Solution Brief

Energy, Financial Services, Healthcare, Enterprise, Government, Telco

Server Design Brings Open, High-Density Compute to Edge

Supermicro, Red Hat, Senao, Lanner and Intel join forces to create a family of edge computing solutions that are AI-ready and scalable for workload consolidation

Deploying workloads on the edge is an expensive and daunting task for most enterprises. Legacy servers were designed for data center deployments, but the needs of edge computing are different from those of data centers in many ways.

And yet, there is great and growing demand for edge as more workloads need real-time processing and as organizations want to move compute closer to the edge where the data is being created and consumed. Market research done by industry analyst firm STL Partners[i] shows that the addressable opportunity of the edge computing market will exceed $500 billion by 2030.

To show that scalable and industry standard edge servers are a possibility, Supermicro, Red Hat, Senao Networks and Lanner - all of which are Intel® Industry Solutions Builders Partners - and Intel have developed a server reference design.

Edge Compute Reference Design

The partners are committed to configuring the server using standard off-the-shelf software and hardware. While the server configuration in this paper will be targeted at a particular use case there is a lot of flexibility to evolve the system to meet different needs. This reference design provides the end user with full control of how they configure the system for each use case.

These systems need to be flexible because they can handle a large number of different workloads. Some common workloads include AI inferencing, security, networking, security camera video analysis, SCADA, POS, and others.

The solution proposed for this reference design is modular and delivers a more cost-effective way for workload deployment. This approach is scalable, more cost effective and uses familiar infrastructure. The hardware foundation of this reference design are Intel®-powered servers from Supermicro and PCIe based resource augmentation cards.

Choose from Wide Range of Servers

Supermicro has specified more than 90 servers, some new and specifically designed for edge use cases, that can be part of the reference design allowing the end solution to be customized by form factor or by compute power. With support for full-height, full-length (FHFL) PCIe slots, additional functionality can be added to the servers to allow for flexibility and scalability across many use cases. These servers fall into three Supermicro product families:

Rackmount Edge: These are single-processor edge servers. They can be configured with Intel® Core™ processors, Intel® Xeon® processors or Intel® Xeon® SoCs. The servers are workload-optimized for the edge. They are rack mountable and are available in either 1U or 2U heights and are available in a short depth form factor with front I/O for space-constrained cabinets.

Rackmount 1U Dual Processor: These are adaptable systems with next-generation performance for a variety of workloads. These systems are powered by dual Intel Xeon 6 processors or Intel® Xeon® Scalable processors. They are all-in-one servers and come complete with configurable storage. The servers support PCIe 5.0 and PCIe 4.0 for extended functionalities.

Rackmount 2U Dual Processor: These servers have all the benefits and flexibility of the 1U systems but with a 2U form factor for additional PCIe expansion options. They also have configurable high-speed storage.

Intel Processors Balance Performance, Power

Intel processors that drive these servers include the latest Intel Xeon 6 processors, and Intel Core processors.

Intel Xeon 6 processors are built for diverse power, performance, and efficiency requirements at the edge, in data centers, and in cloud servers. Efficient-core (E-core) and Performance-core (P-core) options are available to meet various workload requirements. The E-core option offers up to hundreds of cores per socket, and P-core variant offers frequencies approaching 4 GHz. Intel Xeon 6 processors enable massive consolidation of edge workloads in space- and power-constrained environments.

Key features for edge networking include integrated Intel® vRAN Boost, Intel® Advanced Matrix Extensions (Intel® AMX) for AI inference, and integration of Intel® QuickAssist Technology (Intel® QAT) to accelerate cryptography and compression.

Intel Xeon Scalable Processors

Intel Xeon Scalable processors deliver a balanced combination of core density, clock speed, and integrated accelerators for mainstream edge and infrastructure deployments.

