

Accelerating HAProxy® on 4th Gen Intel® Xeon® Scalable Processors

HAProxy delivers high-performance HTTPS/TCP load balancing on 4th Gen Intel® Xeon® Scalable processors with Intel® QuickAssist Technology for cryptographic acceleration



Today, businesses are facing ever-increasing demands for network and server performance driven by the increasing number of mobile users and the applications migrating to cloud data centers and edge.

Industry trends are driving the demand for infrastructure scale, increased performance and operational efficiency, ease of management and enhanced security to handle sensitive and mission-critical workloads and data.

As more traffic moves to the cloud and edge, there is an increasing need for advanced and high-performance load balancers for an infrastructure to provide a flexible, secure, and reliable network to deliver business results for users.



HAPROXY

In load balancing environments, SSL/TLS processing is CPU intensive and is the essential driver for using more CPU resources. Performance scalability is never linear due to thread contention that occurs when scheduling work within multi-threaded software. Often, the number of cores available grows faster than the software's ability to utilize them efficiently, and the expected performance level falls short. In addition to these pitfalls, many users try as much as possible to keep a single load balancing layer and prefer not to use more than one active machine for their load balancers, which increases demand on the software and affects performance.

HAProxy is an open-source software load balancer and reliable reverse proxy, optimized to run on 4th Gen Intel® Xeon® Scalable processors with built-in Intel® QuickAssist Technology (Intel® QAT) accelerators. This combination delivers a highly optimized and cost-effective solution with real-time data and robust network security addressing the above-mentioned load balancer systems challenges.

This paper explains the integration of Intel QuickAssist Technology on 4th Gen Intel Xeon Scalable processors that provides a powerful solution for HAProxy so it can address the growing needs of secure load balancing efficiencies and effectiveness.

Authors

Divya Pendyala

Intel Corporation

Joel Schuetze

Intel Corporation

Stanislas Odinot

Intel Corporation

Willy Tarreau

HAProxy Technologies

Emeric Brun

HAProxy Technologies

Industry Leading Load Balancer by HAProxy Technologies

HAProxy is generally recognized as the fastest and most widely used software load balancer¹ whose development is primarily backed by HAProxy Technologies. It is particularly suited for very high traffic web sites and powers a significant portion of the world's most visited sites. It offers high availability, load balancing, traffic offloading, and a broad spectrum of proxy-based features focused on optimizing application delivery and keeping applications responsive under heavy loads.

HAProxy has a robust design and has evolved significantly to meet the changing needs of modern applications while taking advantage of the evolution of operating systems and processors. Being an application layer reverse-proxy deployed at the edge, its role is to make routing decisions based on traffic analysis, which requires it to take care of data security and integrity. This represents heavy processing concentrated inside a single component, and this is made possible thanks to a highly optimized, extremely low-latency event-driven architecture that takes advantage of

systems with multi-core processors by distributing non-blocking operations to the least busy CPU core in a system and sharing as little as possible between them to keep caches hot. These careful design choices have resulted in the best possible combination of both low latency and high throughput, with an extreme robustness.

For reliability, the software features high observability of data processing, and has been developed with protections against software malfunction conditions such as impossible conditions, endless loops, etc. The software checks for these conditions and will gracefully restart providing information on the problem without corrupting data or leading to lengthy outages, which makes HAProxy an excellent solution for various mission-critical network security workloads.

HAProxy Technologies is a member of Intel® Network Builders program delivering high performance load balancing solutions on Intel platforms.

Load Balancing Use Cases that Benefit from 4th Gen Intel Xeon Scalable Processors and Intel QAT

The following use cases highlight how HAProxy delivers load balancing performance with the latest Intel Xeon Scalable processors:

- **Content Delivery Network/Web Server:** Minimize latency and accelerate asset delivery speed for content delivery networks (CDNs) and web servers. HAProxy on 4th Gen Intel Xeon Scalable processors with Intel QAT improves cryptography and data compression performance allowing each core to serve more clients so that CDNs can deliver content as quickly as possible. In addition, by significantly reducing the SSL processing overhead, the infrastructure better handles fast traffic variations resulting from world-wide events or denial-of-service (DoS) attacks.
- **Web Applications:** Applications running on 4th Gen Intel Xeon Scalable processors will handle the same load using fewer CPU cores thanks to the built-in Intel QAT accelerators that speed up cryptography, data

compression, and data movement. This results in the need for less and/or smaller containers or virtual machines for the same task, lowering operation costs.

- **Network Security Appliance:** Network security appliances must keep ahead of current trends so that customers can protect their investment. The high performance of 4th Gen Intel Xeon Scalable processors with its new instructions, faster DDR5 memory, PCIe Gen 5, built-in accelerators speeding up AI, encryption, and compression is a desirable solution for network security appliance designs that will be extensible in field for several years.

