

Business Brief

4th Gen Intel® Xeon® Scalable Processor
Cloud Computing



Clear Skies for Your Cloud Growth



With Intel you can activate mature and agile hybrid, multi-cloud or intelligent edge strategies.

As cloud migration continues to develop enterprise-wide, the landscape of cloud service environments is becoming more complex. Synchronized business and technology strategies driven by industry best practices can optimize the agility, cost-efficiency and other benefits available from cloud-first operations. Organizations need to improve efficiency and make service levels more predictable, while enabling faster innovation and better-supported business decisions.

It's now more critical than ever for technology to deliver business value as organizations look to scale, drive down costs and deliver new services. Instead of customizing systems for new applications, which adds complexity, enterprises can achieve the performance needed to meet a wide variety of deployments — both today and in the future — with a scalable platform.



Accelerate your cloud with Intel technologies

4th Gen Intel® Xeon® Scalable processors have the most built-in accelerators of any CPU on the market to deliver performance and power efficiency advantages across the fastest-growing cloud workloads, including AI, analytics, networking, storage and HPC. With all-new Intel Advanced Matrix Extensions (Intel AMX), 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 in cloud platforms. To enable new built-in accelerator features within a hyper-scaled environment, Intel supports the ecosystem with all of the most common cloud APIs, libraries and OS-level software.

With built-in accelerators and software optimizations, previous-generation Intel Xeon Scalable processors have been shown to deliver leading performance per watt on targeted real-world cloud workloads.³ This results in more efficient CPU utilization, lower cloud electricity consumption and higher services ROI, while helping businesses achieve their sustainability goals.

PERFORMANCE PROOFPOINT

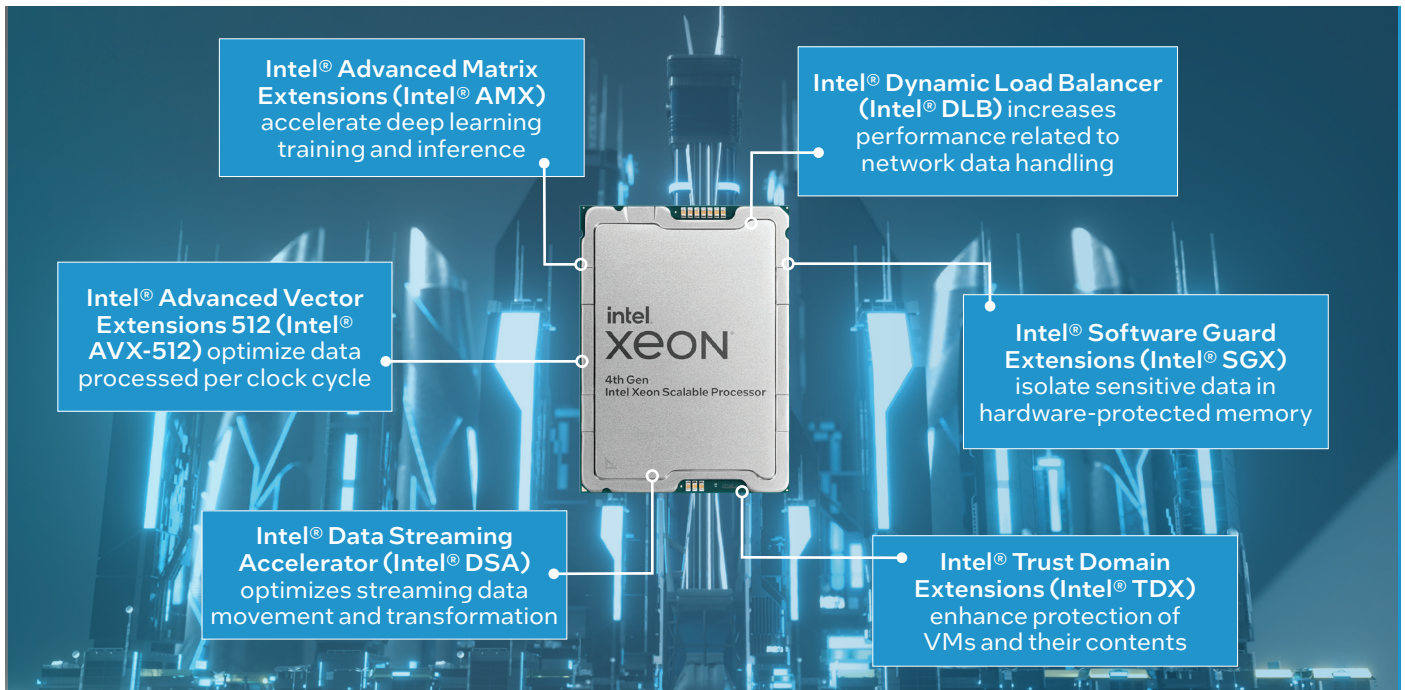
UP TO HIGHER 2.5X THROUGHPUT (RPS)

