Product Brief

Intel® Storage Engines
4th Gen Intel® Xeon® Scalable Processors

New Intel® Xeon® Platform Includes Built-In Accelerators for Encryption, Compression, and Data Movement

"When a CPU can offload storage functions to built-in accelerators, it frees up cores for business-critical workloads—tasks like supporting applications running on hyperconverged infrastructure and virtual machines and containers or spinning up more cloud services for customers. With 4th Gen Intel® Xeon® Scalable processors and Intel® Storage Engines, you can support storage at scale with cores to spare."

—Ronak Singhal, Intel Senior Fellow

4th Gen Intel® Xeon® Scalable processors have built-in accelerators that allow the CPU to offload key storage tasks like encryption and compression plus data movement and transformation. These Intel® Storage Engines improve performance for specific workloads and free up CPU cores for business-critical applications.

Storage workloads are critical and growing exponentially

Data is arguably one of the most important assets of any business—if you can capture, analyze, and act on it. The more data you collect, the more potential insights you can extract. Of course, collecting more data means building out more storage, more bandwidth, and more processing power to encrypt, compress, move, and transform it. As data warehouses grow into the petabytes and beyond, data storage and management are consuming more and more computing power.

Intel Storage Engines on 4th Gen Intel Xeon Scalable processors bring the offload performance of external accelerators onto the processor itself. Building these accelerators into the CPU sidesteps the bottlenecks and latencies that occur when a system shuttles data to external devices while reducing overall complexity and costs.

Storage, security, and data movement put constant demands on CPUs

Storing, moving, and transforming data are never-ending tasks for data centers, web servers, and networking functions. Offloading specific tasks to dedicated accelerators is a proven strategy for increasing overall capacity and performance.

- **Compression/decompression:** Compressing data so that it takes up less storage space and requires less bandwidth is a fundamental performance and capacity strategy. Today, compression/decompression is an intensive task for data center storage, enterprise databases, content delivery networks, and networking.
- **Encryption/decryption:** Protecting data at rest, in motion, and during processing requires encryption at every step. Files must be encrypted before they are written to disk and decrypted when accessed. Servers must encrypt and authenticate every read and write of data to storage. Encrypting and hashing are nonstop, processor-intensive tasks.
- **Data movement:** Moving data from storage to memory, memory to CPU, and node to node around the network are major workloads. As organizations seek to analyze and derive insights from more and more data, data sets have become larger and larger, increasing demand for processing, storage capacity, and network bandwidth.
- **In-memory operations:** Running enterprise databases in-memory has become standard procedure for near-real-time workloads such as recommendation engines, credit checks, and fraud detection. As memory arrays have grown into the terabytes and beyond, data movement between memory and the CPU has become a bottleneck of its own.

Intel® Storage Engines: Built-in accelerators for storage-specific workloads

Integrating workload accelerators into the CPU has three major benefits. First, built-in accelerators sidestep the I/O bottlenecks and latency inherent with drop-in accelerator cards and external appliances. Second, they process their specific workloads faster than a CPU alone. Third, they allow the CPU to offload tasks and preserve headroom for the workloads that need higher performance computing resources.

4th Gen Intel Xeon Scalable processors can offload the bulk of compression, encryption, and data transport to Intel Storage Engines that take performance beyond the capabilities of the CPU alone. Implementation of these accelerators is largely built into the processor's instruction set architecture (ISA). Shifting workloads involves making relatively simple code changes, enabling drivers, and configuring BIOS settings.

Enhanced storage performance with 4th Gen Intel® Xeon® Scalable processors

Fewer cores, faster compression

Intel® QuickAssist Technology (Intel® QAT) ›

Up to

95% fewer cores

and up to

2x higher throughput

for Level 1 compression[1]

More-efficient read/write

Intel® Data Streaming Accelerator (Intel® DSA) ›

Up to

1.21x higher IOPS

and up to

18% latency reduction

for small packet random read vs. Intel® Intelligent Storage Acceleration Library[2]

Faster databases

Intel® In-Memory Analytics Accelerator (Intel® IAA) ›

Up to

3x higher RocksDB

performance[3]

1. See [N16] at intel.com/processorclaims: 4th Gen Intel® Xeon® Scalable processors. Results may vary.
2. See [N18] at intel.com/processorclaims: 4th Gen Intel® Xeon® Scalable processors. Results may vary.
3. See [D1] at intel.com/processorclaims: 4th Gen Intel® Xeon® Scalable processors. Results may vary.

Intel® QuickAssist Technology: Encryption and compression offload engine

Previously available as an external accelerator, Intel QuickAssist Technology (Intel® QAT) is now a built-in accelerator on the 4th Gen Intel Xeon Scalable processor, reducing latency versus prior generations. The CPU offloads encryption, compression, and public key exchange workloads to Intel QAT, freeing up clock cycles for other workloads. On-the-fly compression shrinks storage footprints and reduces network traffic, while fast encryption increases throughput for web connections and API calls.