5th Gen Intel Xeon Scalable processors support up to 60 cores per socket with maximum base frequencies up to 3.9 GHz, enabling strong single-thread and multi-thread performance for mixed workloads. Built-in acceleration technologies such as Intel QAT, Intel® Dynamic Load Balancer (Intel® DLB), and Intel® Data Streaming Accelerator (Intel® DSA) are particularly valuable for virtualized networking, security, and software-defined infrastructure at the edge. Broad ecosystem support and mature software optimization make Intel Xeon Scalable a versatile foundation for distributed edge platforms.

Intel Core Processors

The Intel Core processor provides a cost-effective, power-efficient option for edge nodes, branch locations, and compact network appliances that still require high performance. With configurations offering up to 24 cores and maximum turbo frequencies reaching 5.8 GHz, Intel Core processors excel in latency-sensitive workloads, containerization and virtualization.

Workload Orchestration Platform

For orchestration of virtual machines and containers, the reference design specifies Red Hat OpenShift Container Platform (OCP), an enterprise-grade, hybrid cloud platform built on Kubernetes. The software is used to develop, deploy, and manage containerized applications consistently across on-premises, public, and edge environments.

OCP provides support for virtual machine workloads, allowing enterprises to migrate, manage and deploy VMs. The software is also a comprehensive application platform with a full set of operators, developer services and developer tools.

OCP comes with Red Hat Enterprise Linux CoreOS, over-the-air updates, container runtime, networking, ingress, monitoring, logging, container registry, authentication, and authorization solutions. These components are tested together for unified operations on a complete Kubernetes platform spanning every cloud environment.

Red Hat OpenShift Container Platform running on Intel processors is key to edge computing strategy and to start building hybrid and cloud-native applications. Supermicro has worked with Red Hat to certify OCP on many edge server platforms. The proven platform includes a complete set of services that empower developers to code with speed and agility for applications while providing more flexibility and efficiency for IT operations teams.

Intel® NetSec Accelerator Reference Design

The Intel NetSec Accelerator Reference Design enables an Intel processor-based server in a PCIe-based add-in-card. These add-in modules provide additional compute power for applications running on the server system, including network infrastructure and security function processing, offloading the tasks from the host CPU. The add-in-card can also be operated in isolation from the host server where separation of workloads from server host CPU is desired. Since Intel NetSec Accelerator Reference Design specifies Intel processors, it provides coherent computing architecture where the same Intel architecture optimized software can be executed on both the server host CPU and on the add-in-card.

In this paper, examples of Intel NetSec Accelerator Reference Design based add-in-card servers are added to the solution.

Senao SmartNICs were added to the solution to enable seamless and efficient scalability for AI-driven workloads and to improve infrastructure headroom for edge applications, network throughput, and security workloads by augmenting compute resources available on the server host CPU.

The SmartNICs that are a part of the solution are the SX904 and SX906 (see Figures 1 and 2). Both SmartNICs are built on Intel NetSec Accelerator Reference Design. The foundation for both SmartNICs is the Intel® Ethernet Controller and Intel SoCs.

The SX904 SmartNIC, powered by an Intel® Xeon® D processor with dual 25G SFP28 connectivity, delivers efficient compute and networking offload in a PCIe Gen4 x8 form factor. The SX906 SmartNIC builds on this architecture with an Intel® Xeon® 6 SoC and dual 100G QSFP28 interfaces, enabling higher-throughput networking and greater performance for AI-driven, security, and data-intensive edge workloads. Both models incorporate an onboard BMC for out-of-band management and Intel® Platform Firmware Resilience (PFR) to protect, detect, and recover from firmware-level attacks.

In addition, Senao SmartNICs are designed to integrate with Red Hat OpenShift Container Platform, allowing the cards to be deployed as bare-metal nodes within an OpenShift cluster. This enables containerized workloads to run directly on the SmartNICs while benefiting from OpenShift's centralized orchestration, automated scaling, and high-availability features, extending cloud-native application deployment consistently across host systems and edge compute extensions.

