White Paper

Communication Service Providers (CoSPs), Telecommunications

Varnish Enterprise Shows Near-Line Rate Performance, Energy Efficiency on Intel® Xeon® 6

In tests with Intel® Xeon® 6 Processor-based servers, Varnish demonstrates up to 1.5Tbps performance with excellent power efficiency at up to 1.4 Gbps/Watt for live linear and video on demand (VoD) use cases

Throughout the decade starting in 2010, content delivery networks (CDNs) progressed through at least four maturity phases that added new features and increased performance. From a starting point as a static content cache, CDNs evolved to supporting dynamic and interactive content to adding cybersecurity protections to being the key workload for emerging edge networks.

It's this most recent phase, which hit in about 2020, that is driving a requirement for a big boost in performance to process the added workloads as well as increased CDN throughput. This performance was fueled by servers using the 2nd Gen Intel® Xeon® Processor family that offered up to 28 cores and up to 3.90 GHz processor base frequency.

Flash forward four years and it's time to refresh these 2020-era servers with the latest Intel processors – the Intel® Xeon® 6 Processor family. In addition to providing up to nearly line rate performance, these CPUs are much more power efficient to reduce energy and cooling costs and associated environmental impact. When refreshing from 2nd Gen Intel Xeon Scalable Processor-based servers, the Intel Xeon 6 Processor reduces a 200 rack server application to 72 rack servers and reduces power consumption by more than 1 Megawatt.[1]

To demonstrate just how much CDN performance and energy efficiency can be delivered with Intel Xeon 6 Processors, Intel teamed with Varnish Software, a leading CDN software provider and Intel® Partner Alliance Gold Tier community member, to reprise a performance test using servers based on the latest CPU technology.

Varnish Content Delivery Software

The testing described in this paper used the Varnish Enterprise Content Delivery software, which adapts the web cache and HTTP(S) features of Varnish Enterprise for deployment at the network edge. As shown in Figure 1, the benefits of deploying content delivery at the network edge include significantly reduced latency and reduced backhaul network data traffic to origin or mid-tier servers.

Varnish Enterprise is optimized for deployment on multi-access edge computing (MEC) servers or from a telco cloud data center, network aggregation site or wireless base stations. Varnish Enterprise can be deployed in virtualized, containerized, or bare metal environments. This deployment flexibility makes it an ideal solution for providing CDN services over 5G networks.

Varnish Enterprise is built on top of open source CDN frameworks with enterprise resiliency features and is enhanced with robust features for high performance and scalability. It allows companies to deliver low-latency, high bitrate content even during periods of peak demand. Varnish Enterprise provides capabilities for web and API acceleration, streaming, and private CDN deployments.

Table of Contents

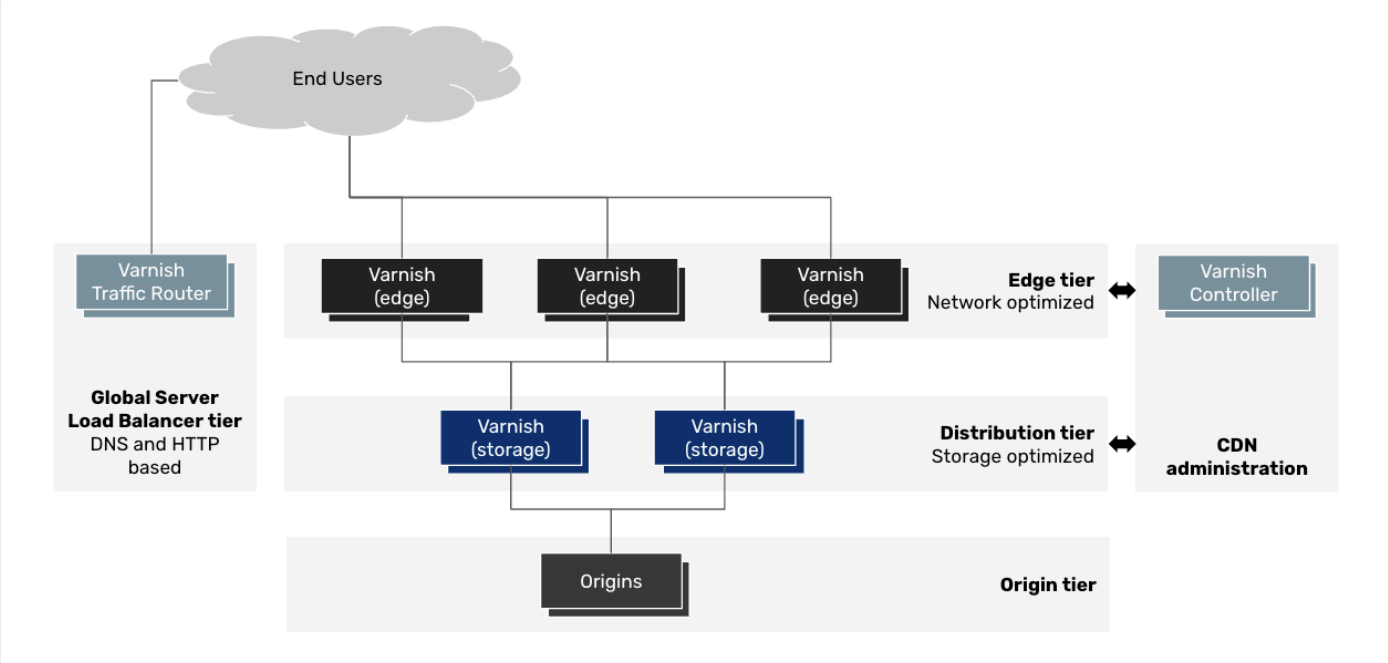

Figure 1. Varnish Enterprise deployment block diagram.

