White Paper

Industrial system designers
Manufacturing, industry

Harnessing Industrial AI with 4th Gen Intel® Xeon® Scalable CPUs

As manufacturers embrace Industry 4.0 technologies for smart manufacturing, 4th Gen Intel® Xeon® Scalable processors have the power and features to drive efficiency and autonomous manufacturing decision-making

Discrete and process manufacturing industries are the new targets for efficiency improvements as companies seek to use the latest technologies to lower costs and improve sustainability.

A global initiative is under way to advance a new era of manufacturing digitalization. The market for smart manufacturing is estimated to grow from $214.7 billion in 2020 to $384.8 billion in 2025, growing at a compound annual growth rate of 12.4%.[1]

Also known as Industry 4.0, these technologies connect un-connected equipment, integrate smart technologies, and increase automation. All this is informed by analytics and uses artificial intelligence and machine-learning technologies to create on-demand manufacturing and interconnection of the manufacturing operation to the entire value chain. Three technologies are the driving forces behind this shift: artificial intelligence (AI) or machine learning (ML), the internet of things (IoT), and private 5G. When used in combination, they can improve manufacturing in unprecedented ways.

However, these advances require an increase in manufacturing system compute power as companies collect increasing amounts of data on their manufacturing processes. And to ensure efficiency, this data must be processed with very low latency. The 4th Gen Intel® Xeon® Scalable processor family has features that improve the performance of a wide range of Industry 4.0 applications.

Edge servers: A new compute delivery model

With 75% of all data projected to be created outside of central data centers by 2025[2], better Industry 4.0 outcomes depend on locating compute close to the network edge where data is generated. The benefits of edge computing include faster downloading and analyzing of data, decreased latency and connectivity costs, and increased bandwidth, security, management, and intelligence.

For manufacturers, edge servers can improve process quality by making real-time decisions, integrating optimized data, and handling complex data source analysis. Edge servers' potential to reduce latency also dramatically increases the benefits of utilizing IoT devices. In addition, the AI features of edge servers accelerate AI applications and provide high virtual machine density for edge computing.

Edge servers are also flexible, serving as connecting points between information and operational technology. And they offer secure and reliable storage when storage functions are distributed in multiple servers around a network perimeter.

There are two primary types of edge servers. On-premises edge servers can be enterprise servers located in a factory floor data center. They deploy compute and storage resources at the same location where the data is produced, solving latency issues for large data sets. On-premises edge servers have AI, machine vision, fleet management, and legacy factory applications.

Table of Contents

Multi-access edge computing (MEC) servers are entwined with 5G applications but are not exclusive to them. Like on-premises servers, they solve latency issues by increasing reliability and overall network efficiency. Their standard architecture helps enhance the functionality and capabilities of networks. MEC servers have numerous applications for smart manufacturing and smart factories.

Using Intel® processors at the edge provides a solution-level approach that incorporates performance optimizations, security, with industrial use conditions and long availability, and security processing and storage of sensitive factory data.

A family of CPUs for smart manufacturing

To support Industry 4.0 applications, 4th Gen Intel® Xeon® Scalable processors offer a new microarchitecture designed to address dynamic and demanding workloads at the industrial edge. With 4th Gen Intel® Xeon® Scalable processors, companies can accelerate their most advanced edge workloads, including AI delivering fast query throughput, data movement, and data compression with better CPU utilization.

The 4th Gen Intel® Xeon® Scalable processors can have up to 56 cores along with multiple built-in accelerators to help maximize performance efficiency for emerging workloads (see Figure 1).