Figure 1. Senao Networks SX904 is a PCIe Ethernet NIC with up to 10 Intel® Xeon® D cores.

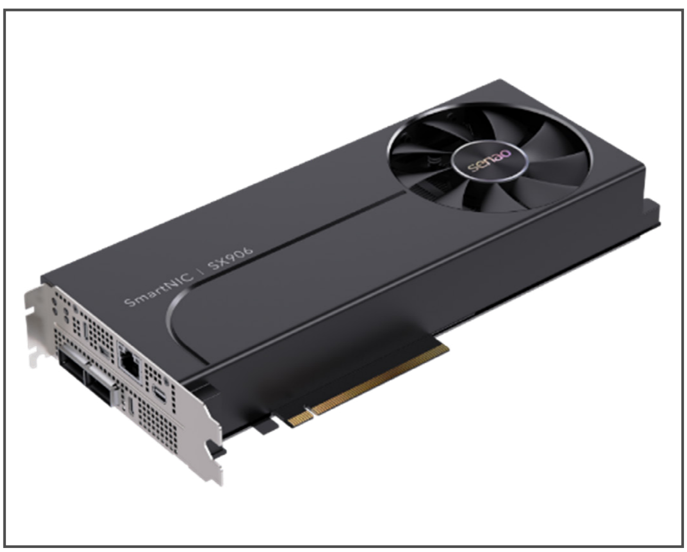

Figure 2. Senao Networks SX906 is a PCIe Ethernet NIC with up to 36 Intel® Xeon® 6 SoC cores.

SmartNIC Option: Lanner IAC-PTL301A

The reference architecture has also specified the Lanner IAC-PTL301A SmartNIC that is based on the Intel NetSec Accelerator Reference Design.

The IAC-PTL301A (see Figure 3) is an Intel Xeon 6 SoC-based server in a FHFL PCIe card form factor that supports up to 32GB of 2CH ECC DDR5 memory. The card features one 1 GbE RJ45 port and two 100G QSFP56 ports. An NVMe slot can accommodate up to 128GB of onboard storage.

The IAC-PTL301A is optimized for high-density virtualized workloads, offering flexible port configurations, advanced time synchronization capabilities and superior security features.

Figure 3. Lanner IAC-PTL301A SmartNIC is a PCIe Ethernet NIC with an Intel® Xeon® 6 SoC for hardware acceleration.

The SmartNIC supports packet synchronization via IEEE 1588 PTP, SyncE, and GNSS integration to deliver sub-microsecond alignment of clocks across networked devices, ensuring deterministic operation in time-critical applications.

The IAC-PTL301A benefits from all the capabilities of the Intel Xeon 6 SoC including AI acceleration, cryptography acceleration, confidential computing, as well as vRAN and media transcoding accelerations. For hardware security, Lanner leverages Intel® Secure Boot, Intel® Secure Firmware Upgrade and dual hardware roots of trust.

Supermicro Demos SYS-322GA

Figure 4 depicts the physical layout of a demonstration design based on the SYS-322GA. Each Senao SmartNIC has two optical pluggable (QSFP) ports and a copper Ethernet BaseT port supporting up to 100 GbE.

The external networking using the SFP and RJ45 ports requires supporting switches. The various SmartNICs are plugged into the PCIe slots allowing for a high-speed backplane between the host platform and the cards. Additional cards such as the Intel® Arc™ Pro B60 Graphics could also be installed if additional AI resources are required.