The software is also optimized for non-uniform memory access (NUMA) architectures. Historically, NUMA has been associated with multi-processor systems, but this is not its only use case. In modern processor architectures, both single and dual-processor systems may use NUMA, allowing the processor(s) and operating system to be aware of the relative distance between a CPU and memory or I/O.

For these tests, all Intel® Xeon® Scalable CPU-based systems are leveraging sub-NUMA clustering (SNC, for 2nd Gen Intel Xeon Scalable Processor-based systems) or Virtual NUMA (Intel Xeon 6 Processor-based systems) to split the CPU into two or four NUMA regions. This is beneficial as it allows the operating system to parallelize work so that this workload doesn't become bottlenecked by Linux page reclaim. By enabling SNC or Virtual NUMA, it's possible to reduce the impact of that Linux behavior, enabling higher total throughput.

Varnish has also incorporated an in-core TLS implementation that is built upon OpenSSL. This TLS implementation makes efficient use of standard OpenSSL APIs, and, when combined with the NUMA awareness features, VCL and threading model, results in more efficient caching software. When deployed on well-designed servers powered by Intel® Xeon® D Processors or Intel® Xeon® Processors, an end user can observe excellent performance and energy efficiency. The Intel Xeon 6 Processors provide larger caches, more memory and PCIe bandwidth, and more cores than prior generations.

Intel® Xeon® CPUs Enable Efficient and High-Performance CDNs

The Intel Xeon 6 Processor family represents a significant advancement in CPU performance and efficiency. Designed to meet the demands of modern workloads, including CDNs, these processors offer a compelling combination of features and benefits.

The Intel Xeon 6 Processor with Efficient-cores (E-cores) features a microarchitecture that delivers performance gains across a wide range of applications. Enhanced instruction set capabilities, coupled with high core count (up to 144 cores) and up to 3.2 GHz turbo clock frequencies, enable these processors to handle the demanding computational tasks associated with CDN operations, such as content delivery, and cache management.

Energy efficiency is a critical consideration for CDN operators. The Intel Xeon 6 Processors used for this testing incorporates E-cores to deliver high core density and exceptional performance per watt, offering distinct advantages for workloads that demand high throughput. Performance-cores (P-cores) excel at a wide range of workloads and can provide better performance for compute intensive workloads. Varnish Enterprise can extract exceptional performance and energy efficiency from the Intel Xeon 6 Processors with E-cores in part because Varnish is optimized for industry leading performance and efficiency with modern, multi-core, processors.

To address the specific needs of network-intensive workloads like CDNs, the Intel Xeon 6 Processor family integrates advanced accelerators. These hardware-based accelerators offload certain processing tasks from the CPU, freeing up valuable resources for other critical functions. By optimizing for efficient performance, these processors enable CDNs to handle increased traffic volumes, improve efficiency, and deliver a superior user experience.

Test Setup

To show performance and increased power efficiency, Intel and Varnish Software teamed up to test six servers using three models of Intel Xeon 6 Processors in single- and dual-processor servers (see Table 1 for details). The tests were also conducted using 2nd Gen Intel Xeon Scalable Processors to show the improvements when refreshing these servers that are reaching the end of their depreciation timeline.

The tests measured network throughput and energy consumption to show performance and energy efficiency improvements in the latest Intel® Xeon® Processor family. Each server was tested with the open source wrk HTTP(S) benchmarking tool, using 48, 100 and 800 connections per client.

The tests focused on two popular use cases for Varnish Enterprise content delivery software, a live linear service and a video on demand (VoD) service. In both cases, the clients requested randomly distributed 2MB files, with a target cache node hit rate of 98.8% for live linear and 100% for VoD.

Server	CPU(s)	PCIe Lanes & Gen used for CDN Networking & Storage	Networking	Storage	DRAM
1RU Supermicro SYS-1019P-WTR	1x Intel® Xeon® 6252N	40 Gen3	2x 100GbE Up to 200Gbps	2x PCIe Gen3x4 Intel® P4510 4TB	192GB (6x 32GB DDR4 3200 @ 2933 MT/s)
1RU Supermicro SYS-1029U-TN10RT	2x Intel® Xeon® 6252N	72 Gen3	2x 100GbE Up to 200Gbps	10x PCIe Gen3x4 Intel® P4510 4TB	384GB (12x 32GB DDR4 3200 @ 2933 MT/s)
2RU Supermicro SYS-212H-TN	1x Intel® Xeon® 6710E	64 Gen5	2x 2x200GbE Up to 800Gbps	8x PCIe Gen5x4 Kioxia CM7 1.92TB	512GB (8x 64GB DDR5 6400 @ 5600 MT/s)
	1x Intel® Xeon® 6766E				
	1x Intel® Xeon® 6780E				512GB (8x 64GB DDR5 6400 MT/s)
2RU Supermicro SYS-222H-TN	2x Intel® Xeon® 6710E	128 Gen5	4x 2x200GbE Up to 1600Gbps	16x PCIe Gen5x4 Kioxia CM7 1.92TB	512GB (16x 32GB DDR5 6400 @ 5600 MT/s)
	2x Intel® Xeon® 6766E				
	2x Intel® Xeon® 6780E				512GB (16x 32GB DDR5 6400)