UP TO LOWER P99 74% LATENCY

UP TO LOWER CPU 12% UTILIZATION

on 4th Gen Intel Xeon Platinum 8480+ processors with two Intel QAT devices compared to solution with no acceleration⁴





Key cloud workloads

Confidential computing

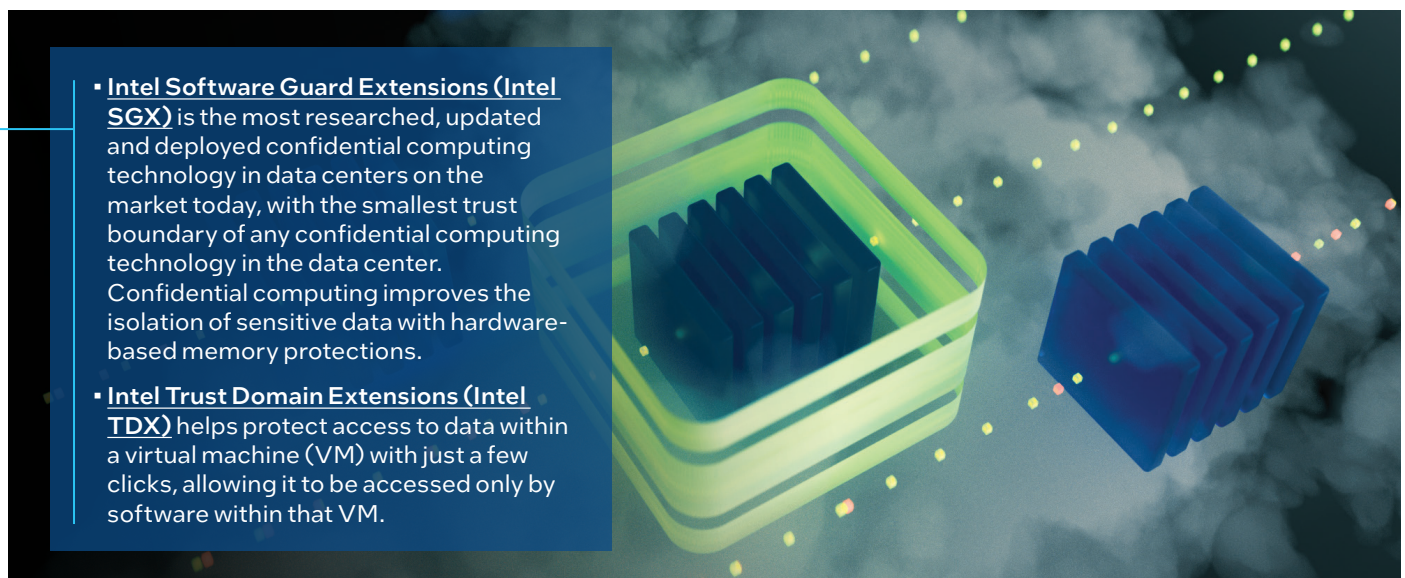
Organizations are focusing on rigorous security, identity and compliance management as part of their zero-trust security strategies. Confidential computing improves the isolation of sensitive data with hardware-based memory protections. The confidential computing market is projected to grow at a CAGR of 90%-95% to reach \$54 billion in 2026.⁵