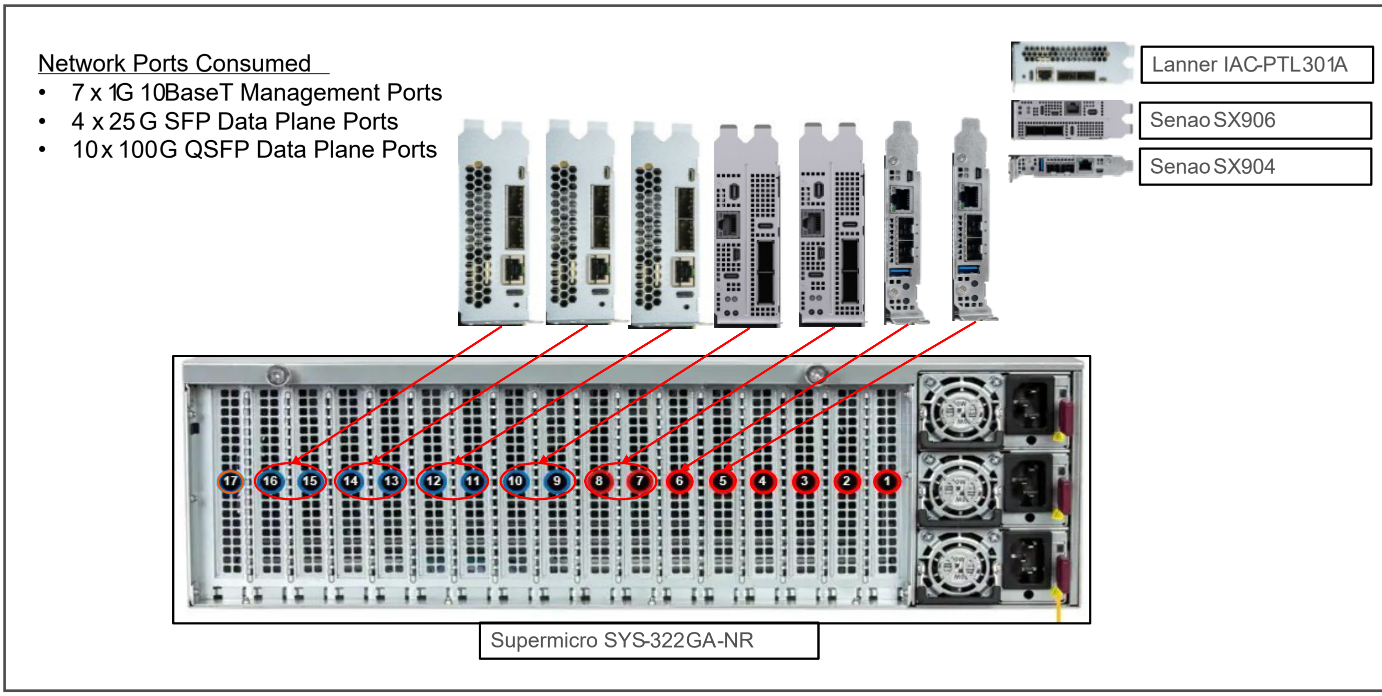

Figure 4. This picture is a logical depiction showing how the PCIe cards in the demonstration system are networked with the host system using both the cabled connections as well as a virtual switch running on the host platform.

The architecture of using a form factor that can accept multiple PCIe cards enables a very high-density compute environment when leveraging CPU augmentation cards like the NetSec Accelerator. Each card can run independently as a stand alone server but also be networked over the backplane allowing for chained services. The host platform can leverage Intel® Xeon® 6900 processors with up to 256 cores on a single motherboard. By using OpenShift the whole platform can be orchestrated via Kubernetes containerization.

Total Cost of Ownership

The intent of this reference design is to illustrate a highly dense CPU architecture that provides maximum scalability and flexibility in a very small rack footprint. This modular approach to deploying compute resources enables a very robust edge deployment in a very typical, but space-constrained environment.

