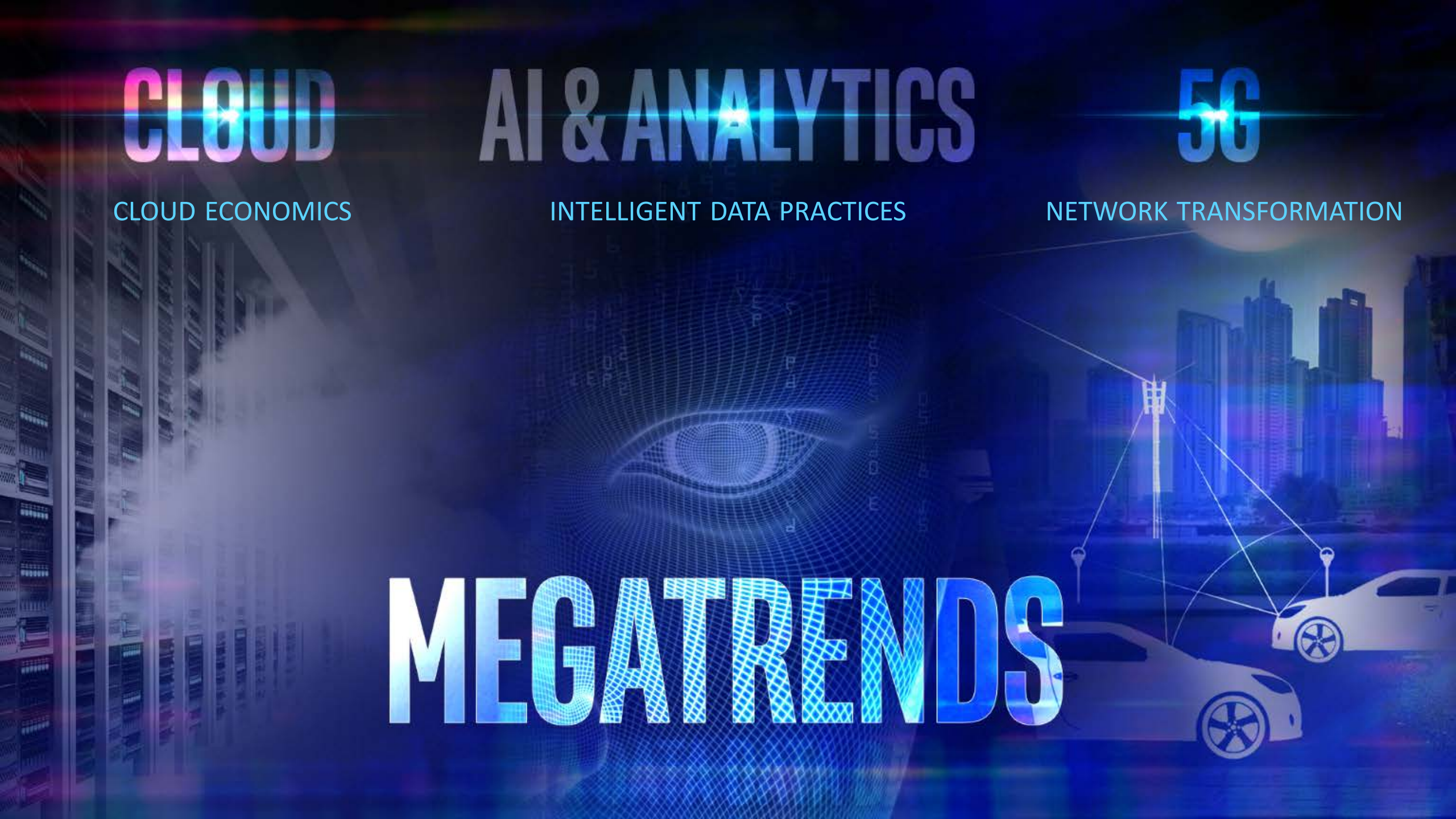
Artur Dlugosz

Server Sales Manager Intel Corporation


CLOUD
AI & ANALYTICS
5G
CLOUD ECONOMICS
INTELLIGENT DATA PRACTICES
NETWORK TRANSFORMATION
MEGATRENDS

TRANSFORMATION STARTS
ON THE INSIDE

Intel® Xeon® Scalable Platform

Biggest platform advancement

The industry's Biggest platform advancement In a decade

INTEL® XEON® SCALABLE Platform

Performance

1.65X

AVERAGE GENERATION AL

Security

2X

DATA PROTECTION PERFORMANCE

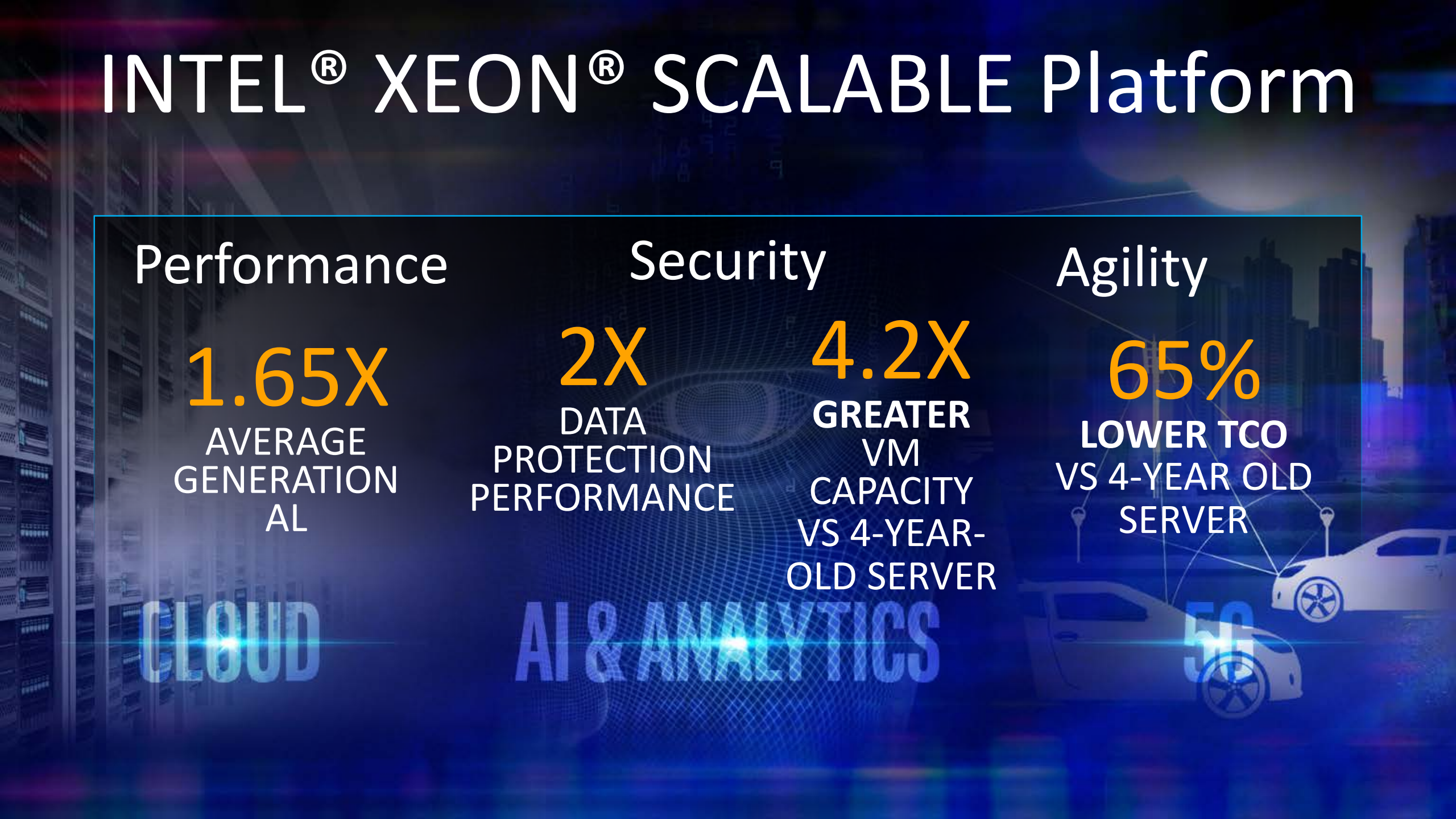
4.2X

GREATER VM CAPACITY VS 4-YEAR-OLD SERVER

Agility

65%

LOWER TCO VS 4-YEAR OLD SERVER



Performance

Intel® AVX-512

NEW LEVEL OF VECTOR PERFORMANCE

UP TO

2X

FLOPS PER CLOCK CYCLE[1]