THE NEED:

Sharing data among third parties is often made difficult or impossible because of privacy, regulatory or intellectual property considerations. Overcoming those obstacles provides potential benefits such as enabling research or training AI models with massive distributed datasets, taking advantage of cloud cost, scalability and agility benefits.

THE ANSWER:

Confidential computing powered by 4th Gen Intel Xeon Scalable processors improves the isolation of data by enabling only trusted code to access it, within hardware-protected memory enclaves. These privacy protections enable innovation around regulated workloads in hyper-scaled platforms and other distributed networks.



Microservices

Microservices adoption continues to be a major contributor to transformation within development organizations, especially for hyper-scaled platforms and intelligent edge environments. A recent study conducted by Intel revealed that more than 83% of all new cloud-native applications, along with Software as a Service (SaaS) solutions, are using microservices.⁶

THE NEED:

Businesses need to operate seamlessly across multi-cloud environments, securing their apps and working faster and better at scale. Cloud hardware and software must be optimized for throughput to support key microservices challenges such as quality of service, infrastructure overhead and observability.

PERFORMANCE PROOFPOINT

UP TO HIGHER
1.88X THROUGHPUT

with 4th Gen Intel Xeon Scalable processors versus prior generation on DeathStarBench — social network microservices workload (read home timeline) with 100ms SLA⁷

THE ANSWER:

Accelerators built into 4th Gen Intel Xeon Scalable processors speed up data movement through and within cloud platforms, safeguarding operational targets for quality of service, infrastructure overhead and data/process observability. They help ensure fast spin-up and shut-down of cloud business services, hardware-assisted queuing to help load-balance millions of incoming operational requests and high-speed distributed network communications.

- **Intel Data Streaming Accelerator (Intel DSA)** drives high performance for storage, networking and data-intensive workloads by improving streaming data movement and transformation operations. Designed to offload the most common data movement tasks that cause overhead in data center-scale deployments, Intel DSA helps speed up data movement across the CPU, memory and caches, as well as all attached memory, storage and network devices.
- **Intel Dynamic Load Balancer (Intel DLB)** improves the system performance related to handling network data on multi-core Intel Xeon Scalable processors. Intel DLB enables the efficient distribution of network processing across multiple CPU cores/threads and dynamically distributes network data across multiple CPU cores for processing as the system load varies. Intel DLB also restores the order of networking data packets processed simultaneously on CPU cores.

PERFORMANCE PROOFPOINT

UP TO LOWER
96% LATENCY

at the same throughput (RPS) with 4th Gen Intel Xeon Scalable processors and Intel DLB versus software for Istio ingress gateway working on six cores (12 threads)⁸

PERFORMANCE PROOFPOINT

UP TO **2.33X** HIGHER SINGLE-CORE THROUGHPUT
[MTU 1500, 1500B]

AND

UP TO **2.63X** HIGHER SINGLE-CORE THROUGHPUT
[MTU 9000, 2400B]

on 4th Gen Intel Xeon Scalable processor with Intel DSA memory copy compared with software memory copy⁹

AI and advanced data analytics

Embracing the convergence of AI and analytics helps businesses gain visibility into their data that drives operational insights and data-driven strategic decision making. 84% of business executives believe they need to use AI to achieve their growth objectives but 76% acknowledge they are struggling with how to scale AI across their business and services.¹⁰