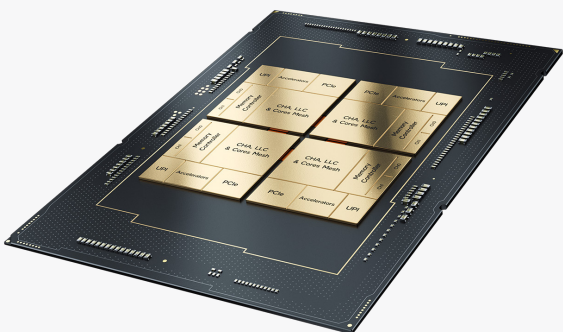
Performance Provided by 4th Gen Intel Xeon Scalable processor

The 4th Gen Intel Xeon Scalable processor is designed to accelerate performance across the most demanding workloads. The new processor has integrated the most built-in accelerators (see Figure 1)² of any CPU on the market to help maximize performance efficiency for emerging workloads, including those powered by AI. Advanced data security technologies help protect data in an ever-changing landscape of threats while unlocking new opportunities for business collaboration and insights. Together with its ecosystem of partners, Intel makes it easier for enterprises to stay competitive, offering the most choices to scale infrastructure and achieve business value quickly.

Network-optimized 4th Gen Intel Xeon processors are the next step in accelerating load balancing workloads while increasing energy efficiency, with a high-throughput, low-latency platform engineered for data centers, network core, and scalable to the edge for on-prem or cloud deployments.

Servers built with 4th Gen Intel Xeon Scalable processors further enhance the scalability and performance of software-defined infrastructure with an enhanced instruction set architecture (ISA). Intel® Ethernet 800 Series Network Adapters complement Intel architecture processors for load balancer deployments, with enhanced timing synchronization that helps prevent service disruptions. For cloud deployments, optimized acceleration software packages enhance the performance seamlessly.

Purpose Built for Accelerated Network Workloads



4th Gen Intel Xeon Scalable processors³ offer:

- **Advanced execution resources** in a range of core counts and feature sets, with improved per-core performance compared to the previous generation, enhanced by the most built-in accelerators in the industry.
- **Enhanced memory subsystem** with up to eight DDR5 channels operating at up to 4800 MT/s, a 1.5x improvement in memory bandwidth and speed compared to the predecessor platform.
- **Faster, higher capacity I/O** based on up to 80 lanes of PCIe 5.0 per socket, compared to 64 lanes of PCIe 4.0 per socket in the prior generation.

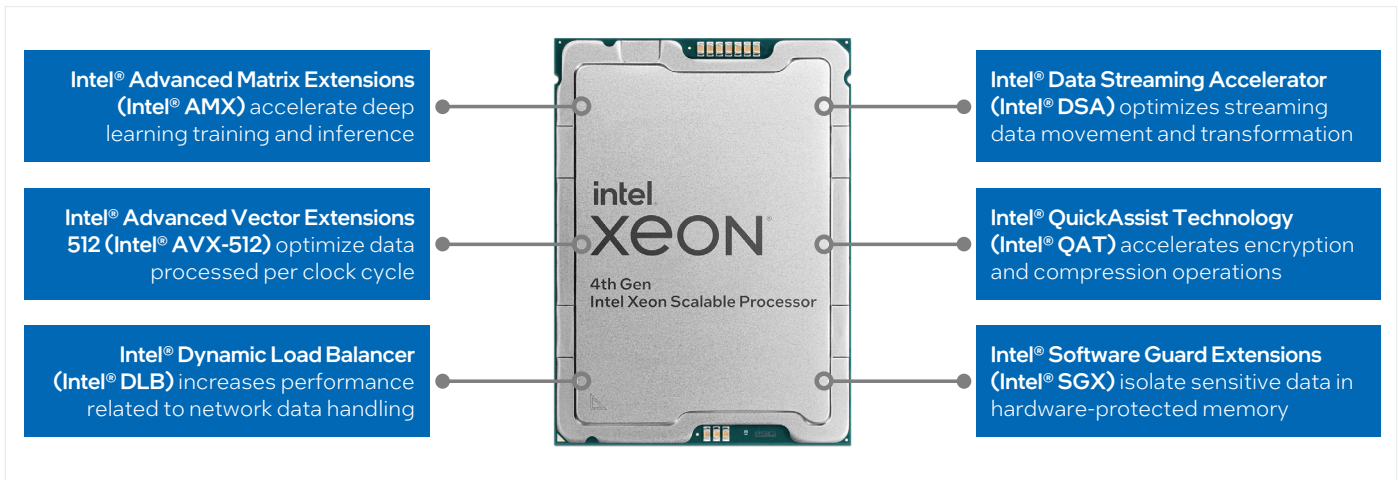


Figure 1. Key accelerators built into the 4th Gen Intel Xeon Scalable processor.