Intel® Mesh Architecture

OPTIMIZED

CORES, CACHE, MEMORY, I/O

Intel® QuickAssist Technology

UP TO

100 GB/S

ENCRYPTION
@1K PACKETS

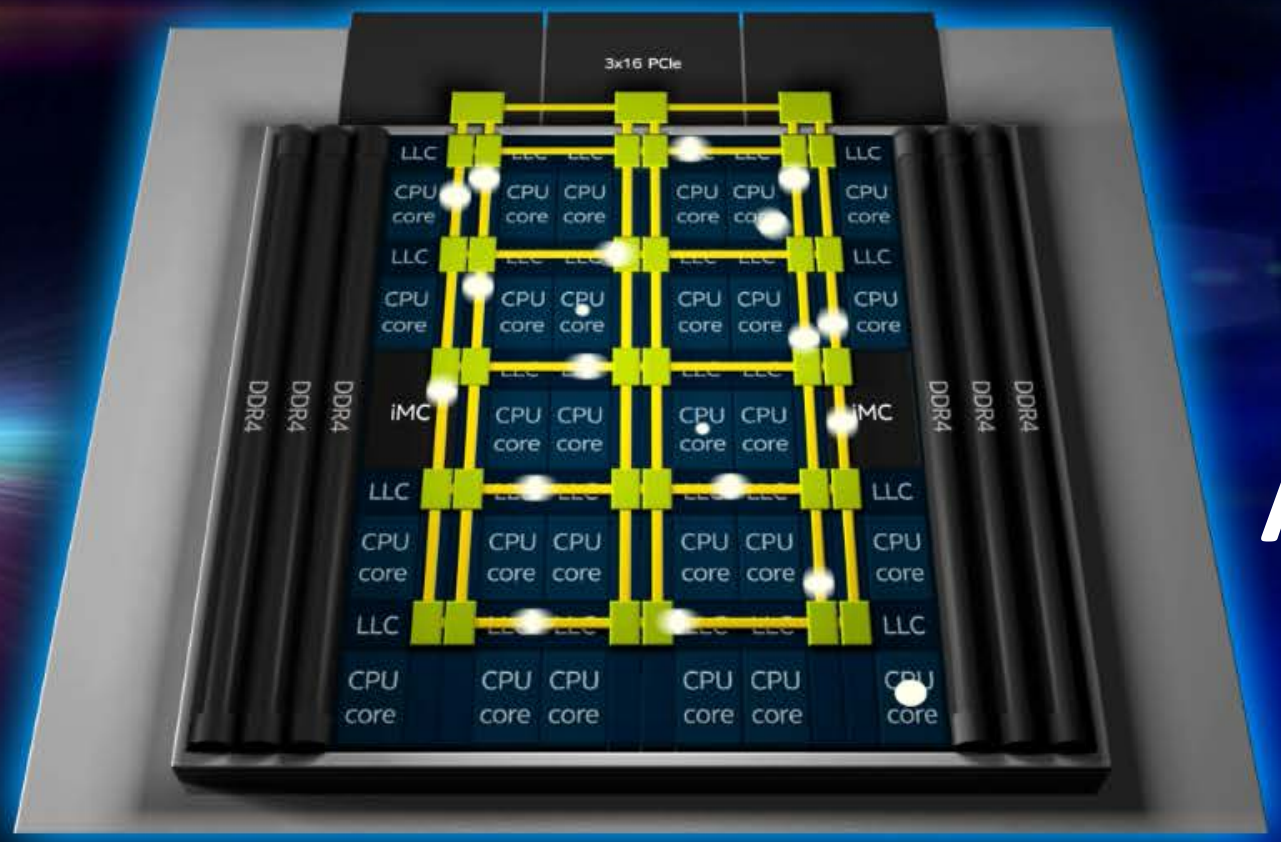
UP TO

100 GB/S

COMPRESSION
@8K BLOCK

Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products. For more complete information visit http://www.intel.com/performance.

1. Up to 2X flops per cycle: when compared with Intel® AVX2 available on previous E5/E7 v3, v4 families



Intel® Mesh Architecture

The Intel Xeon Scalable:

Intel® Xeon® BRONZE

31XX Series Bronze (2 Socket)
- Up to **8 cores**
- 2S configuration
- **Improved core interconnect (UPI)** over past gen
- **48 PCIe 3.0 lanes**
- **Intel® AVX-512** feature enabled
- **Standard RAS** features

Entry Performance and security for price sensitive deployments

Intel® Xeon® SILVER

41XX Series Silver (2 Socket)
+ Up to **12 cores**
+ 2S configuration with **Improved Memory channel performance**
+ **Intel® Turbo Boost Technology** for higher frequency capability
+ **Intel® HT Technology** for hyper threaded workloads

Efficient Performance at Low Power. Provides more horsepower for single purpose workloads

Intel® Xeon® GOLD

61XX Series Gold (2 and 4 Socket)
+ Up to **22 Cores**
+ **Added 3rd UPI link** for increased dataflow across cores
+ Increased performance across memory channels[6]
+ Intel® **AVX-512 with additional FMA**
+ Added **Node Controller Support** to assist in scaled node management

51XX Series Gold (2 and 4 Socket)
+ Up to **14 cores**
+ Supports **2S and 4S configuration** for increased scalability
+ Increased core interconnect speed to boost data flow in multi-processor workloads
+ **Advanced RAS** features

Mainstream Performance, Fast Memory, More Interconnect Engines, Advanced Reliability

Intel® Xeon® PLATINUM

81XX Series Platinum (2, 4, and 8 Socket)
+ Up to **28 Cores**
+ **2,4, or 8 socket configurations** for best performance and scalability[5]
+ **Topline memory channel performance** (1.5 TB memory bandwidth on select SKUs)
+ **3 UPI links option across 2S,4S,8S** for improved scalability and inter-core data flow

The Best Performance, Scalability, Core options, and all Hardware-Enhanced Security features for the most robust capability