Table 1. System under test (SUT) configurations.

Supermicro Rackmount Servers

The Supermicro servers used in this test were from the company's Rackmount Server Solution product family, which are designed as data center servers optimized for modern workloads including AI, high performance computing, cloud, storage and edge.

The rack mount products are available in 1U-high or 2U-high form factors with single or dual processors. All of the servers use Intel Xeon Processors for outstanding performance.

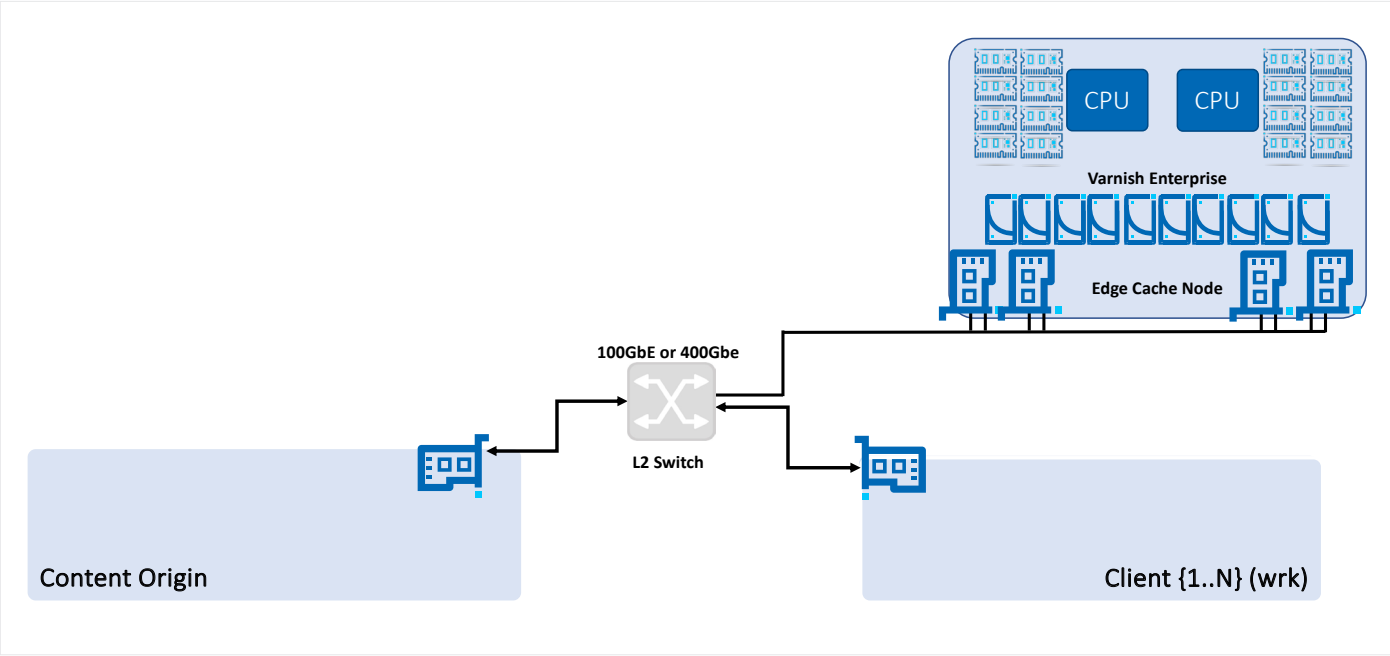

Figure 2. Varnish Enterprise CDN test configuration.

Figure 2 shows the network design for the tests with the origin server and clients connected to the Varnish Enterprise server via a non-blocking 100GbE or 400GbE layer 2 switch. Each server had either 100GbE or 200GbE network adapters that were bonded together to offer up to 1.6 Tbps (see Table 1).

In each of the SUT configurations, all memory channels were populated at the highest frequency supported, an optimal number and speed of network interface cards, and the optimal number of NVMe storage devices were installed in the server. For both CPU generations, the servers and PCIe devices were selected to be representative of what may have been deployed if that server was configured for CDN performance. Depending upon capacity and performance requirements, as well as deployment location constraints, an end user may prefer either a single or dual processor configuration.

In all tests, network throughput was measured at the system under test with dstat and power consumption was measured at the PSU inlets with a Raritan iPDU, and includes all power consumed by the server, from the processors to the cooling fans.

