SOLUTION BRIEF

Intel Corporation

Modular Single-Socket Reference Design for 4th Generation Intel® Xeon® Scalable Processor

This reference design (based on the 4th Gen Intel® Xeon® processor) is a modular and single socket hardware solution for all markets including Enterprise, Cloud, and Network Edge.

Introductions

The single-socket modular reference design is a new, innovative platform architecture created to meet the needs of a broad range of network edge, mainstream enterprise, and cloud markets. The flexible, modular platform can be configured for a wide variety of servers for the 4th Generation Intel® Xeon® Scalable processor, and future generations of processors.

The reference design architecture combines new technology and open standards in a modular, flexible design, optimized for the 4th Gen Intel Xeon processor. The single-socket compute module is designed to support XCC, MCC, and EE SKU variants, addressing the wide variety of markets mentioned earlier.

PCIe* cable connections provide flexibility in I/O or storage configurations. The architecture accommodates choices through standards: Open Compute Project (OCP) NICs 3.0, for growing 3rd party board products, and provisioning for a pluggable BMC module for solution flexibility. In addition, the modular architecture will support the next generation Intel® Xeon®. Hence, engineering investment is greatly reduced while ROI increases because an ODM or OEM can offer multiple products over at least two CPU generations based on the same architecture.

The introduction of the single socket reference design offers operational efficiencies through I/O balance, simplified CPU pinning, and simpler workload orchestration. It is an excellent fit for a variety of workloads across different markets in Networking and Communications, Enterprise and Cloud, and Internet of Things (IOT). A small sample of use cases includes:

- Quick storage collection for video analytics
- Lightning-fast analytics in financial trading
- Efficient access to streaming video, and 5G core and optimized edge data processing in telecommunications.

The modular architecture provides TEMs, OEMs, and ODMs a product-ready design to bring any of these solutions to market.

Technical Details

The modular single-socket reference design is based on a single motherboard for 1U or 2U systems, both standard and short depth, and allows many I/O and storage configurations. Cabled PCIe*, growing in popularity, provides PCIe* flexibility for OEM servers and simplified and less expensive PCIe* Gen5 board routing. The operating system boots from the PCH to maximize CPU lane flexibility. Almost any type of storage is supported including HDD, U.2 and E.x drives. The BMC (Board Management Controller) provides server control, security, and management access. This design uses a modular BMC. This will accommodate DC-SCM v2.0 in future generations in alignment with those platforms.