Better performance, interconnectivity, scalability, and memory

Intel® Xeon® Scalable Processor

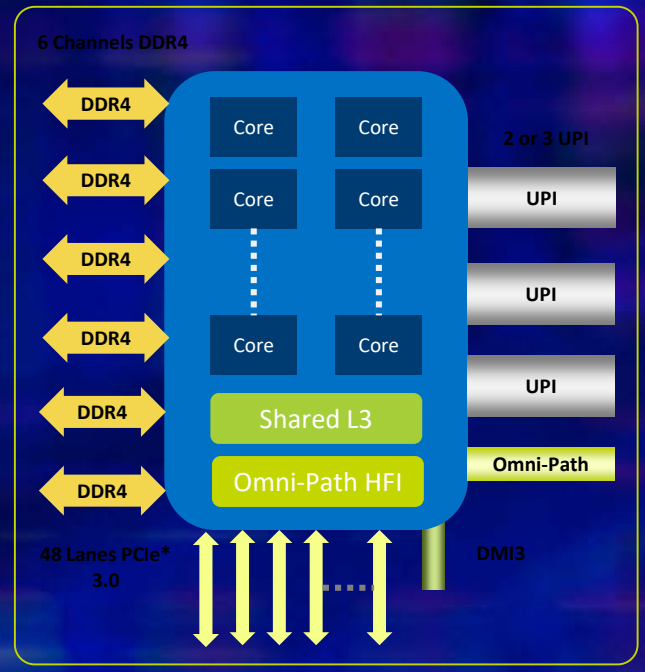
Re-architected from the Ground Up

- Skylake core microarchitecture, with data center specific enhancements
- Intel® AVX-512 with 32 DP flops per core
- Data center optimized cache hierarchy – 1MB L2 per core, non-inclusive L3

- New mesh interconnect architecture
- Enhanced memory subsystem
- Modular IO with integrated devices
- New Intel® Ultra Path Interconnect (Intel® UPI)

- Intel® Speed Shift Technology
- Security & Virtualization enhancements (MBE, PPK, MPX)
- Optional Integrated Intel® Omni-Path Fabric (Intel® OPA)

Features	Intel® Xeon® Processor E5-2600 v4	Intel® Xeon® Scalable Processor
Cores Per Socket	Up to 22	**Up to 28**
Threads Per Socket	Up to 44 threads	**Up to 56 threads**
Last-level Cache (LLC)	Up to 55 MB	**Up to 38.5 MB (non-inclusive)**
QPI/UPI Speed (GT/s)	2x QPI channels @ 9.6 GT/s	**Up to 3x UPI @ 10.4 GT/s**
PCIe* Lanes/ Controllers/Speed(GT/s)	40 / 10 / PCIe* 3.0 (2.5, 5, 8 GT/s)	**48 / 12 / PCIe 3.0 (2.5, 5, 8 GT/s)**
Memory Population	4 channels of up to 3 RDIMMs, LRDIMMs, or 3DS LRDIMMs	**6 channels of up to 2 RDIMMs, LRDIMMs, or 3DS LRDIMMs**
Max Memory Speed	Up to 2400	**Up to 2666**
TDP (W)	55W-145W	**70W-205W**



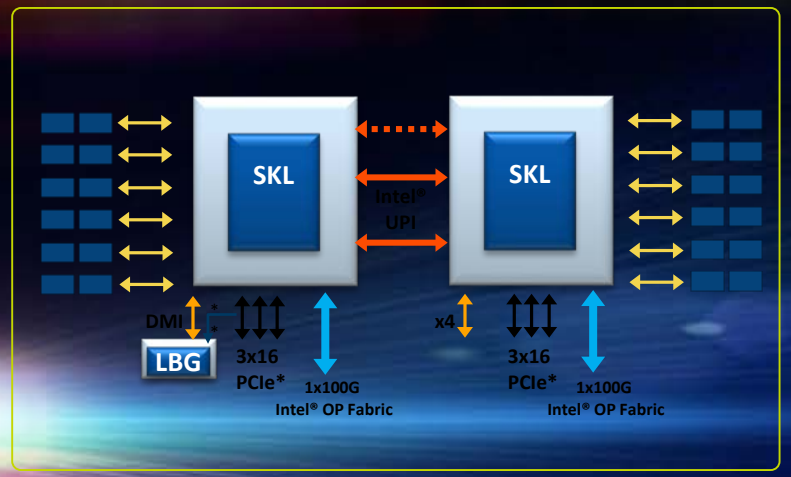
Re-Architected L2 & L3 Cache Hierarchy

- On-chip cache balance shifted from shared-distributed (prior architectures) to private-local (Skylake architecture):
  - Shared-distributed ➔ shared-distributed L3 is primary cache
  - Private-local ➔ private L2 becomes primary cache with shared L3 used as overflow cache
- Shared L3 changed from inclusive to non-inclusive:
  - Inclusive (prior architectures) ➔ L3 has copies of all lines in L2
  - Non-inclusive (Skylake architecture) ➔ lines in L2 ***may not*** exist in L3