The systems were subjected to analogues of video on demand and live-linear workloads using the wrk benchmarking tool. In the live-linear tests, the NVMe drives were installed (and thus consuming power), but in an idle state as data was cached entirely in memory. The results show that in general, live linear tests deliver a higher data rate, at lower system power, because the data is kept in memory and does not need to be retrieved from the NVMe drives.

Live Linear Test Results

The first set of test results show the performance of the servers, organized by processor generation, in tests of the Varnish live linear capability with single processor servers (see Figure 3) and dual processor servers (see Figure 4).

In both server types, the Intel Xeon 6 Processors offered a significant improvement over the 2nd Gen Intel Xeon Scalable Processor-based systems. For single-processor servers, the Intel® Xeon® 6766E offered the best power efficiency with up to 1.57 Gbps / watt, whereas the highest throughput was provided by the Intel® Xeon® 6780E with up to 788 Gbps at 1.3 Gbps / watt.

In the dual-processor configurations, the Intel Xeon 6766E-based server still offered the best power efficiency with up to 1.53 Gbps / watt. The highest throughput was provided by the Intel Xeon 6780E Processors with up to 1,556 Gbps at 1.41 Gbps / watt, showing that dual-processor configurations can provide excellent energy efficiency for deployments where more performance or capacity is desired than can be delivered by a single-processor configuration.

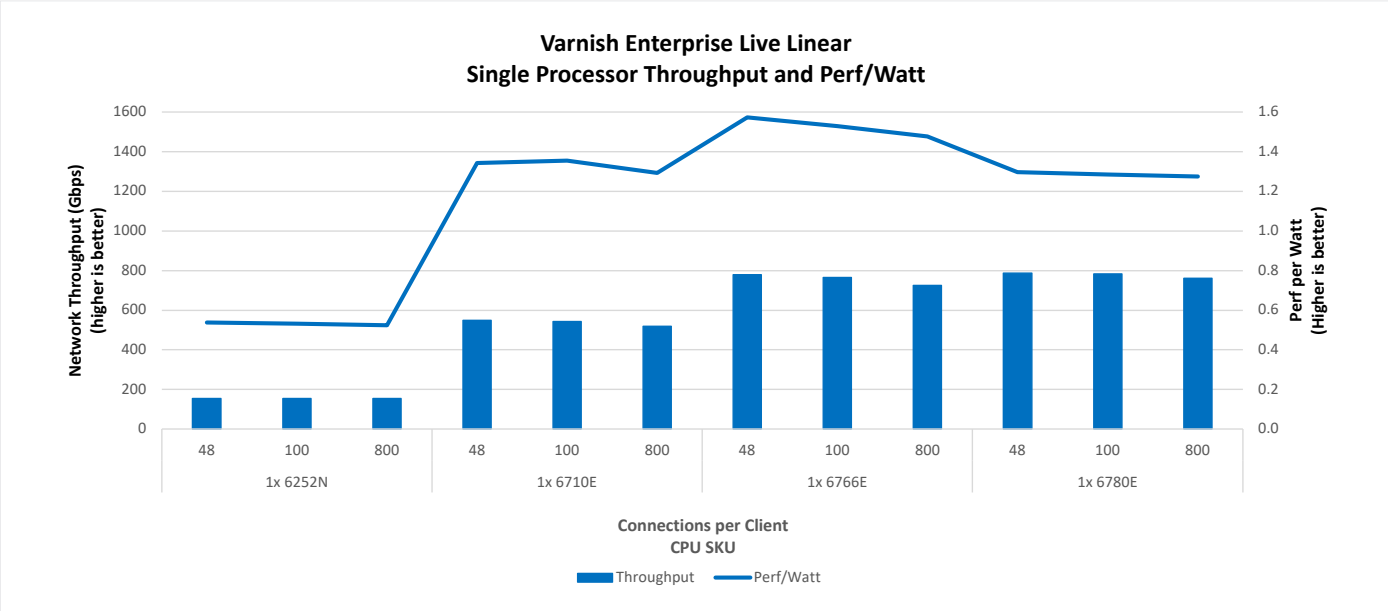

Figure 3. Live linear test results comparing single-processor servers using 2nd Gen Xeon Scalable Processors and Intel Xeon 6 Processors.