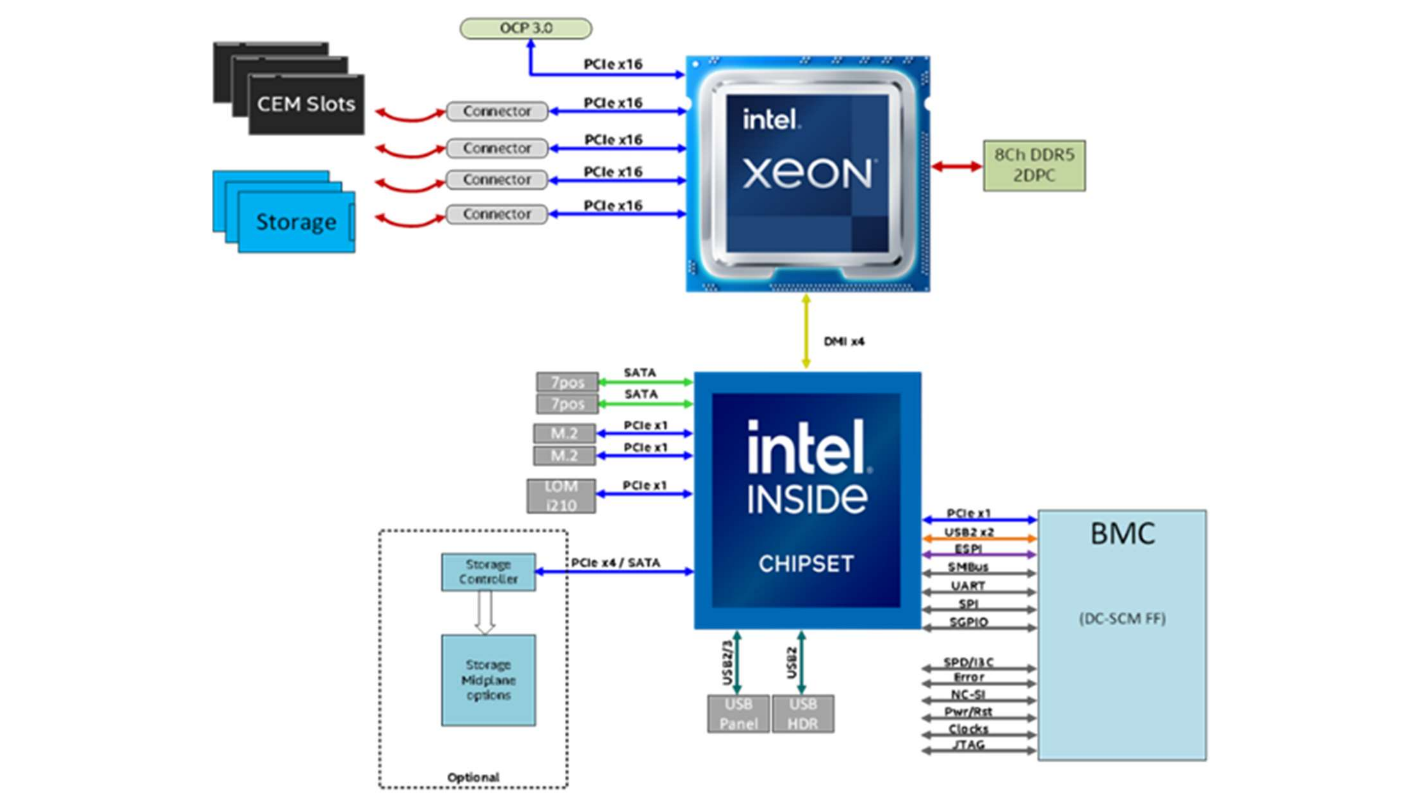


Figure 1. Modular Single-Socket Architecture

The motherboard is combined with modular system elements. An ODM or OEM may layout the elements based on component accessibility (i.e. front and/or rear in a rack), desired air flow, or other physical requirements.