Skylake-SP cache hierarchy architected specifically for Data center use case

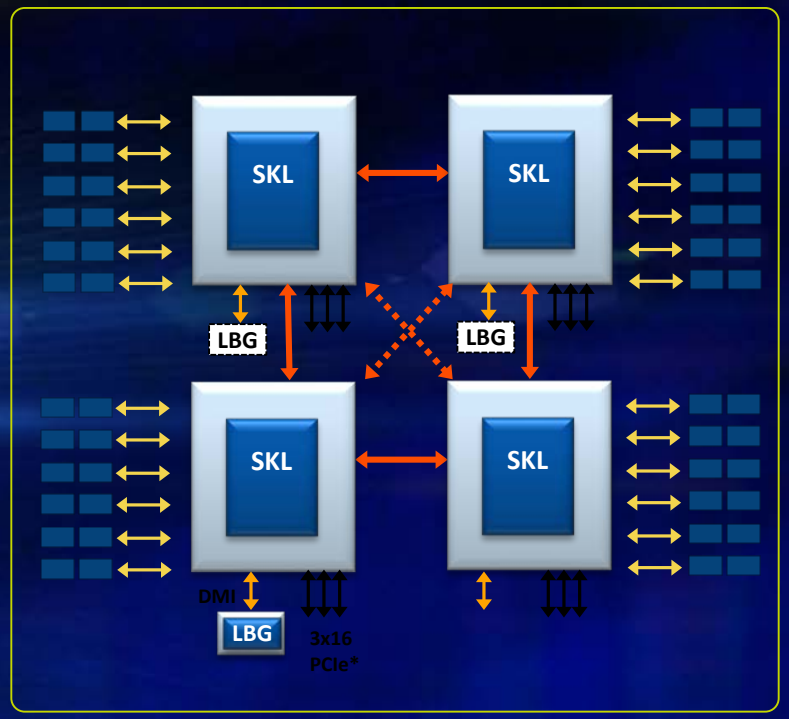
Platform Topologies

2S Configurations



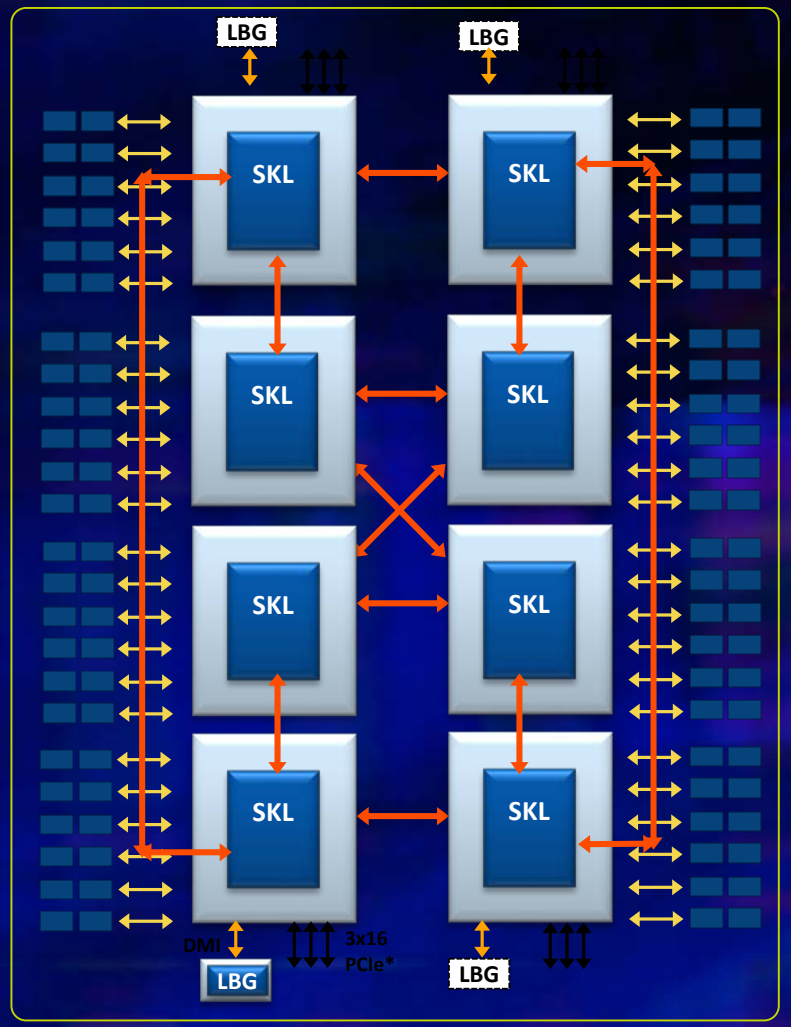
(2S-2UPI & 2S-3UPI shown)

4S Configurations



(4S-2UPI & 4S-3UPI shown)

8S Configuration



Intel® Xeon® Scalable Processor supports configurations ranging from 2S-2UPI to 8S

Most Agile, Scalable AI Platform

Built-in ROI

Potent Performance

Production Ready

cut training time from **days** to **hours**

up to

113X

performance with optimized software

VS INTEL XEON E5 V3[1]

Real-time workloads, like **inferencing**, run on Xeon

up to

2.2X

deep learning performance

VS PRIOR GEN[2]

Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products. For more complete information visit http://www.intel.com/performance.

1. Platform: 2S Intel® Xeon® Platinum 8180 CPU @ 2.50GHz (28 cores), HT disabled, turbo disabled, scaling governor set to "performance" via intel_pstate driver, 384GB DDR4-2666 ECC RAM. CentOS Linux release 7.3.1611 (Core), Linux kernel 3.10.0-514.10.2.el7.x86_64. SSD: Intel® SSD DC S3700 Series (800GB, 2.5in SATA 6Gb/s, 25nm, MLC).Performance measured with: Environment variables: KMP_AFFINITY='granularity=fine, compact', OMP_NUM_THREADS=56, CPU Freq set with cpupower frequency-set -d 2.5G -u 3.8G -g performance. Compared with Platform: 2S Intel® Xeon® CPU E5-2699 v3 @ 2.30GHz (18 cores), HT enabled, turbo disabled, scaling governor set to "performance" via intel_pstate driver, 256GB DDR4-2133 ECC RAM. CentOS Linux release 7.3.1611 (Core), Linux kernel 3.10.0-514.el7.x86_64. OS drive: Seagate* Enterprise ST2000NX0253 2 TB 2.5" Internal Hard Drive.Performance measured with: Environment variables: KMP_AFFINITY='granularity=fine, compact,1,0', OMP_NUM_THREADS=36, CPU Freq set with cpupower frequency-set -d 2.3G -u 2.3G -g performance. Intel Caffe: (http://github.com/intel/caffe/), revision b0ef3236528a2c7d2988f249d347d5fdae831236. Inference measured with "caffe time --forward_only" command, training measured with "caffe time" command. For "ConvNet" topologies, dummy dataset was used. For other topologies, data was stored on local storage and cached in memory before training. Topology specs from https://github.com/intel/caffe/tree/master/models/intel_optimized_models (GoogLeNet, AlexNet, and ResNet-50), GCC 4.8.5, MKLML version 2017.0.2.20170110. BVLC-Caffe: https://github.com/BVLC/caffe, Inference & Training measured with "caffe time" command. For "ConvNet" topologies, dummy dataset was used. For other topologies, data was st ored on local storage and cached in memory before training BVLC Caffe (http://github.com/BVLC/caffe), revision 91b09280f5233cafc62954c98ce8bc4c204e7475 (commit date 5/14/2017). BLAS: atlas ver. 3.10.1.
2. Platform: 2S Intel® Xeon® Platinum 8180 CPU @ 2.50GHz (28 cores), HT disabled, turbo disabled, scaling governor set to "performance" via intel_pstate driver, 384GB DDR4-2666 ECC RAM. CentOS Linux release 7.3.1611 (Core), Linux kernel 3.10.0-514.10.2.el7.x86_64. SSD: Intel® SSD DC S3700 Series (800GB, 2.5in SATA 6Gb/s, 25nm, MLC).Performance measured with: Environment variables: KMP_AFFINITY='granularity=fine, compact', OMP_NUM_THREADS=56, CPU Freq set with cpupower frequency-set -d 2.5G -u 3.8G -g performance. Compared with Platform: 2S Intel® Xeon® CPU E5-2699 v4 @ 2.20GHz (22 cores), HT enabled, turbo disabled, scaling governor set to "performance" via acpi-cpufreq driver, 256GB DDR4-2133 ECC RAM. CentOS Linux release 7.3.1611 (Core), Linux kernel 3.10.0-514.10.2.el7.x86_64. SSD: Intel® SSD DC S3500 Series (480GB, 2.5in SATA 6Gb/s, 20nm, MLC). Performance measured with: Environment variables: KMP_AFFINITY='granularity=fine, compact,1,0', OMP_NUM_THREADS=44, CPU Freq set with cpupower frequency-set -d 2.2G -u 2.2G -g performance. Neon: ZP/MKL_CHWN branch commit id:52bd02acb947a2adabb8a227166a7da5d9123b6d. Dummy data was used. The main.py script was used for benchmarking , in mkl mode. ICC version used : 17.0.3 20170404, Intel MKL small libraries version 2018.0.20170425; Inference and training throughput uses FP32 instructions

Intel® C620 Series Chipset

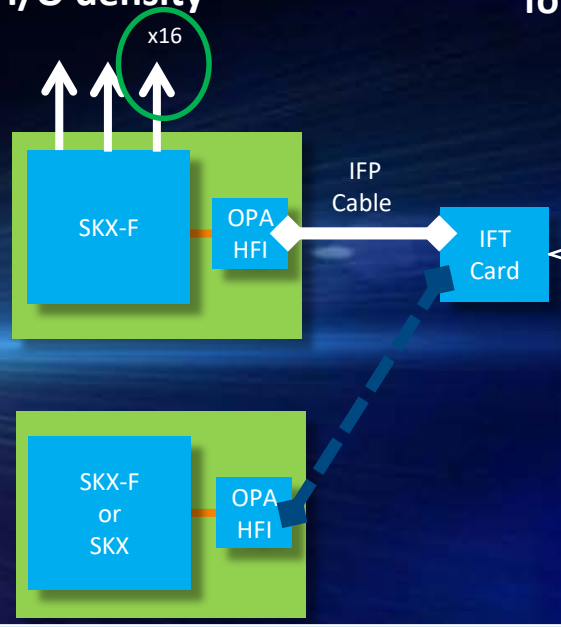
Integrated Intel® QuickAssist Technology