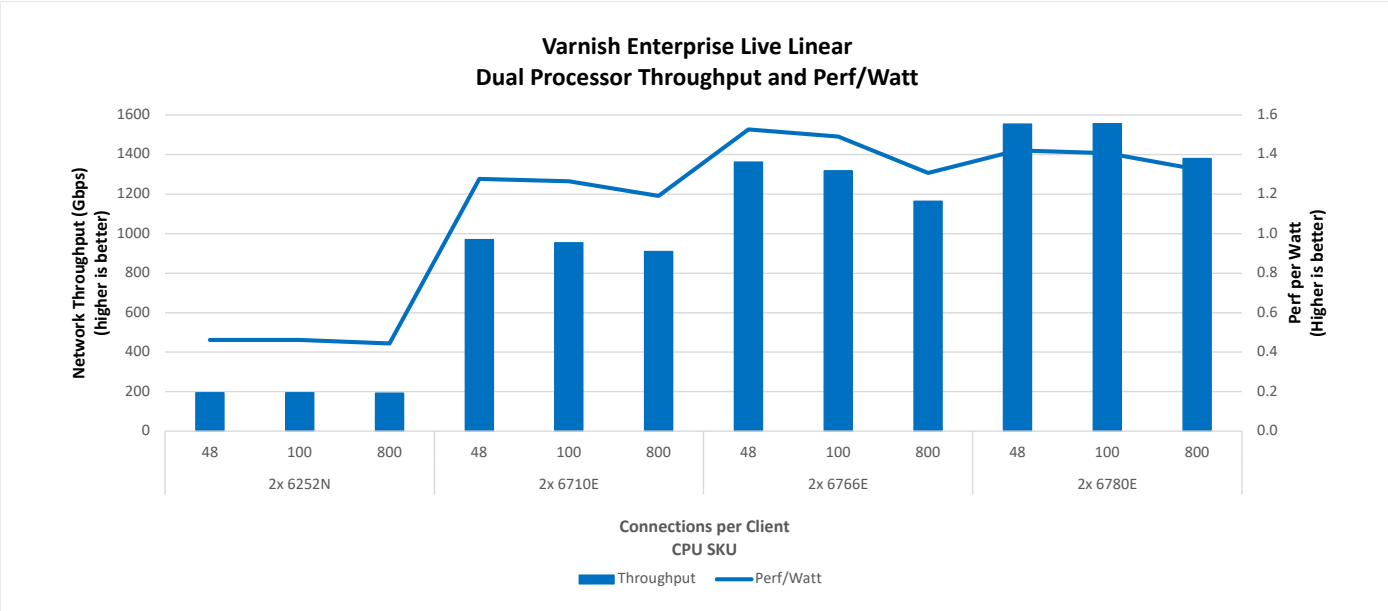

Figure 4. Live linear test results comparing dual-processor servers using 2nd Gen Xeon Scalable Processors and Intel Xeon 6 Processor.

VoD Test Results

The second set of test results show the performance of the servers for a VoD use case using both single-processor servers (see Figure 5) and the dual-processor servers (see Figure 6).

As in the live linear tests, the Intel Xeon 6 Processors offered significant improvement over the 2nd Gen Intel Xeon Scalable Processor-based systems. In the single processor-based servers, the Intel Xeon 6766E offered the best power efficiency with up to 0.93 Gbps / watt, whereas the highest throughput was provided by the Intel Xeon 6780E with up to 551 Gbps throughput at 0.89 Gbps / watt.

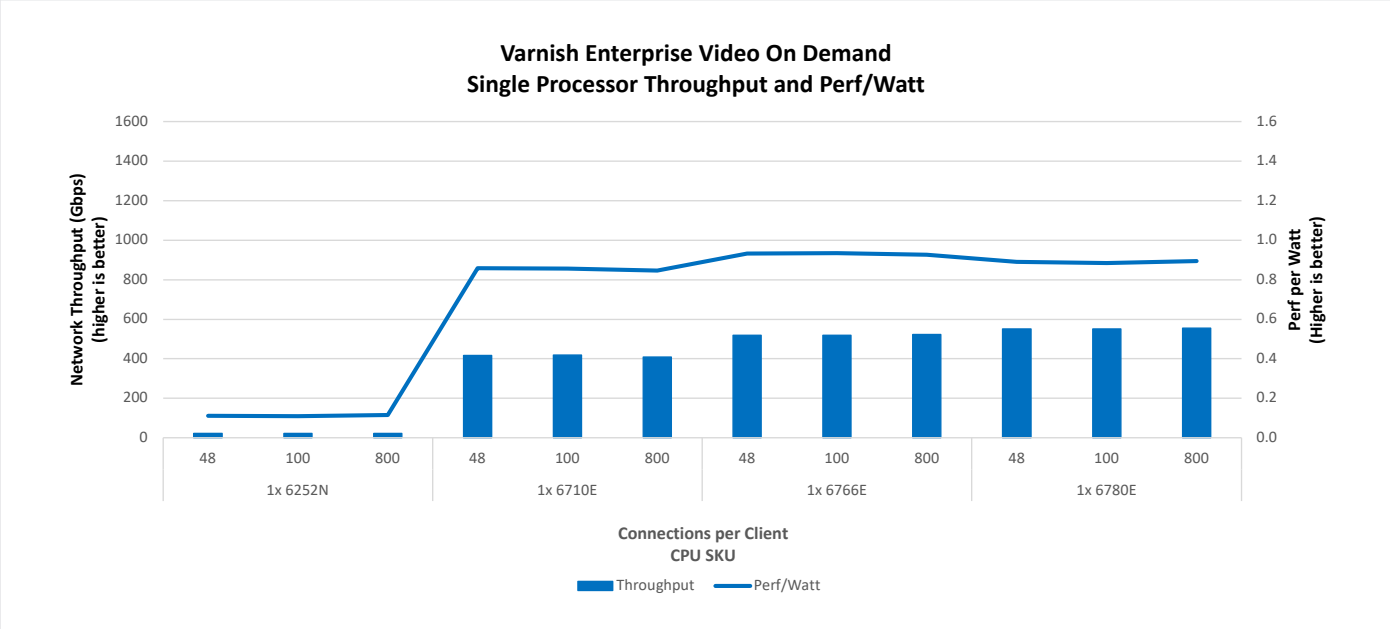

Figure 5. VoD test results comparing single-processor servers using 2nd Gen Xeon Scalable Processors and Intel Xeon 6 Processors.