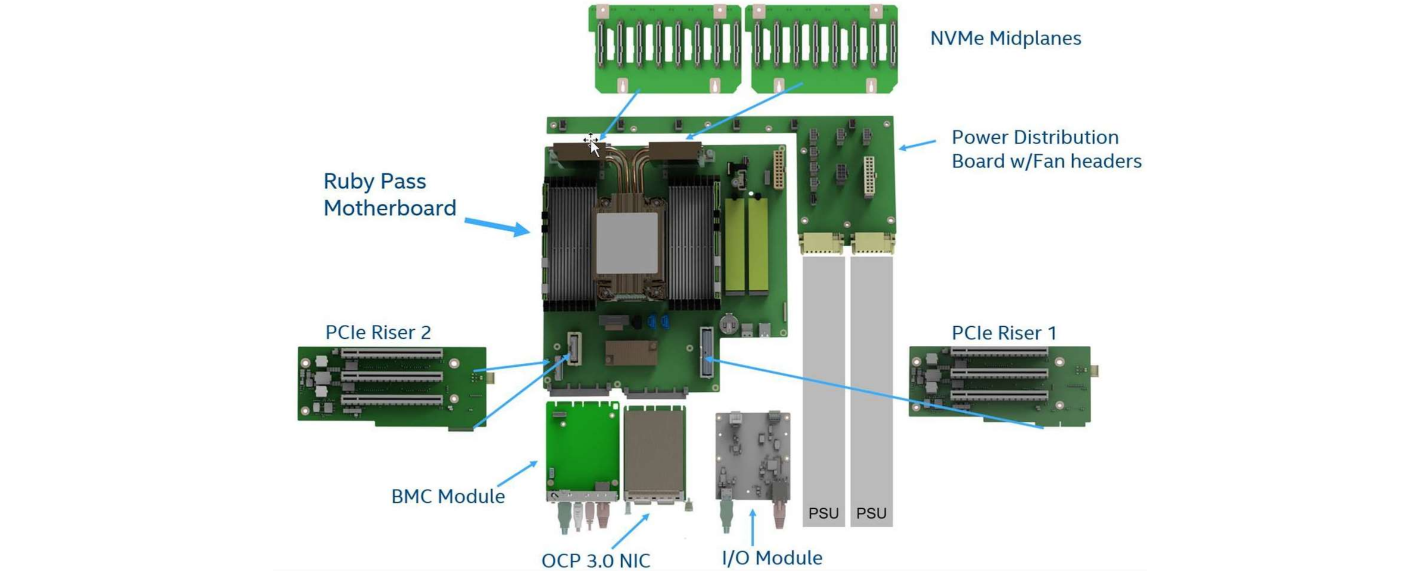
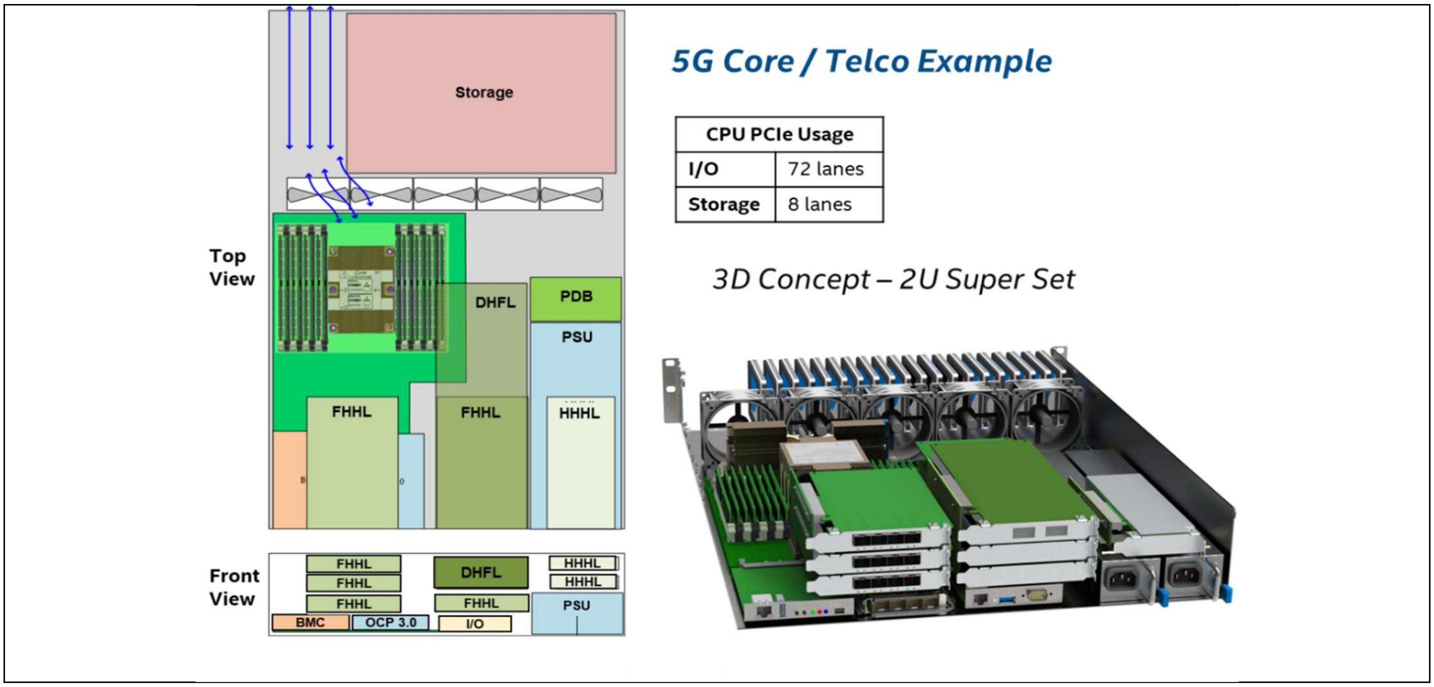


Figure 2. System Elements

The following examples are just four potential configurations based on the modular single-socket reference design.

The system in Figure 3 supports 72 I/O lanes with 8 lanes of storage. Note, the 2U super set 3D concept represents all possible accessories in this 2U chassis and includes OCP modules, legacy I/O modules, full length, and ½ length PCIe* modules, etc.