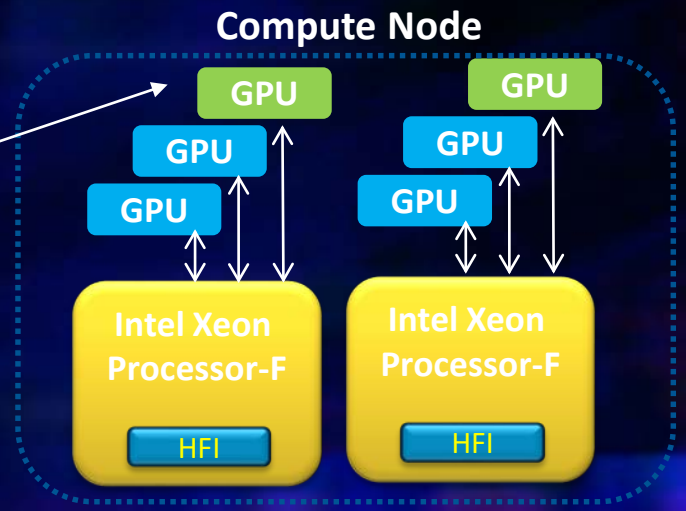
PRODUCT NAME	SKU	10Gb/1Gb ETHERNET PORTS	COMPRESSION	ENCRYPTION	RSA
			Intel® QuickAssist Technology		
Intel® C621 Chipset	LBG-1G	0/4	N/A	N/A	N/A
Intel® C622 Chipset	LBG-2	2/4	N/A	N/A	N/A
Intel® C624 Chipset	LBG-4	4/4	N/A	N/A	N/A
Intel® C625 Chipset	LBG-E	4/4	20 Gb/s	20 Gb/s	20K Ops/s
Intel® C626 Chipset	LBG-M	4/4	40 Gb/s	40 Gb/s	40K Op/s
Intel® C627 Chipset	LBG-T	4/4	100 Gb/s	100 Gb/s	100K Ops/s
Intel® C628 Chipset	LBG-L	4/4	100 Gb/s	100 Gb/s	100K Ops/s

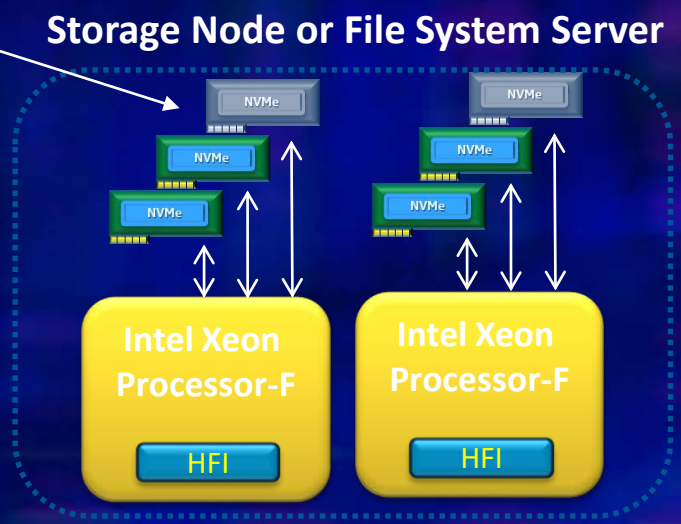
Integrated Intel® Omni-Path Architecture

Up to TWO additional PCIe x16 slots are available for maximizing I/O density[1]

Significantly more I/O capacity for compute or storage nodes[1]







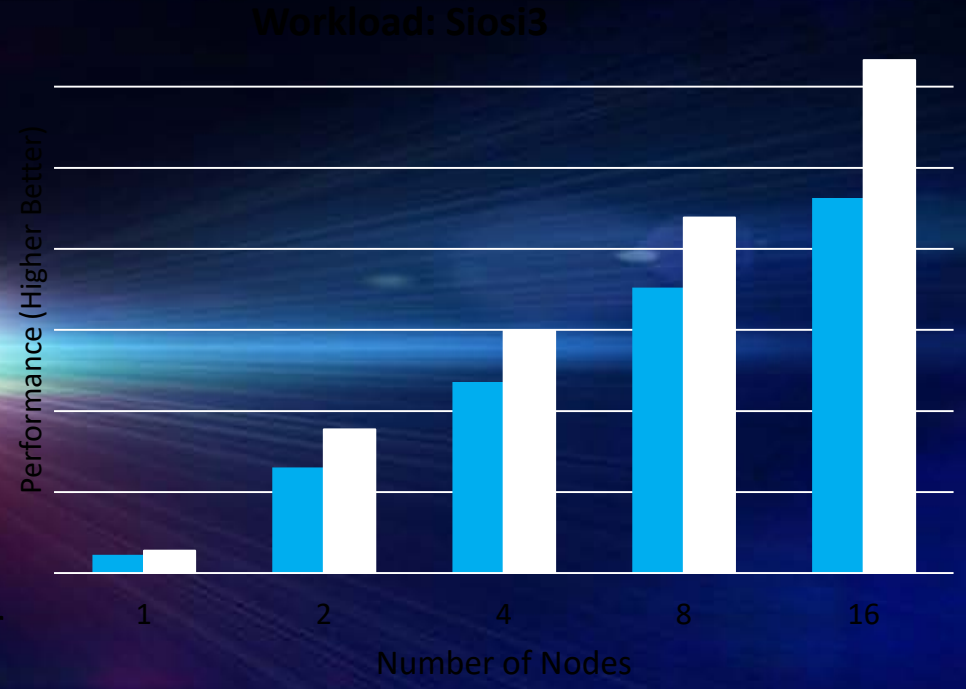
SKUS WITH INTEGRATED INTEL® OMNI-PATH ARCHITECTURE FABRIC

Class	SKU	Cores	Base Non-AVX Speed (GHz)	TDP (W)
Platinum	8176F	28	2.1	173
Platinum	8160F	24	2.1	160
Gold	6148F	20	2.4	160
Gold	6142F	16	2.6	160
Gold	6138F	20	2.0	135
Gold	6130F	16	2.1	135
Gold	6126F	12	2.6	105

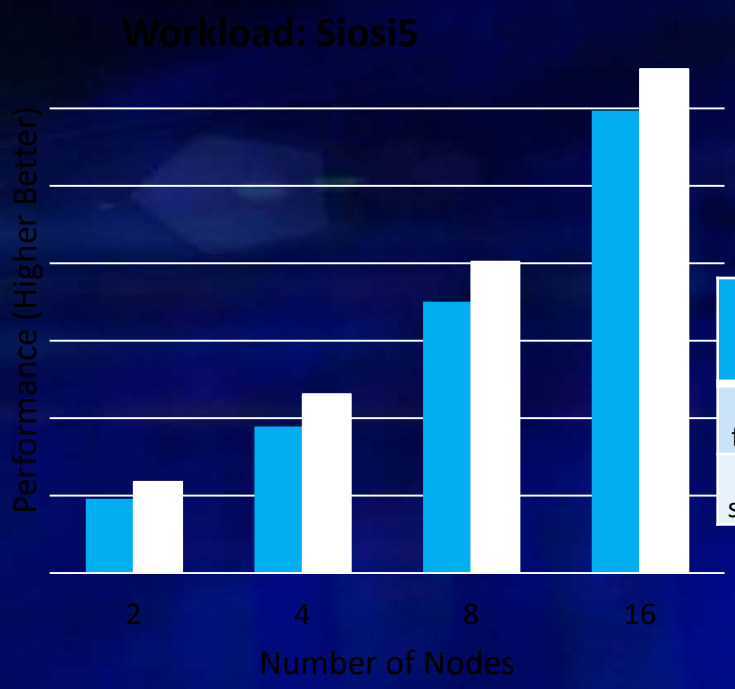
[1] For illustrative purposes only. Assumes each CPU socket is configured with all 48 PCIe lanes routed to three x16 slots, or 96 total lanes for a 2S Purley platform. PCIe slot count and PCIe device support will vary by OEM platform, so check with your OEM for more details.