Built-in accelerators improve performance

Integration of accelerators into the processor redefines CPU architecture. Using accelerators provides a more efficient way to achieve higher performance than relying solely on increasing the CPU core count for workload processing.

With all-new accelerated matrix multiply operations, 4th Gen Intel Xeon Scalable processors have exceptional AI training and inference performance. Other seamlessly integrated accelerators speed up data movement and compression for faster networking, boost query throughput for more responsive analytics, and offload scheduling and queue management to dynamically balance loads across multiple cores. To enable new built-in accelerator features, Intel supports the ecosystem with OS-level software, libraries, and APIs. This architecture allows users to run cloud and networking workloads using fewer cores with faster cryptography.

Intel QuickAssist Technology (Intel QAT)

Intel QAT (see Figure 1), is one of the built-in accelerators and is designed for speeding up cryptography including private key protection, and data compression. It performs fixed-function acceleration for asymmetric and symmetric encryption, hashing, compression/decompression, and PKE. It also enables fusing of multiple operations to improve latency and memory bandwidth such as decompression, encryption and tagging data with hash.

Intel QAT speeds up these common application and infrastructure functions, while also freeing up a significant number of CPU cores. Because it offloads cryptography functions from the processor, the accelerator helps systems serve a larger number of clients.

Unlike previous generation Intel Xeon processors, Intel QAT acceleration comes as an integrated feature in specific SKUs of the 4th Gen Intel Xeon Scalable processor. This enables customers to use the acceleration technology for cryptographic workloads or compression / decompression with no additional physical components.

Intel® Advanced Vector Extensions 512 (Intel® AVX-512)

Intel® Advanced Vector Extensions 512 (Intel® AVX-512) helps accelerate the performance of scientific simulations, financial analytics, AI/deep learning, 3D modeling and analysis, image and audio/video processing, cryptography, data compression, and other intensive workloads. Intel AVX-512 is the latest Intel architecture processor vector instruction set, with up to two fused-multiply add (FMA) units and other optimizations to help accelerate the performance of demanding computational tasks.

These vectorized instructions used in software acceleration libraries, namely Intel® IPP Cryptography and Intel® Multi-Buffer Crypto for IPsec with Intel QAT Engine for OpenSSL, provide the computational power for HAProxy.

Intel® QuickAssist Technology Engine for OpenSSL* (Intel® QAT Engine for OpenSSL*)

Intel® QAT Engine for OpenSSL is a software package that supports acceleration for both hardware and optimized software based on vectorized instructions. The advancement in cryptographic acceleration provides users more options to accelerate their workloads. The **Intel® QAT Engine for OpenSSL*** supports the ability to accelerate the standard OpenSSL using basic Intel instruction set to either the hardware acceleration path (via the Intel® QAT hardware (qat_hw) path) or via the optimized software path (qat_sw lib).

HAProxy makes use of the Intel® QAT Engine for OpenSSL. Applications such as HAProxy interface to OpenSSL, a toolkit for TLS/SSL protocols that includes a modular system to plugin device-specific engines. With this implementation, HAProxy can be deployed on-prem or in the cloud for network security workloads using fewer cores with faster cryptography allowing each core to serve more clients.

HAProxy, using Intel® QAT crypto engine, reduces the number of cores needed for the software depending on the implementation used Intel® QAT hardware versus Intel® QAT software optimizations. For users, this means faster crypto processing and more cores available to run more applications.

Conclusion

HAProxy has incorporated features and capabilities introduced in 4th Gen Intel Xeon Scalable processors including built-in Intel QAT. The new capabilities deliver better performance using fewer cores for the same workload, reducing contention, and leaving more room for application layer processing and improving TCO. With the built-in accelerators available on 4th Gen Intel Xeon Scalable processors, users may get the benefit of higher performance and efficiency across network security workloads at lower costs.

Learn More

[HAProxy.com](#)

[4th Gen Intel Xeon Scalable](#)

[Intel QuickAssist Technology](#)

[Intel QuickAssist Technology software](#)

[Intel 800 Series Network Adapters](#)

[Intel Network Builders](#)

[GitHub for Intel QuickAssist Technology Engine](#)

[GitHub for Intel IPP Cryptography](#)

[GitHub for Intel Multi-Buffer Crypto for IPsec Library](#)



Notices & Disclaimers

¹<https://www.haproxy.com/blog/haproxyconf-2022-recap/>

<https://press.aboutamazon.com/2022/11/aws-announces-three-amazon-ec2-instances-powered-by-new-aws-designed-chips>

²<https://www.intel.com/content/www/us/en/products/details/processors/xeon/scalable.html>

³All 4th Gen Intel Xeon Scalable processor benchmarks are here: <https://edc.intel.com/content/www/us/en/products/performance/benchmarks/overview/-SPR>

Performance varies by use, configuration and other factors. Learn more on the [Performance Index](#) site.

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.

Your costs and results may vary.

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

© 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.