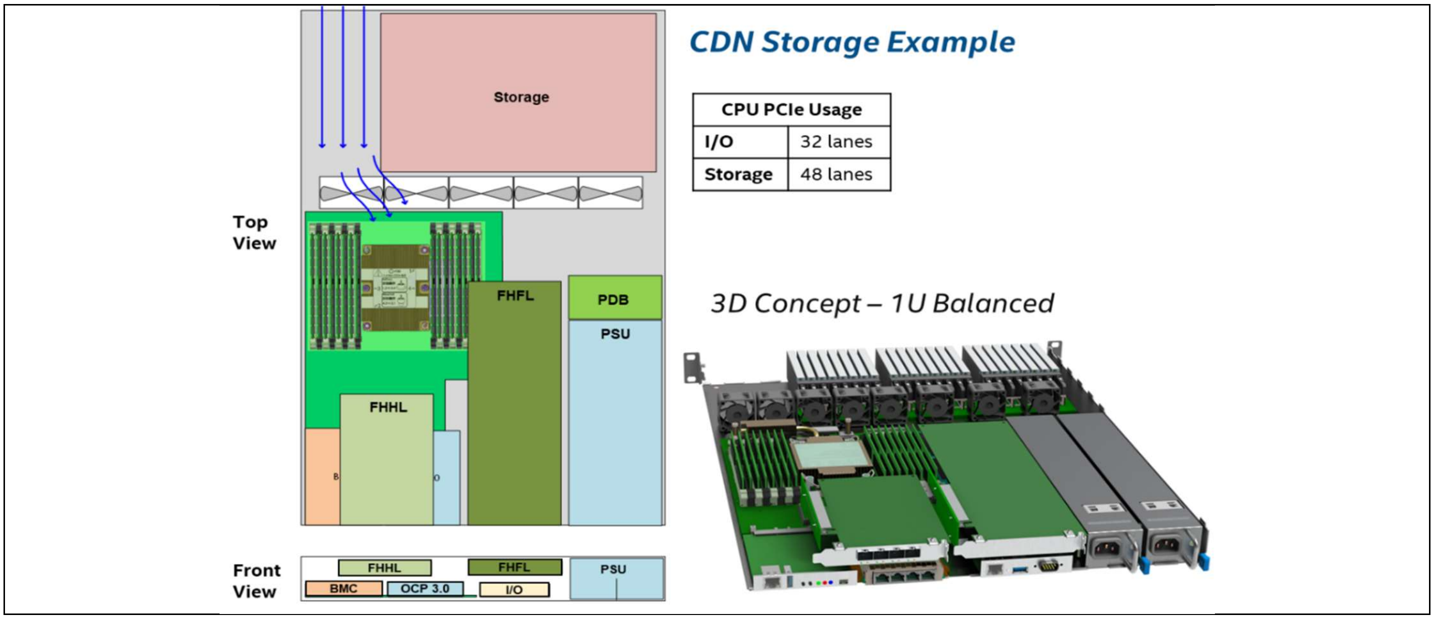
5G Core / Telco Example

CPU PCIe Usage	
I/O	72 lanes
Storage	8 lanes

3D Concept – 2U Super Set

Figure 3. 2U Standard Depth Configuration

Figure 4 shows a 1U, standard depth balanced system with 32 PCIe* lanes for I/O. It also has 48 PCIe* lanes for storage, which is useful for a content delivery network server.