Intel® Omni-Path Architecture Performance

Intel® Xeon® **Platinum 8170** processor vs Intel® Xeon® processor **E5-2697a v4**



HIGHER is Better



Intel® Xeon® E5-2697a v4 processor[2]

Intel® Xeon® Platinum 8170 processor[1]

	E5-2697A processor	8170 Platinum processor	Xeon Scalable Advantage
Base frequency	2.6	2.1	X
Memory size/speed	64 GB 2133 MHz	192 GB 2666 MHz	✓

1 MPI rank per node

1. Intel® Xeon® Platinum 8170 dual socket servers, 2.10 GHz, 26 cores/socket, 64 GB 2666 MHz DDR4 memory per node. RHEL* 7.3, 3.10.0-514.el7.x86_64 kernel.
2. Intel® Xeon® Processor E5-2697A v4 dual socket servers, 2.10 GHz, 16 cores/node, 64 GB 2133 MHz DDR4 memory per node. RHEL 7.3. BIOS settings: Snoop hold-off timer = 9, Early snoop disabled, Cluster on die disabled. Common configurations: Intel® Turbo Boost Technology enabled, Intel® Hyper-Threading Technology disabled. Intel Fabric Suite 10.3.1.0.22. Intel Corporation Device 24f0 – Series 100 HFI ASIC. OPA Switch: Series 100 Edge Switch – 48 port. OPA parameters: -genv I_MPI_FABRICS shm:tmi -genv I_MPI_TMI_PROVIDER psm2. HFI parameters: krcvqs=4 eager_buffer_size=8388608 max_mtu=10240
NWChem release 6.6. Binary: nwchem_armci-mpi_intel-mpi_mkl with MPI-PR run over MPI-1. Workload: siosi3 and siosi5. Intel MPI 2017.1.132. 2 ranks per node, 1 rank for computation and 1 rank for communication. http://www.nwchem-sw.org/index.php/Main_Page

Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products. For more complete information visit http://www.intel.com/performance. Copyright © 2017, Intel Corporation.*Other names and brands may be claimed as the property of others.

Intel® Quick Assist Technology (Intel® QAT)

Now integrated directly into the Intel C620 series chipset

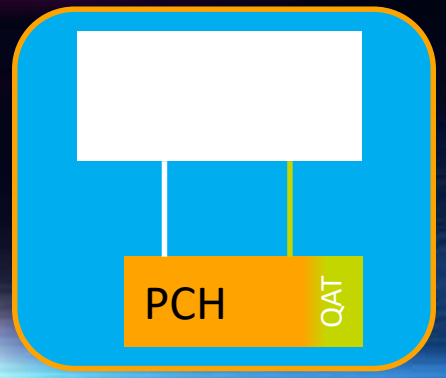
- Set of scalable hardware accelerators exposed to IA. More efficient packet processing, and Functions acceleration. Crypto (public and Pvt keys), and Compression (codecs)
  - Server: secure browsing, email, search, big-data analytics (Hadoop), secure multi-tenancy, IPsec, SSL/TLS, OpenSSL
  - Networking: firewall, IDS/IPS, VPN, secure routing, Web proxy, WAN optimization (IP Comp), 3G/4G authentication
  - Storage: real-time data compression, static data compression, secure storage.
- Intel® QAT on Purley offers outstanding capabilities: 100Gbs Crypto, 100Gbs Compression, 100kops RSA , 2k Decrypt
- Technology genesis in the IXP28xx Network Processors; continuously enhanced since then, in every generation

Symmetric cryptography functions including cipher operations and authentication operations
Public key functions including RSA, Diffie-Hellman, and elliptic curve cryptography
Compression and decompression functions including DEFLATE

http://www.intel.com/content/www/us/en/embedded/technology/quickassist/overview.html

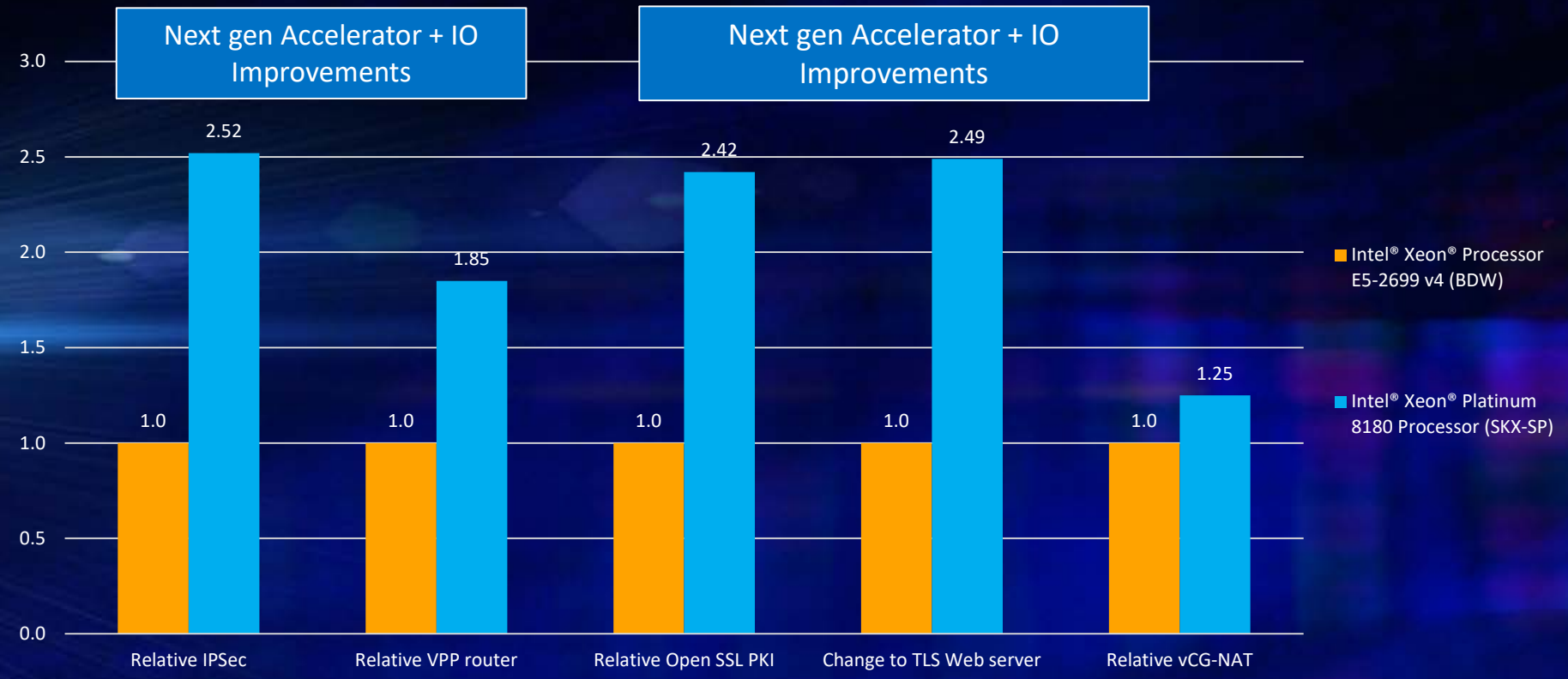
Packet processing & functions acceleration

Converged “highly integrated” Platform for the network



Intel® Xeon® Scalable processor with Server PCH + QAT

- comms
- Cloud
- Enterprise



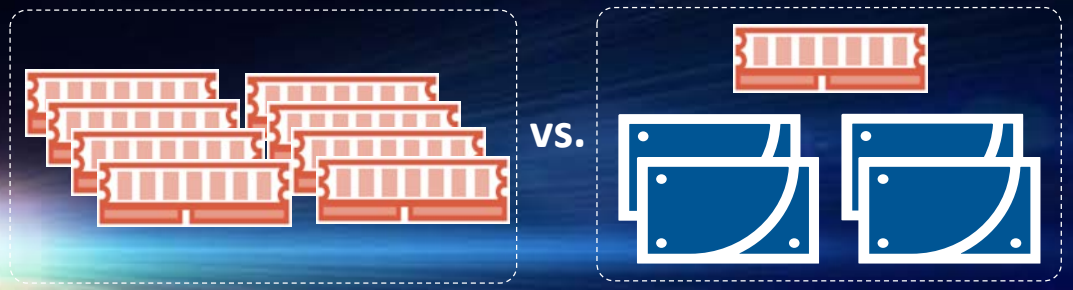
Application & Architectural Level Performance Comparison

Results have been estimated based on internal Intel analysis and are provided for informational purposes only. Any difference in system hardware or software design or configuration may affect actual performance. Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products. For more information go to http://www.intel.com/performance/datacenter.