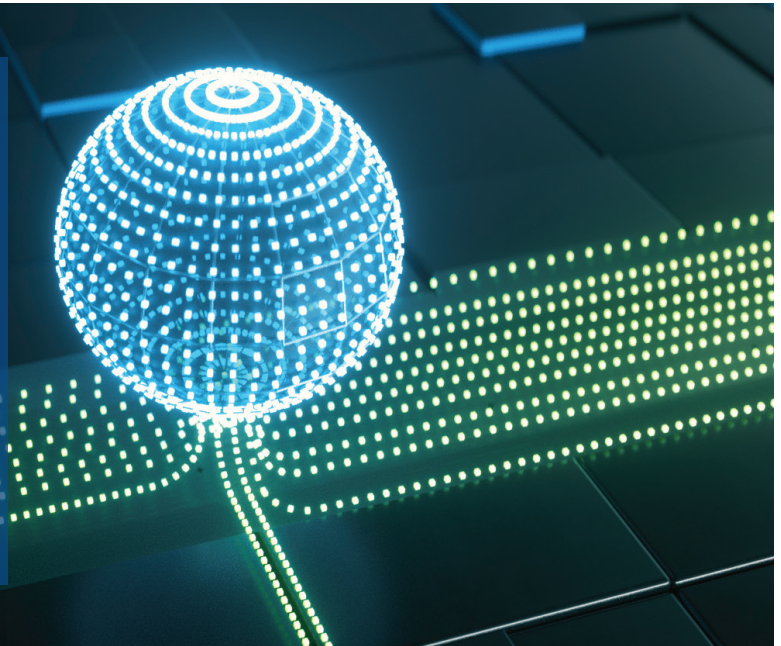
THE NEED:

As businesses of all types and sizes increase their use of AI and analytics, they must optimize performance and other operational outcomes with positive ROI.

THE ANSWER:

4th Gen Intel Xeon processors give businesses the ability to drive greater insight and value from their data, using AI for data-driven decision making. Financial services can make better investment and risk management decisions. Healthcare service organizations can help improve patient care, reduce medical costs through more efficient billing and pre-approvals and create more accurate predictions of post-surgery complications. Online and retail establishments will gain unique understanding and linkages into specific customers' needs and grow revenue through better demand forecasting, proactive and personalized marketing and real-time logistics and inventory management.

- **Intel Advanced Matrix Extensions (Intel AMX)** accelerates AI capabilities on 4th Gen Intel Xeon Scalable processors, speeding up training and inference without additional hardware. This accelerator is ideal for workloads such as natural language processing, recommendation systems and image recognition.
- **Intel Advanced Vector Extensions 512 (Intel AVX-512)** accelerates performance for intensive workloads. It is the latest x86 vector instruction set, with up to two fused multiply-add (FMA) units and other optimizations to accelerate performance for the most demanding computational tasks.



PERFORMANCE PROOFPOINT

5.7X TO 10X HIGHER PYTORCH REAL-TIME INFERENCE PERFORMANCE¹¹

3.5X TO 10X HIGHER PYTORCH TRAINING PERFORMANCE¹²

with built-in Intel AMX (BF16) versus the prior generation (FP32)

Do more

Increase the performance of cloud services and applications hyper-scaled across many compute nodes.

Leading customers, including [Google](#), [Meta](#), [Amazon](#) and [Microsoft](#), recognize the value of dramatically improving time-synchronization accuracy across data centers and many have deployed distributed applications that rely on precision network time-synchronization (IEEE Std. 1588-2019 and similar). 4th Gen Intel Xeon Scalable processors include Precision Time Measurement (PTM) that extends precision network time *within the server* from PCIe network device all the way up to the application software running on the CPU, complementing the orders-of-magnitude improvement achievable *between servers*.

For more information on PTM, see "[Adopting Time-Sensitive Networking \(TSN\) for Automation Systems.](#)"

Ease of integration with existing infrastructure

With Intel, businesses can speed up time to deployment with the largest ecosystem of partners they know and use. Hardware and software vendors and solution integrators around the world build their products on Intel Xeon Scalable processors, offering maximum choice and interoperability with the reassurance of thousands of real-world implementations.