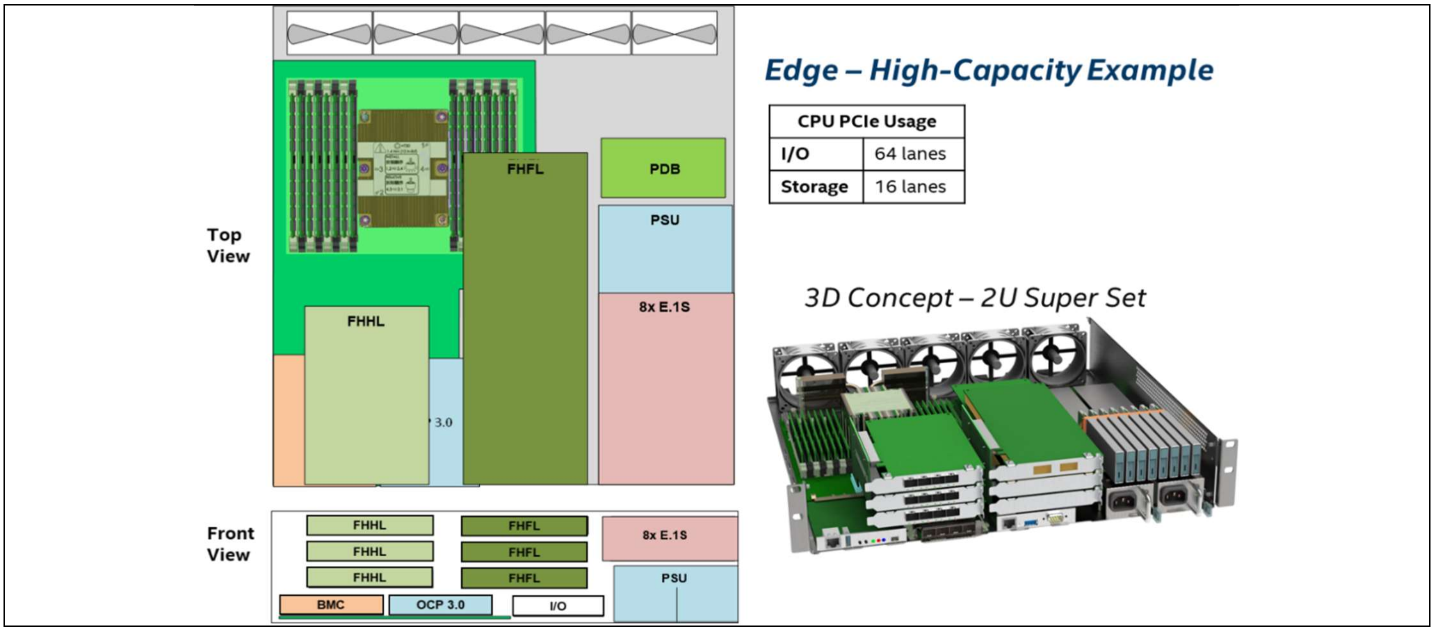
CDN Storage Example

CPU PCIe Usage	
I/O	32 lanes
Storage	48 lanes

3D Concept – 1U Balanced

Figure 4. IU Standard Depth Storage Configuration

Short depth configurations (shown in Figures 5 and 6) pull the storage to the "front" for installation in servers typically located in telco remote edge access locations especially where space is a premium.

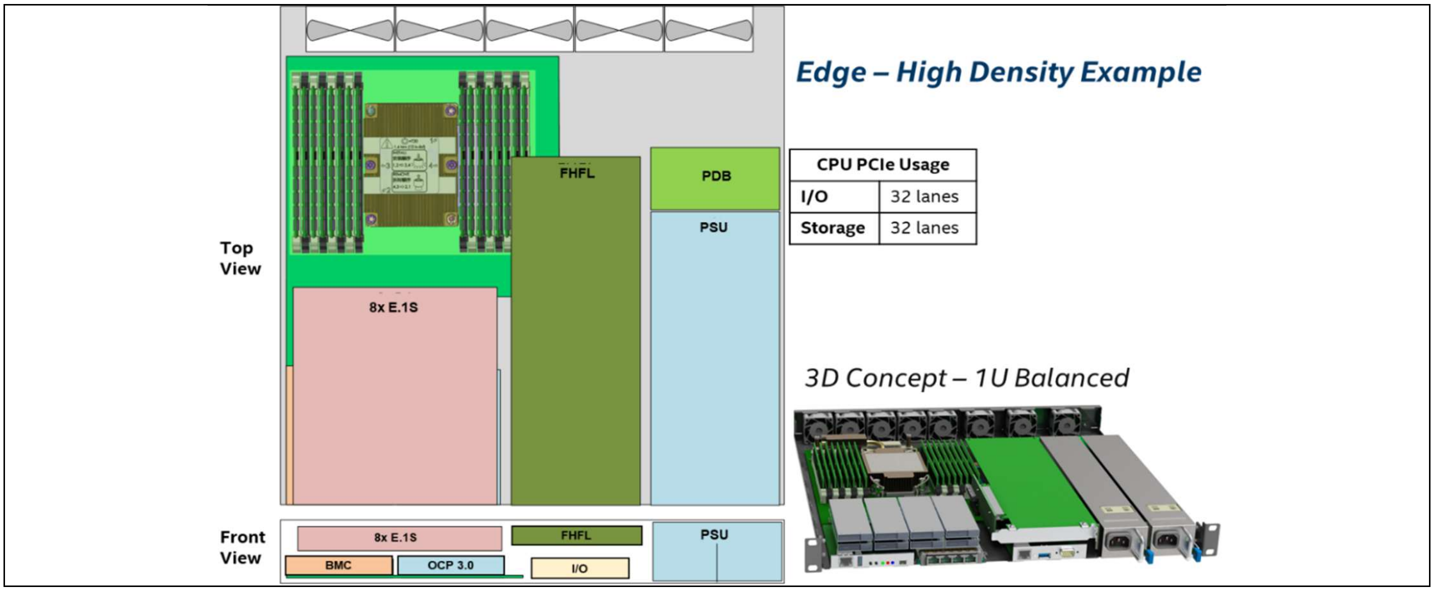


Edge – High-Capacity Example

CPU PCIe Usage	
I/O	64 lanes
Storage	16 lanes

3D Concept – 2U Super Set

Figure 5. 2U Short Depth – I/O Priority



Edge – High Density Example

CPU PCIe Usage	
I/O	32 lanes
Storage	32 lanes

3D Concept – 1U Balanced

Figure 6. IU Short Depth – Balanced

Technologies Implemented

Combining new technology and open standards in its modular, flexible design, this modular single-socket architecture is optimized for the 4th Gen Intel Xeon processor. The compute module is designed with a broad range of features that support XCC, MCC, and EE SKU variants.