In the dual-processor VoD tests, the Intel Xeon 6766E server offered the best power efficiency with up to 0.79 Gbps / watt. The highest throughput was provided by the dual Intel Xeon 6780E with up to 838 Gbps throughput at 0.78 Gbps / watt.

Table 2 also shows the results from all of the tests.

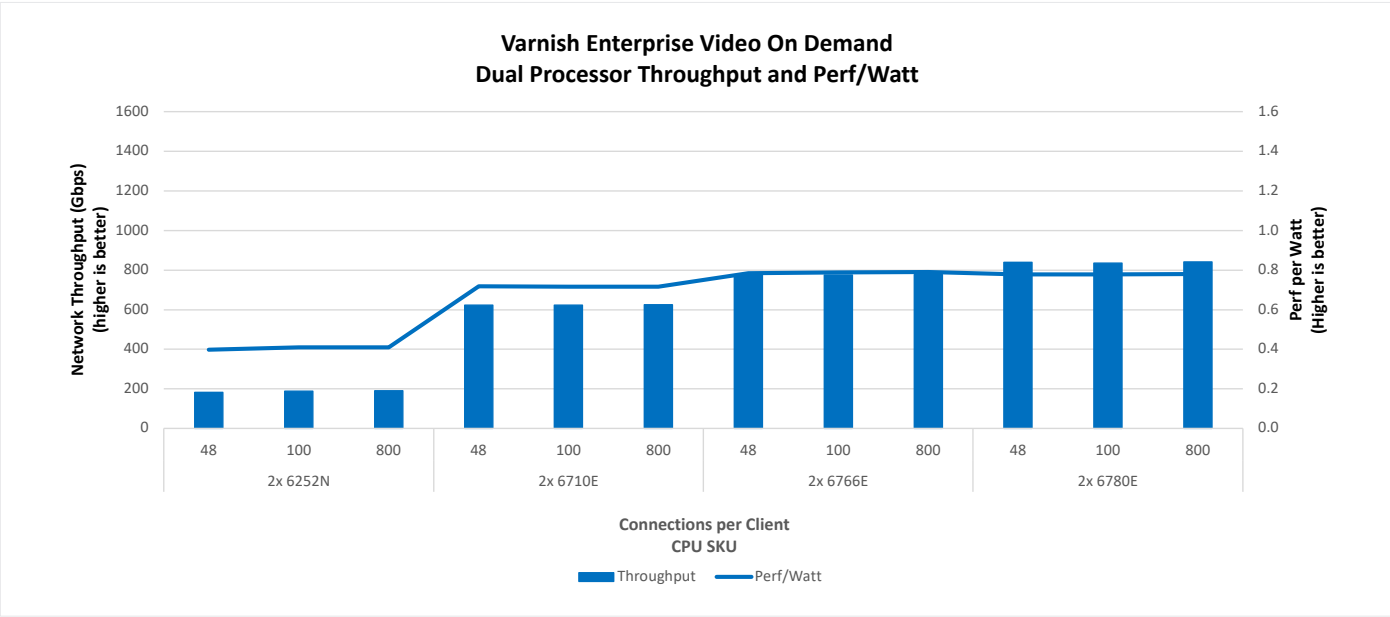

Figure 6. VoD test results comparing dual-processor servers using 2nd Gen Xeon Scalable Processors and Intel Xeon 6 Processors.

Server	CPU(s)	Connections / Client	Live Linear Network Throughput (Gbps)	Live Linear Gbps/W (System)	VoD Network Throughput (Gbps)	VoD Gbps/W (System)
1RU Supermicro SYS-1019P-WTR	1x Intel® Xeon® 6252N	48	154.18	0.54	21.96	0.11
		100	154.00	0.53	21.91	0.11
		800	154.02	0.52	22.04	0.12
1RU Supermicro SYS-1029U-TN10RT	2x Intel® Xeon® 6252N	48	195.64	0.46	182.91	0.40
		100	195.69	0.46	189.65	0.41
		800	191.87	0.44	191.27	0.41
2RU Supermicro SYS-212H-TN	1x Intel® Xeon® 6710E	48	549.84	1.34	416.91	0.86
		100	543.42	1.36	418.93	0.86
		800	518.13	1.29	409.44	0.85
	1x Intel® Xeon® 6766E	48	780.24	1.57	518.60	0.93
		100	765.76	1.53	519.02	0.93
		800	725.20	1.48	522.58	0.93
	1x Intel® Xeon® 6780E	48	786.47	1.30	551.09	0.89
		100	783.40	1.29	550.07	0.88
		800	761.64	1.28	555.27	0.89
2RU Supermicro SYS-222H-TN	2x Intel® Xeon® 6710E	48	969.62	1.28	622.64	0.72
		100	952.85	1.26	622.62	0.72
		800	909.73	1.19	625.49	0.72
	2x Intel® Xeon® 6766E	48	1361.40	1.53	775.13	0.78
		100	1317.64	1.49	774.64	0.79
		800	1163.91	1.31	778.36	0.79
	2x Intel® Xeon® 6780E	48	1553.62	1.42	838.93	0.78
		100	1556.67	1.41	835.53	0.78
		800	1380.82	1.33	841.75	0.78

Table 2. Summary of all configurations.