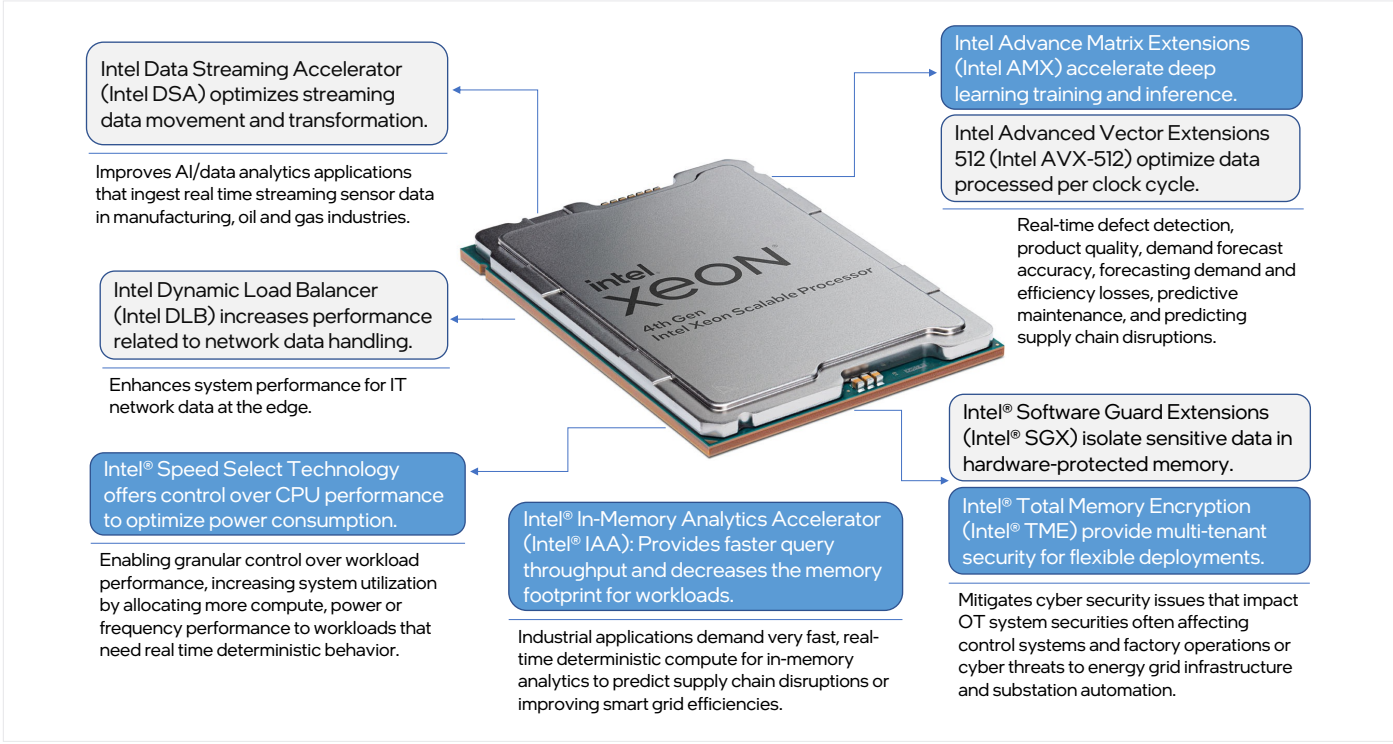

Figure 1. The 4th Gen Intel® Xeon® Scalable processor features six accelerators and up to 56 cores for performance as well as specialized features for industrial usage.

These latest processors offer a breakthrough in I/O capabilities, with up to 80 lanes of PCIe 5.0 connectivity and Intel® Scalable I/O Virtualization for more virtual machines per server. The CPUs also support Compute Express Link (CXL), a cache-coherent interconnect for processors, and accelerators. In many industrial applications, low power consumption is a difference maker and the 4th Gen Intel Xeon Scalable processor has a TDP of 125W for all 56 cores, compared to its competitor's TDP. The CPU also offers a smaller trust boundary to support the most security-conscious industrial applications.

Software tools and compatibility guide development

Intel® architecture CPUs feature software compatibility with other Intel devices, enabling a develop once, deploy everywhere approach. Intel CPUs allow support for the widest range of commercial and open-source operating systems, virtual machines or containers, and cloud platforms.

Intel-developed software that is compatible with 4th Gen Intel® Xeon® Scalable processors includes the OpenVINO™ toolkit to deploy high-performance, deep learning inference. OneAPI is an open, cross-architecture programming model that frees developers to use a single code base across multiple architectures.

To smooth the development of edge and AI/ML applications, Intel has developed Intel® Developer Cloud for the Edge, designed to help evaluate, benchmark, and prototype AI and edge solutions on Intel hardware. The Intel® Edge Software Hub allows users to download pre-validated software to learn, develop, and test solutions for the edge. And the Intel Training and Learning Suite 2.0 is an application toolkit to perform end-to-end training and inferencing.

The Intel® Edge Insights for Industrial (EII) software package is a pre-validated, ready-to-deploy containerized software reference solution for video and time series data ingestion. It includes AI analysis and can publish to local applications or the cloud.