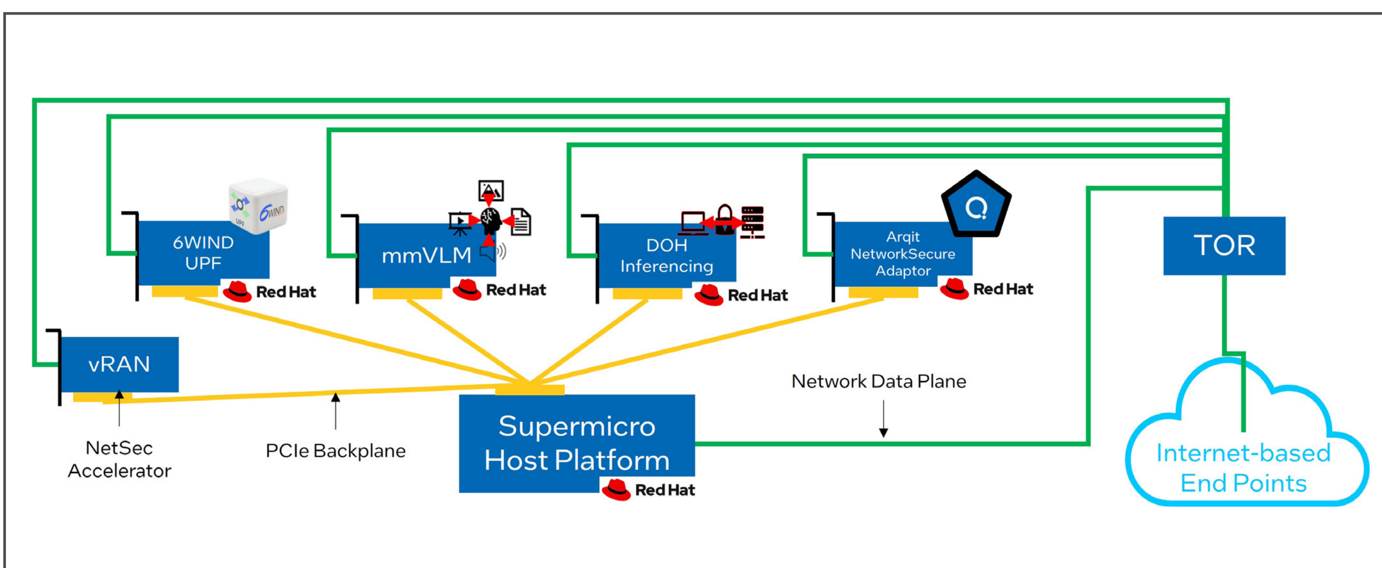

Figure 5. This image illustrates an example of a software deployment on the reference platform with each component providing network functions for a private 5G deployment on a MEC with network security functions and inferencing solutions.

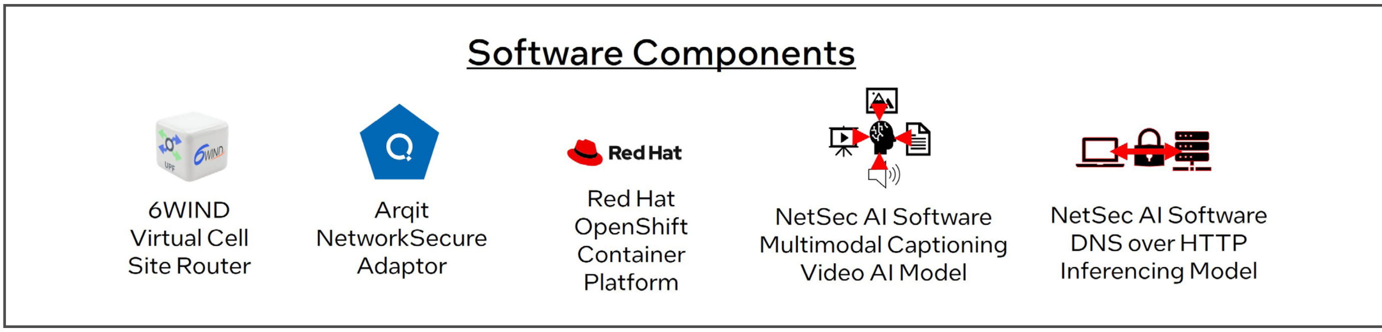

Figure 6. This highlights several examples of software validated on products based on Intel® Netsec Accelerator Reference Design that could be used in the deployment model highlighted in Figure 5.