The reference design features are detailed in the Table 1:

Table 1. Features of the modular single-socket reference design

Feature	Description
CPU	1S 4th Generation Intel® Xeon® Scalable processor SP–XCC, MCC (any non-HBM SKU) · Up to 350W TDP on 2U systems · Up to 240W TDP on 1U systems
PCH	Emmitsburg
Form Factor / Dimension (Board)	~11" x ~11"
Memory	16x DDR5 DIMMs, 4800MT/s (8 channels, 2DPC), VRoD Optane Persistent Memory
DMI	Gen3 x4 (4GT/s)
LOM	Intel® i210
PCIe* Generation / Lanes	PCIe* Gen5, 80 lanes 64 lanes provided through cabled connections
Rear PCIe* Risers	Two SFF-TA-1016 x 16 PCIe* Gen5
Front Drive Cable Connectors	Two (x16)/Four (x8) SFF-TA-1016 PCIe* Gen5
Front Storage Devices	8x U.2 NVMe drive slots
On-board SATA Ports	Minimum 4
OCP NIC	One OCP NIC 3.0 with a x16 G5 connection
M.2 Boot Drives	Dual Gen3 PCIe* x4
BMC	AST2600
Firmware Security (RoT)	Intel® PFR
Host USB Connection	One Front 3.0 / One Rear 3.0
TPM	Plug-in 2.0 Module
Rear Serial Console	Type B USB
Video Port	Front and rear connectors

4th Gen Intel Xeon processors offer new integrated features for security, improved acceleration and performance, and next generation memory support.

These features include:

Table 2. Features of 4th Generation Intel® Xeon® Scalable processors

Feature	Description
AMX	Built-in AI Acceleration engine for improved performance in deep learning inference and training • Target workloads and usages: • Image recognition • Recommendation systems • Machine/language translation • Reinforcement learning • Natural language processing (NLP) • Media processing and delivery • Media analytics
DLB	New integrated IP to increase throughput with efficient load balancing across multiple cores • Target workloads/Usages: • IPSec security gateway • VPP router • UPF • vSwitch • Streaming to data processing • Elephant flow handling
DSA	New integrated IP to accelerate applications reliant on data movement **Target workloads/usages:** • Virtualization: VM fast-checkpoint analysis • Network: vSwitch network vitalization • Storage: fast replication across non-transparent bridge • Application usage examples: messaging, ERP, In-Memory Databases, Analytics
IAX	New integrated IP to accelerate applications reliant on data improvement **Target workloads/usages:** • Commercial in-memory databases • Columnar Formats Big Data Analytics, Apache Parquet, Apache ORC

Feature	Description
	• Open-Source in-memory database/data stores, RocksDB, Redis, Cassandra, MySQL, PostfreSQL, MongoDB, Memached and more
QAT	Integrate IP to accelerate cryptography and data (de) compression
	Target workloads/usages: • Distributes storage systems (Ceph) • File systems (BTRFS, ZFS) • MSFT Azure Cosmos DB • RocksDB • Data lakes • Apache spark, Hadoop RDBMS • http compression • Memory infrastructure optimization
SGX	Trusted execution environment for increased protection of confidential data
	Target workloads/usages: Multi-party compute • Blockchain • Trusted multi-party compute • Federated learning/Secure analytics • Secure native application hosting • Secure database • Key management • Secure networking
CXL 1.1.	Improve accelerator performance via memory coherency and direct accelerator memory access
	Target workloads/usages: • Accelerator Attach: – Type 2 CXL device (.io, .mem, .cache - accelerator w/ private memory) – Type 1 CXL device (.io, .cache - accelerator w/o private memory)
Intel® Optane Persistent Memory 300 Series	Enables larger capacities and performance improvements
	Target workload/usages: • Hybrid cloud, IaaS, and Virtualization • Fast storage solutions • AI/Analytics, Machine Learning Analytics • IMDB and data analytics services
Next-gen IO Integrated PCIe* 5.0	Increased IO bandwidth and support for coherent interface with Compute Express Link v1.1
DDR5	Next generation memory support with higher speeds and increased memory bandwidth for memory intensive workloads

Edge workloads, like those listed in the Table 3 , take advantage of the 4th Gen Intel Xeon processor's integrated features and the modular single-socket design, making it ideal for many current Reference architectures and Intel® Select Solutions.