Figure 2. Industry 4.0 applications for Intel Xeon Scalable processors.

Some of the ways Intel is helping developers to easily develop and deploy AI solutions on 4th Gen Intel® Xeon® Scalable processors include:

- Support for a wide range of commercial and open-source OSes, virtual machines and containers
- Open tool kits such as OpenVINO™ and OneAPI programming model
- Intel® Edge Insights for Industrial (EII) reference software solutions
- Intel® Developer Cloud for prototyping and benchmarking

How 4th Gen Intel® Xeon® Scalable processors benefit industrial processes

Using 4th Gen Intel® Xeon® Scalable processors in manufacturing settings that have embraced Industry 4.0 can result in numerous benefits.

For example, one industrial die caster was experiencing rising costs from labor, defective products, and a lack of technical support to undertake a smart factory transformation. With the help of Intel scientists, the die caster designed a high-precision computer vision algorithm using edge servers for four identified defects, resulting in a near 100% detection rate.

In another real-world implementation, the manufacturing plant of a European car maker was manually inspecting welds on only one out of 1,000 cars produced in the plant per day. With the help of Intel scientists, the manufacturer wrote, trained, and implemented a quality control analytics model on a live production line using Intel edge servers and a third-party edge software platform. These real-time analytics evolved quality control from a manual sampling of one car (with 5,000 welds) per day to inline to 100% automated inspection of all plant welds (5 million welds), thereby reducing labor costs by up to 50%.

Workload consolidation

As factories seek optimization and new efficiencies, they often turn to workload consolidation. The industry is shifting away from myriad disaggregated systems to a more centralized approach in which multiple workloads, including those that are very latency sensitive, are consolidated onto a single Intel processor. This gives operators greater levels of control and access to insights.

One evolving use case in industrial manufacturing centers on improving defect detection and quality inspection with input from a variety of cameras, all of which require swift and accurate image recognition capabilities. Manufacturers can integrate emerging machine vision and deep learning workloads into their current infrastructure with Intel® architecture CPUs while avoiding expensive GPUs with high costs and thermal and power requirements.

These improvements are made possible through advancements in enhanced AI acceleration for training and inferencing using Intel® Advanced Matrix Extensions (AMX), OneAPI, and OpenVINO™, available on the 4th Gen Intel® Xeon® Scalable processors. By accelerating inference time, the result is improved performance and a higher value density of the solution.

Applications also extend beyond manufacturing to oil and gas, energy and utilities, and logistics. One power and water utility company serving a major U.S. city was experiencing appliance sprawl at substations, including custom hardware and software tools, limited security and manageability, and increasing capital-intensive operation and maintenance (O&M) costs. Implementing Intel-based edge servers with real-time virtualization software, numerous operational technology applications such as high-voltage protection and control and fault analysis and classification, and IT applications such as virtual firewalls, resulted in a 76% reduction in O&M costs and a 50% reduction in devices used.

Conclusion

A new era of manufacturing digitalization is using advanced technology for autonomous decision-making. To support Industry 4.0 applications, 4th Gen Intel® Xeon® Scalable processors offer a new microarchitecture to address dynamic and demanding workloads at the industrial edge. These processors offer breakthroughs in server performance, hardware accelerators, engines for AI capability, memory capacity and I/O capabilities, multitenant security, reliability, and long-term availability. Intel CPUs also support the widest range of commercial and open-source operating systems, virtual machines or containers, and cloud platforms. Using 4th Gen Intel® Xeon® Scalable processors in manufacturing settings that have embraced Industry 4.0 can result in numerous benefits, in use cases such as manufacturing defect detection, digital twins, predictive analysis, quality control, and semiconductor testing.

Learn More

Intel® Xeon® Scalable processors

Intel® Software Guard Extensions

Intel® Total Memory Encryption

Intel® Speed Select Technology

Intel® Resource Director Technology

Intel® Advanced Matrix Extensions

Intel® Data Streaming Accelerator

OpenVINO™

OneAPI

Intel® Developer Cloud for the Edge

Intel® Edge Software Hub

Intel® Edge Insights for Industrial

[1] https://www.researchandmarkets.com/reports/5013081/smart-manufacturing-market-by-enabling-technology
[2] https://www.gartner.com/smarterwithgartner/what-edge-computing-means-for-infrastructure-and-operations-leaders

0323/DJA/H09/PDF ♻Please Recycle