Intel® Volume Management Device: Integrated NVMe device control

Intel Volume Management Device (Intel® VMD) brings NVMe device control into the CPU. It is a mainstay of Intel® Xeon® Scalable processors that's widely used in hyperconverged infrastructure to consistently manage NVMe storage devices. Intel VMD supports Intel® Virtual RAID on CPU (Intel® VROC), plus hot plug, surprise removal, and blink status LED for NVMe devices. Using Intel VMD to manage SSD storage can improve reliability, availability, and serviceability.

Intel® Data Streaming Accelerator: Faster data movement within and beyond the CPU

Intel Data Streaming Accelerator (Intel® DSA) is our next-generation direct memory access (DMA) engine. It delivers significant throughput improvements by accelerating data movement and transformation operations, such as data integrity checks and deduplication. Intel DSA works on the CPU—between memory, caches, and processor cores—and beyond the CPU to attached memory, storage, and networked resources. The impact on performance yields high efficiency for I/Os, data transfers, and packet processing.

Intel® In-Memory Analytics Accelerator: Compression for in-memory databases

In-memory databases are fundamental for insight-driven transactions and decision-making. However, moving large data sets from memory to CPU and back takes time and processing power, even when held completely in memory or persistent memory. The Intel® In-Memory Analytics Accelerator (Intel® IAA) delivers quick data-swap times and workloads by compressing and decompressing data as it moves between the processor, memory, and persistent memory. Shrinking the in-memory footprints means shorter data transfer times and less work for the CPU. Plus, the accelerator handles compression/decompression, so these operations don't place extra demand on CPU cycles.

Intel® Storage Engines – Performance beyond the CPU

Workload	Accelerator
Compression/decompression: Reduces storage footprints, shrinks files before transport	**Intel® QuickAssist Technology:** Accelerates on-the-fly compression, frees up CPU cores
Encryption/decryption: Ubiquitous, CPU-intensive task that secures data at rest and in motion	**Intel® QuickAssist Technology:** Accelerates encryption and authentication, frees CPU cores
Data movement/transformation: Common storage functions and operations such as data integrity checks and deduplication	**Intel® Data Streaming Accelerator:** Moves more data faster[2] through the data path, accelerating reads, writes, and replication, while freeing CPU cores
In-memory operations: Large-scale data movement between memory and the CPU	**Intel® In-Memory Analytics Accelerator:** Compresses and decompresses data as it moves between the processor, memory, and persistent memory, increasing throughput[3]

Conclusion: Intel Storage Engines amplify performance and value

Building storage accelerators into 4th Gen Intel Xeon Scalable processors adds significant performance improvements and capabilities that create business value. Everyone from system builders, enterprises, and cloud service providers to the architects and developers who build, use, and optimize data-intensive services will benefit.

Learn more

See how built-in accelerators can help improve the performance of your fastest-growing workloads at intel.com/4thgenxeon.

Explore how to get the most out of Intel Xeon Scalable processors with built-in accelerators at intel.com/xeonscalable.

Learn more about Intel Storage Engines

Intel QuickAssist Technology ›

Intel Volume Management Device ›

Intel Data Streaming Accelerator ›

Intel In-Memory Analytics Accelerator Architecture Specification ›

What's the difference?

Intel® In-Memory Analytics Accelerator vs. Intel® QuickAssist Technology

In 4th Gen Intel® Xeon® Scalable processors, both Intel® IAA and Intel® QAT offload compression workloads from the processor cores. However, the types of data compression are different for each accelerator.

- Intel IAA processes compression/decompression for data in-memory to help speed data movement between the CPU and memory. It is suited for columnar compression and database queries and can provide dramatic performance gains for in-memory computing.
- Intel QAT processes bulk compression/decompression for data in storage and transit. It also processes encryption workloads, which makes it ideal for accelerating chained decompress-encrypt-decrypt-compress operations using standards such as LZ4 lossless compression and OpenSSL.

In general, Intel QAT is for local, cloud, and hybrid storage, whereas Intel IAA is for in-memory computing.

intel XEON

1. Up to 95 percent fewer cores and up to 2x higher throughput for Level 1 compression. See [N16] at intel.com/processorclaims: 4th Gen Intel® Xeon® Scalable processors. Results may vary.
2. Up to 1.6x higher IOPs and up to 37 percent latency reduction for large packet sequential read. See [N18] at intel.com/processorclaims: 4th Gen Intel® Xeon® Scalable processors. Results may vary.
3. Up to 3x higher RocksDB performance. See [D1] at intel.com/processorclaims: 4th Gen Intel® Xeon® Scalable processors. Results may vary.

0123/TT/CMD/PDF