Table 3. Edge Workloads Supported

Workload	PCIe AIC 0	PCIe AIC 1	OCP AIC	Memory	Optane Persistent Memory	Storage NVMe	Boot Storage	LOM LAN on Motherboard	QAT Required
5G Core	E810-2CQDA2	Not Used	E810-CQDA2	256GB – 32GB/Ch	512GB	Not Used	2 x 480GB SSD	1G or 10G -> 25G	No
vRAN - centralized virtualized DU	E810-2CQDA2	vRAN ACC100	E810-CQDA2	128GB – 16GB/Ch	Not Used	Not Used	2 x 480GB SSD	1G or 10G	No
vRAN - CU	E810-2CQDA2	Not used	E810-CQDA2	128GB – 16GB/Ch	Not Used	Not Used	2 x 480GB SSD	10G	Recommended
vBNG	E810-2CQDA2	Not Used	E810-CQDA2	256GB – 32GB/Ch	Not Used	2x 8TB NVMe	2 x 480GB SSD	1G or 10G	No
CDN	E810-2CQDA2	Not Used	E810-CQDA2	256GB – 32GB/Ch	1.5TB	8x 16TB NVMe	2 x 480GB SSD	1G or 10G	Yes
MEC	E810-2CQDA2	vRAN ACC100	E810-CQDA2	256GB – 32GB/Ch	Not Used	4x 2TB or Greater	2 x 480GB SSD	1G or 10G	Yes
SASE	E810-2CQDA2	Not Used	Not Used	256GB – 32GB/Ch	Not Used	Not Used	1 x 480GB SSD	1G or 10G	Yes
SD-WAN	E810-2CQDA2	Not Used	Not Used	128GB – 16GB/Ch	Not Used	Not Used	1 x 480GB SSD	1G or 10G	Yes
OSS/BSS	E810-2CQDA2	Not Used	E810-CQDA2	256GB– 32GB/CH	Not Used	2x 8TB or greater	2 x 480GB SSD	1G or 10G	No
Open Cloud	Not Used	Not Used	E810-CQDA2	256GB– 32GB/CH	Not Used	4x 2TB or greater	1x 480GB SSD	1G or 10G	Recommended
AI	Not Used	Not Used	E810-CQDA2	256GB– 32GB/CH	512GB	1x 1.6TB NVMe	1x 480GB SSD	1G or 10G	Recommended
Media Analytics	Not Used	Not Used	E810-CQDA2	512GB– 64GB/CH	Not Used	4x 4TB or greater	2 x 480GB SSD	1G or 10G	Yes

Benefits of Solution

Single-socket solutions have become increasingly prevalent in the past few years, with companies embracing the benefits of these solutions. For example, Lenovo, HPE, Dell, and Supermicro offer one socket platforms for enterprise and network servers.

The use of single socket offers the following benefits:

- Operational efficiencies through I/O balance and guaranteed NUMA affinity
- Simplified CPU pinning, workload placement, no stranded capacity, and efficient vCPU allocation
- Simpler workload orchestration
- Efficient space, TDP, and thermals especially at the edge
- Deployment consistency and simplicity
- Single socket delivers a performance/watt advantage when compared to 2S
- Optimized 1-Socket solutions allow cost savings of approximately $150 per board design versus 2-Socket with one CPU de-populated

Furthermore, this modular architecture allows customers to reduce development costs over multiple generations, as the same, or very similar, compute modules can be deployed across a variety of server product lines. Future-proof, the same design can be re-used with the next Intel Xeon processor family by simply replacing the motherboard. This means that Intel's ODMs' and OEMS' initial investments in overall system design are good for at least five years, significantly increasing ROI from a "one and done" product.

Use Case Examples

The single-socket server has a broad range of uses across markets. For example, in networking and communications, it can be used for Edge, Telco cloud, data forwarding, data routing, 5G RAN and 5G Core.

In addition, across the Cloud and enterprise markets, the architecture can be implemented in IT Infrastructure/IaaS, value/mainstream digital services for next wave CSPs/FSIs etc., and general purpose/compute virtualization. Other use case examples include medical imaging, edge-based video analytics, and retail store analytics for the IOT market.

Table 4. Single Socket: Target Market Segments and Application Examples

Markets	Applications
Networking and Communications	• Edge (analytics, security, storage, CDN) • Telecommunications company cloud • Data forwarding • Data routing • 5G RAN • 5G Core
Cloud and Enterprise	• IT Infrastructure / IaaS • Digital services for next wave SPs, • FSI • General Purpose Compute • Edge
Internet of Things	• Medical imaging • Edge-based video analytics • Retail store analytics

Single socket optimized designs are ideal for next generation Edge Reference Architectures. Today, there are tens of centralized data centers with a large number of servers. However, in the edge evolution, the world is shifting to distributed data centers, combined with a centralized infrastructure to improve latency, and application locality (i.e. closer content distribution). A single-socket, 2U short depth configuration can address many applications. 5G vRAN servers are thermally challenged and require short depth systems with front access. With this design, multiple cloud workloads can be addressed. For the enterprise, this cost-optimized system lets additional systems be easily added to a rack in "pay as you grow" scenarios. Networking and communications require simplified deployment and easy orchestration. Easy peripheral and accelerator expansion was also a design goal for IOT and SASE locations. With the newest AMX integrated instructions in the 4th Gen Intel Xeon processor for machine learning and analysis, edge servers can enhance applications with AI.