Conclusion

The test results show that Varnish Enterprise is able to leverage the capabilities of Intel Xeon 6 Processors with E-cores to simultaneously improve energy efficiency and performance when compared to previous generation deployments.

With a growing demand for rich content, constrained budgets, and increasing focus on sustainability, CDN operators need to extract more value than ever from their infrastructure. This requires excellent software that can make efficient use of the precious resources available in the server.

Learn More

Varnish home page

Varnish Enterprise

Intel® Xeon® 6 Processors

Intel® Partner Alliance

[1] https://www.intel.com/content/www/us/en/products/details/processors/xeon.html

Notices & Disclaimers

Single Intel® Xeon® 6252N: 1-node, Super Micro Superserver SYS-1019P-WTR, 1x Intel(R) Xeon(R) Gold 6252N CPU @ 2.30GHz, 24 cores, HT On, Turbo On, Total Memory 192GB (6x32GB DDR4 2933 MT/s [2933 MT/s]), BIOS 4.2, microcode 0x5003605, 2x Ethernet Connection X722 for 10GBASE-T, 4x MT28800 Family [ConnectX-5 Ex], 1x 477G INTEL SSDPEKNW512G8, 2x 3.7T INTEL SSDPE2KX040T8, Red Hat Enterprise Linux 8 (Ootpa), 6.6.16, Test by Intel as of 05/10/24. Software: Workload: wrk master 02/07/2021, varnish-plus-6.0.12r6 revision b9a3af68bd9c9ecf253a4c2aeda757ad6e30e71f, OpenSSL 1.1.1k FIPS 25 Mar 2021, keep alive, 800 connections per client, throughput measured with 100% Transport Layer Security (TLS) traffic with 98.8% (Live Linear) and 100%(Video-On-Demand) target cache hit ratio.

Dual Intel® Xeon® 6252N: 1-node, Super Micro Superserver SYS-1029U-TN10RT, 2x Intel(R) Xeon(R) Gold 6252N CPU @ 2.30GHz, 24 cores, HT On, Turbo On, Total Memory 384GB (12x32GB DDR4 2933 MT/s [2933 MT/s]), BIOS 4.2, microcode 0x5003605, 2x Ethernet Controller 10-Gigabit X540-AT2, 4x MT28800 Family [ConnectX-5 Ex], 1x 477G INTEL SSDPEKNW512G8, 10x 3.7T INTEL SSDPE2KX040T8, Red Hat Enterprise Linux 8 (Ootpa), 6.6.16, Test by Intel as of 07/18/24. Software: Workload: wrk master 02/07/2021, varnish-plus-6.0.12r6 revision b9a3af68bd9c9ecf253a4c2aeda757ad6e30e71f, OpenSSL 1.1.1k FIPS 25 Mar 2021, keep alive, 800 connections per client, throughput measured with 100% Transport Layer Security (TLS) traffic with 98.8% (Live Linear) and 100% (Video-On-Demand) target cache hit ratio.

Single Intel® Xeon® 6710E: 1-node, Super Micro Superserver Hyper SYS-212H-TN, 1x Intel(R) Xeon(R) 6710E, 64 cores, HT NA, Turbo On, Total Memory 512GB (8x64GB DDR5 6400 MT/s [5600 MT/s]), BIOS T20240403150118, microcode 0x13000131, 4x MT2910 Family [ConnectX-7], 1x 894.3G Micron_7450_MTFDKBG960TFR, 8x 1.8T KIOXIA KCMY1RUG1T92, Red Hat Enterprise Linux 8 (Ootpa), 6.6.16, Test by Intel as of 05/10/24. Software: Workload: wrk master 02/07/2021, varnish-plus-6.0.12r6 revision b9a3af68bd9c9ecf253a4c2aeda757ad6e30e71f, OpenSSL 1.1.1k FIPS 25 Mar 2021, keep alive, 800 connections per client, throughput measured with 100% Transport Layer Security (TLS) traffic with 98.8%(Live Linear) and 100% (Video-On-Demand) target cache hit ratio.