Figure 6 is an illustration of the workloads that can be run as independent services in a multi-access edge compute (MEC) location. This example shows how the environment is orchestrated using Red Hat OpenShift. A 6WIND Cell Router is used for a distributed user plane function for private 5G with the Arqit software running as a chained service to ensure post quantum compute security for the packet flow. The Intel NetSec AI software for DOH is also chained to identify possible intrusion. The mmVLM is an example of an additional application that is for visual security. This deployment model represents a significant cost savings versus a traditional server deployment.

This reference design does require some planning in order to provide the adequate resources to support the hardware. The main concern is power and cooling availability. The power demand for this architecture requires three 220 VAC power drops. The cooling can be provided based on standard server air conditioning.

If the edge location has limitations on space and cooling other host platforms can be selected. The tradeoff is the number of processor cores that can be supported and thus the workloads that can be deployed.

Figure 7 shows a comparison of resources and cost based on three different deployment options.

Platform	Resources	Server Platform Cost
Supermicro SYS-322GA-NR  + NetSec Accelerator  https://www.thinkmate.com/system/superserver-322ga-nr	One Server Xeon 6 (6952P) 196 Cores 9 NetSec Cards Xeon 6 SOC 36 Cores Ea. Total of 520 Core of Xeon 6 3 RU Consumed Cooling Capacity 3044W (Air) 20 QSFP ports required	Approx. $78,300
Supermicro SYS-122H-TN  https://www.thinkmate.com/system/superserver-122h-tn	Three Servers Xeon 6 (6768P) 172 Cores Ea. Total of 516 Cores of Xeon 6 3 RU Consumed Cooling Capacity 2100W (Air) 6 QSFP Ports	Approx. $179,400
Supermicro SYS-222C-TN  https://www.thinkmate.com/system/superserver-222c-tn	Three Servers Xeon 6 (6768P) 172 Cores Ea. Total of 516 Cores of Xeon 6 6 RU Consumed Cooling Capacity 2100W (Air) 6 QSFP Ports	Approx. $127,800

Figure 7. Modular hardware design cost comparison.

This comparison highlights three deployment options[ii]. The first is based on the reference design described in this paper. The second shows the optimum design based on standard 1U rack mount servers. The third shows a deployment using 2U servers. To achieve a consistent comparison, the compute resources are based on the same generation of Intel Xeon 6 processors, similar NICs, and similar storage devices.

As can be seen in this depiction, the SYS-322GA platform from Supermicro provides a significant TCO benefit with fairly identical CPU resources.

Conclusion

There is a solution for organizations to deploy an open, configurable and flexible edge compute device. Working together, Supermicro, Red Hat, Senao Networks, Lanner and Intel have assembled a reference design of an edge server with the expandable compute performance that can scale security and other edge computing applications. Supermicro servers combined with SmartNICs from Senao and Lanner provide coherent compute architecture that integrate CPUs on both the host server and the SmartNICs, to deliver exceptional scalability in a cost efficient manner.

Learn More

Supermicro Rackmount Edge

Supermicro Rackmount 1U Dual Processor

Supermicro Rackmount 2U Dual Processor

Red Hat OpenShift Container Platform

Senao Networks SmartNICs

Lanner IAC-PTL301A

Intel® Xeon® 6 processors

Intel® Arc™ Pro B60 Graphics

Intel® Xeon® Scalable processors

Intel® Core™ processor family

Intel® Ethernet products

Intel® NetSec Accelerator Reference Design

Intel® Industry Solutions Builders

[i] https://stlpartners.com/articles/edge-computing/edge-computing-challenges/

[ii] The server cost is based on an online quote for these servers using WWW.THINKMATE.COM

0226/MK/DJA/PDF Please Recycle 367910-001US