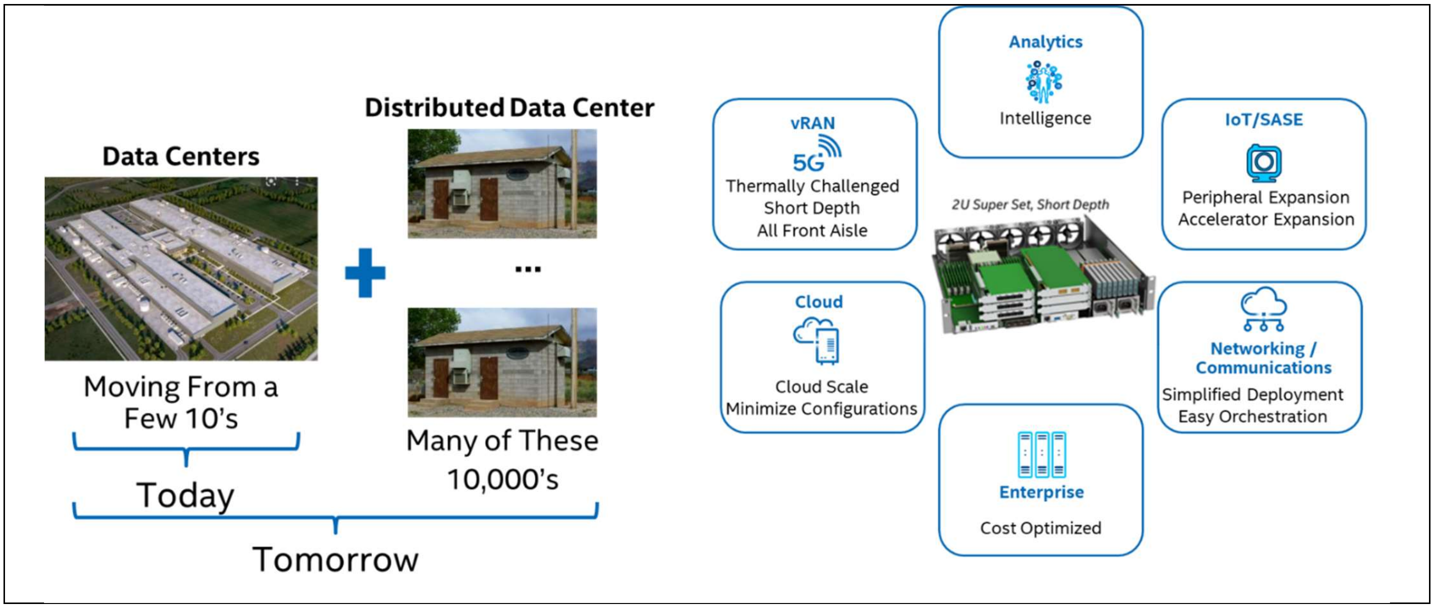


Figure 7. Where and how is single socket being deployed?

Summary

A modular single-socket reference design allows customers to reduce development costs over multiple generations. The same or very similar compute modules can be deployed across a variety of server product lines. Future-proof, the same design can be re-used with the next Intel Xeon processor family by simply replacing the motherboard. This means that Intel's ODMs' and OEMs' initial investments in overall system design are good for several years, significantly increasing ROI from a "one and done" product.

Intel is delivering a boost to accelerate one socket designs across all Intel's target markets. This design is that vehicle – a modular, flexible, standards-based reference design providing a head-start to productize one socket solution for 4th Generation Intel Xeon Scalable processors and next generation Intel Xeon.

The modular single-socket system was designed in collaboration with Jabil. Contact Jabil for information on related products.

The Reference Design (including schematics and board layout files) is available on the Intel® Resource Design Center (Reference ID: 648338).

§

Performance varies by use, configuration, and other factors. Learn more at www.Intel.com/PerformanceIndex.

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 disclaims all express and implied warranties, including without limitation, the implied warranties of merchantability, fitness for a particular purpose, and non-infringement, as well as any warranty arising from course of performance, course of dealing, or usage in trade.

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

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

The products described may contain design defects or errors known as errata which may cause the product to deviate from published specifications. Current characterized errata are available on request.

0123/SY/TL/PDF 733354