Dual Intel® Xeon® 6710E: 1-node, Super Micro Superserver Hyper SYS-222H-TN2x Intel(R) Xeon(R) 6710E, 64 cores, HT NA, Turbo On, Total Memory 512GB (16x32GB DDR5 6400 MT/s [5600 MT/s]), BIOS 1.0, microcode 0x13000132, 8x MT2910 Family [ConnectX-7], 1x 465.8G Samsung SSD 970 EVO Plus 500GB, 16x 1.8T KIOXIA KCMY1RUG1T92, Red Hat Enterprise Linux 8 (Ootpa), 6.6.16, Test by Intel as of 07/12/24. Software: Workload: wrk master 02/07/2021, varnish-plus-6.0.12r6 revision b9a3af68bd9c9ecf253a4c2aeda757ad6e30e71f, OpenSSL 1.1.1k FIPS 25 Mar 2021, keep alive, 800 connections per client, throughput measured with 100% Transport Layer Security (TLS) traffic with 98.8%(Live Linear) and 100% (Video-On-Demand) target cache hit ratio.

Single Intel® Xeon® 6766E: 1-node, Super Micro Superserver Hyper SYS-212H-TN, 1x Intel(R) Xeon(R) 6766E, 144 cores, HT NA, Turbo On, Total Memory 512GB (8x64GB DDR5 6400 MT/s [6400 MT/s]), BIOS T20240403150118, microcode 0x13000131, 4x MT2910 Family [ConnectX-7], 1x 894.3G Micron_7450_MTFDKBG960TFR, 8x 1.8T KIOXIA KCMY1RUG1T92, Red Hat Enterprise Linux 8 (Ootpa), 6.6.16, Test by Intel as of 07/12/24. Software: Workload: wrk master 02/07/2021, varnish-plus-6.0.12r6 revision b9a3af68bd9c9ecf253a4c2aeda757ad6e30e71f, OpenSSL 1.1.1k FIPS 25 Mar 2021, keep alive, 800 connections per client, throughput measured with 100% Transport Layer Security (TLS) traffic with 98.8%(Live Linear) and 100%(Video-On-Demand) target cache hit ratio.

Dual Intel® Xeon® 6766E: 1-node, Super Micro Superserver Hyper SYS-222H-TN2x Intel(R) Xeon(R) 6766E, 144 cores, HT NA, Turbo On, Total Memory 512GB (16x32GB DDR5 6400 MT/s [5600 MT/s]), BIOS 1.0, microcode 0x13000132, 8x MT2910 Family [ConnectX-7], 1x 465.8G Samsung SSD 970 EVO Plus 500GB, 16x 1.8T KIOXIA KCMY1RUG1T92, Red Hat Enterprise Linux 8 (Ootpa), 6.6.16, Test by Intel as of 07/03/24. Software: Workload: wrk master 02/07/2021, varnish-plus-6.0.12r6 revision b9a3af68bd9c9ecf253a4c2aeda757ad6e30e71f, OpenSSL 1.1.1k FIPS 25 Mar 2021, keep alive, 800 connections per client, throughput measured with 100% Transport Layer Security (TLS) traffic with 98.8%(Live Linear) and 100% (Video-On-Demand) target cache hit ratio.

Single Intel® Xeon® 6780E: 1-node, Super Micro Superserver Hyper SYS-212H-TN, 1x Intel(R) Xeon(R) 6780E, 144 cores, HT NA, Turbo On, Total Memory 512GB (8x64GB DDR5 6400 MT/s [6400 MT/s]), BIOS T20240403150118, microcode 0x13000131, 4x MT2910 Family [ConnectX-7], 1x 894.3G Micron_7450_MTFDKBG960TFR, 8x 1.8T KIOXIA KCMY1RUG1T92, Red Hat Enterprise Linux 8 (Ootpa), 6.6.16, Test by Intel as of 05/9/24. Software: Workload: wrk master 02/07/2021, varnish-plus-6.0.12r6 revision b9a3af68bd9c9ecf253a4c2aeda757ad6e30e71f, OpenSSL 1.1.1k FIPS 25 Mar 2021, keep alive, 800 connections per client, throughput measured with 100% Transport Layer Security (TLS) traffic with 98.8%(Live Linear) and 100%(Video-On-Demand) target cache hit ratio.

Dual Intel® Xeon® 6780E: 1-node, Super Micro Superserver Hyper SYS-222H-TN2x Intel(R) Xeon(R) 6780E, 144 cores, HT NA, Turbo On, Total Memory 512GB (16x32GB DDR5 6400 MT/s [5600 MT/s]), BIOS 1.0, microcode 0x13000132, 8x MT2910 Family [ConnectX-7], 1x 465.8G Samsung SSD 970 EVO Plus 500GB, 16x 1.8T KIOXIA KCMY1RUG1T92, Red Hat Enterprise Linux 8 (Ootpa), 6.6.16, Test by Intel as of 05/26/24. Software: Workload: wrk master 02/07/2021, varnish-plus-6.0.12r6 revision b9a3af68bd9c9ecf253a4c2aeda757ad6e30e71f, OpenSSL 1.1.1k FIPS 25 Mar 2021, keep alive, 800 connections per client, throughput measured with 100% Transport Layer Security (TLS) traffic with 98.8%(Live Linear) and 100% (Video-On-Demand) target cache hit ratio.

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. No product or component can be absolutely secure.

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

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

No product or component can be absolutely secure.

Your costs and results may vary.

0924/NK/H09/PDF ♻Please Recycle 361991-001US