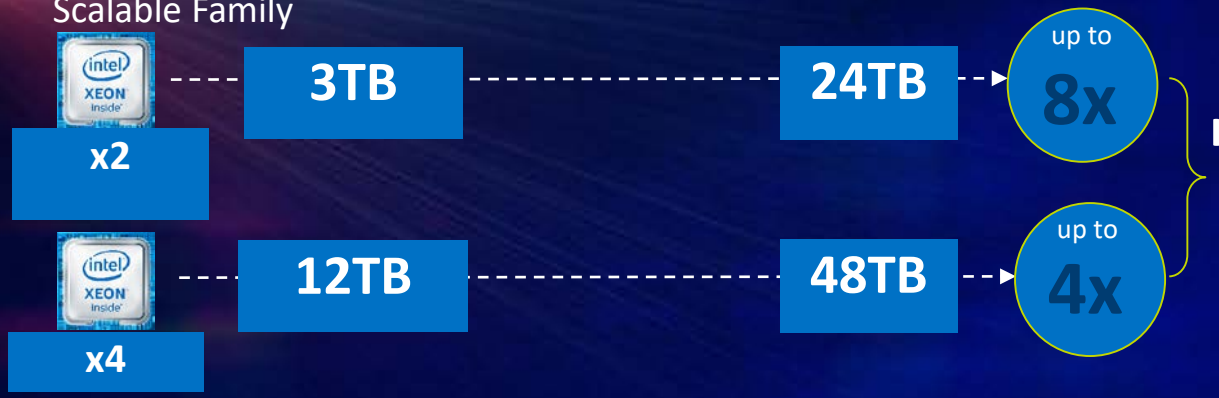
Massively Scalable, Faster[3] Memory

DRAM + Intel® Optane™ SSD + Intel® Memory Drive Technology

All DRAM



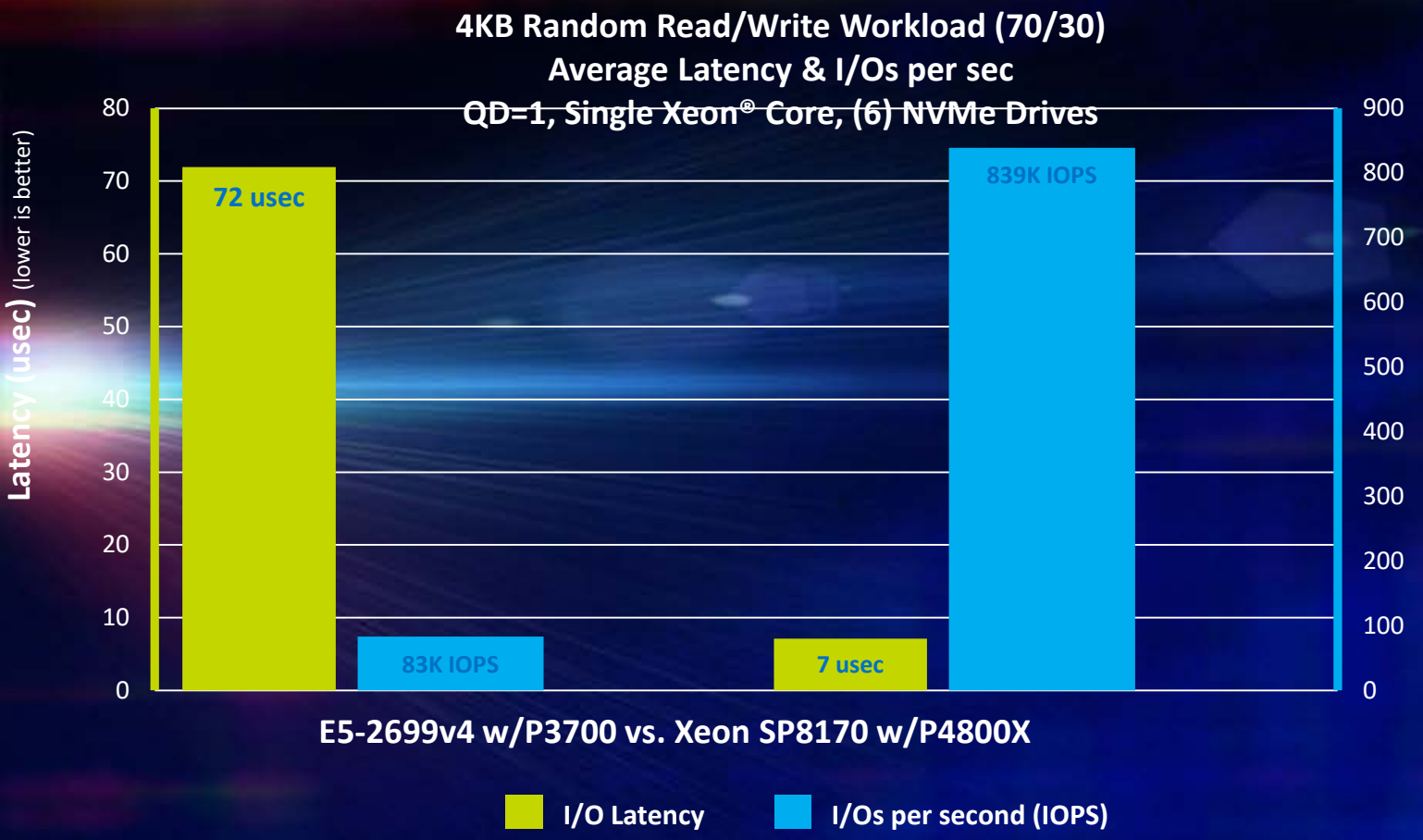
Intel® Xeon® Processor Scalable Family



- Increase memory pool up to **8x**[1]
- Displace DRAM up to **10:1** in select workloads[2]
- **Higher platform** memory & PCIe **bandwidth** with Intel® Scalable Family of Processors[3]
- Accelerate applications and gain new insights from **larger working sets**

Intel® Optane™ SSD DC P4800X for Storage Builders

SPDK Performance: Platform Comparison

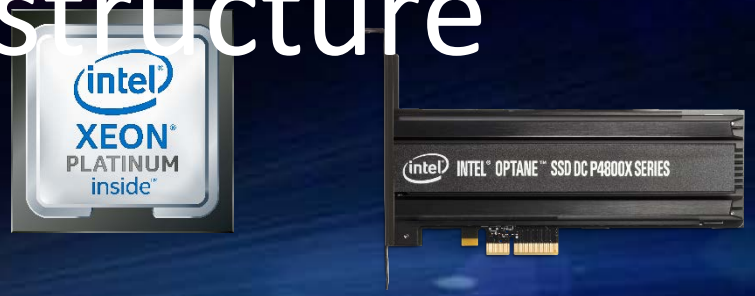


Intel® Xeon® Scalable Processor Platinum Family + Intel® Optane™

- 10X higher throughput
- 10X lower latency
- Up to 27 cores remaining for:
  - Virtual Machines
  - Big Data/Analytics
  - Machine Learning
  - Storage services like erasure coding, de-duplication, compression, or encryption.
- Platform offers RDMA
  - Enables NVMe over Fabrics
  - No more trapped I/O capacity

See notices, configurations, disclaimers

Intel® Xeon® Platinum and Intel® Optane™ SSDs for STORAGE infrastructure



- **Intel Optane SSD P4800 Series:** High performance, low latency storage
- **Intel® Volume Management Device:** Hot-swap of drives with standardized LED management
- **Software tools for optimized storage**
  - Intel® Intelligent Storage Acceleration Library (ISA-L)
  - Intel® Storage Performance Development Kit (SPDK)

2X performance increasgeneration[1]e

Vs prior

Business impact

- ✓ Faster data analytics results
- ✓ More complex analyses
- ✓ Deeper data insights

[1] 2x claim based on SAS Business Analytics: SAS 9.4 m4 application running the 30 session SAS Mixed Analytics workload. OS: CentOS 7.2 kernel 3.10.0. Testing by Intel and SAS May 2017. Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products. For more complete information visit http://www.intel.com/performance. *Other names and brands may be claimed as the property of others.

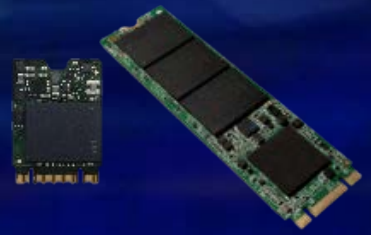
Storage Enhancements for Intel® Xeon® Scalable Platforms

New PCIe storage enhancements in Processor

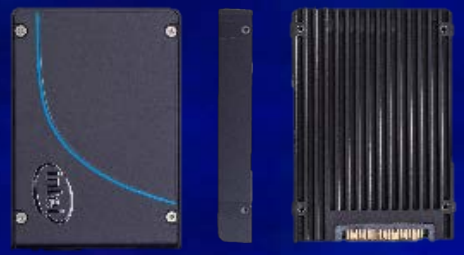
Intel® Volume Management Device (Intel® VMD) provides support for PCIe Solid State Drives (SSD): Hot-plug, enclosure management and error containment functions

New SSDs

M.2 (SATA and PCIe) support – recommend for all server boot drives



U.2 2.5 inch PCIe SSD topologies supported



New RAID and Performance Enhancement Software

PCIe based RAID using Intel® Virtual RAID on Chip (Intel® VROC) technology

Updates to Intel® Rapid Storage Technology Enterprise (Intel® RSTe) and Intel® Cache Acceleration Software (Intel® CAS)

DATA CENTER DESIGNED

Security

Protect the Data

UP TO

2X

DATA PROTECTION PERFORMANCE
GEN OVER GEN[1]

INTEL® KEY PROTECTION TECHNOLOGY

PROTECT KEYS FROM SOFTWARE ATTACKS

Secure the Platform

HARDWARE
ROOT OF TRUST

Security Without Compromise

0.37%

ENCRYPTION PERFORMANCE **OVERHEAD**[2]

DATA CENTER DESIGNED

Agility

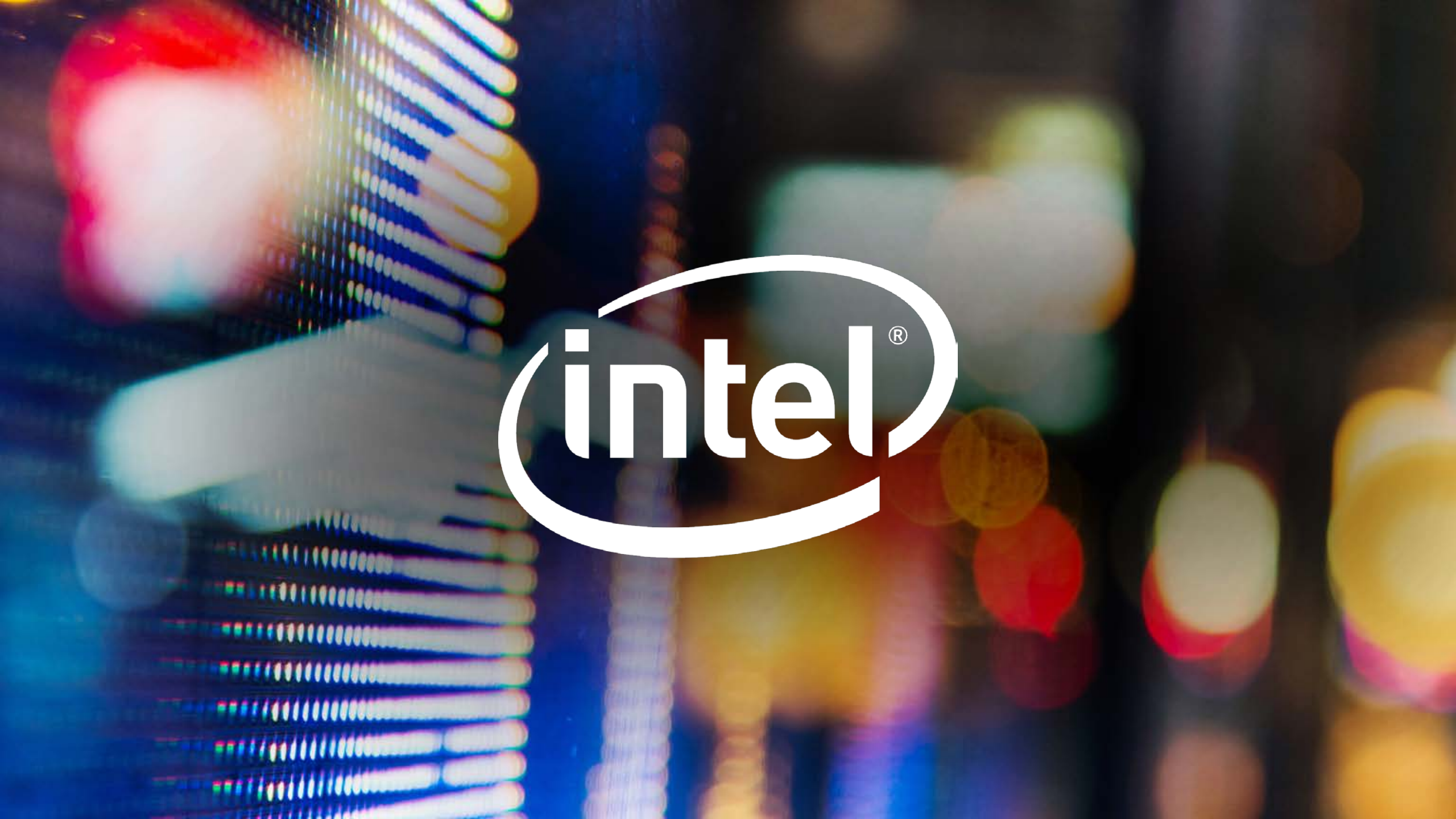
Advanced RAS Features

Intel® Volume Management Device
WITH INTEL® OPTANE™ SSDS

Enhanced Virtualization
WITH MODE BASED EXECUTION

Artificial Intelligence

Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products. For more complete information visit http://www.intel.com/performance.


intel