Decades of ecosystem enablement help extend Intel's trusted computing foundation across the hyper-scaled data center and the new edge frontier, with everything needed to build, scale and transform for operational agility. The openness of this approach fosters unbridled choice among hardware, software, cloud and service providers. Through the [Intel Partner Alliance](#), access exclusive resources for AI, cloud, high performance computing and other solution areas to help plan, build and deliver more value to your customers.

With the most flexibility to choose different cloud services, shapes and sizes to support specific workload demands, Intel architecture scales globally, across all the major cloud providers.

SUPPORTING STAT

Get the most choice with Intel's **OVER 50,000** unique instance types, sizes and regions. **6X GREATER THAN** the competition.¹³

Leadership's top business priorities in the digital transformation journey

Investments in digital transformation by organization leaders (tech and business alike) are expected to total \$6.3 trillion between 2022 and 2024, accounting for as much as 55% of all IT spending by 2024.¹⁴ This business brief is part of a series that illuminates the top business priorities leaders are focused on to achieve their business success in this transformative future, and how Intel hardware, software and services, including the 4th Gen Intel Xeon processor, help achieve these priorities:



- **Cloud (this brief):** Activate strategies across hybrid, multi-cloud and the intelligent edge
- **Security:** Achieve rigorous security and contribute to your zero trust security strategy
- **AI:** Adopt data analytics and AI to drive critical outcomes
- **Redefined worker experiences:** Embrace boundaryless interactive worker experiences
- **ESG:** Foster equitable outcomes and responsibility in environment | social | governance (ESG)

Learn More

www.intel.com/xeon/scalable

www.intel.com/cloud



¹ Flexera. "2022 State of the Cloud Report." <https://resources.flexera.com/web/pdf/Flexera-State-of-the-Cloud-Report-2022.pdf>.

² IDC Market Forecast, September 2021 (IDC #US47397521). "Worldwide Whole Cloud Forecast, 2021–2025: The Path Ahead for Cloud in a Digital-First World." <https://www.idc.com/getdoc.jsp?containerId=US47397521>.

³ 3rd Gen Intel Xeon Scalable Processor vs. AMD EPYC. See configuration details [126-130] at www.intel.com/3gen-xeon-config.

⁴ See [W5] at intel.com/processorclaims: 4th Gen Intel Xeon Scalable processors. Results may vary.

⁵ Everest Group, October 2021. "Confidential Computing – The Next Frontier in Data Security." https://confidentialcomputing.io/wp-content/uploads/sites/85/2021/10/Everest_Group_-_Confidential_Computing_-_The_Next_Frontier_in_Data_Security_-_2021-10-19.pdf.

⁶ Intel Microservices Insights Study, June 2021.

⁷ See [W1] at intel.com/processorclaims: 4th Gen Intel Xeon Scalable processors. Results may vary.

⁸ See [W6] at intel.com/processorclaims: 4th Gen Intel Xeon Scalable processors. Results may vary.

⁹ See [W7] at intel.com/processorclaims: 4th Gen Intel Xeon Scalable processors. Results may vary.

¹⁰ Accenture, November 19 2019. "AI: Built to Scale." <https://www.accenture.com/us-en/insights/artificial-intelligence/ai-investments>.

¹¹ See [A17] at intel.com/processorclaims: 4th Gen Intel® Xeon® Scalable processors. Results may vary.

¹² See [A16] at intel.com/processorclaims: 4th Gen Intel® Xeon® Scalable processors. Results may vary.

¹³ Source: Historical Liftr Insights Component tracker data + Intel internal preliminary analysis as of 09/02/22.

¹⁴ IDC, October 2021. "IDC FutureScape: Worldwide Digital Transformation 2022 Predictions." <https://www.idc.com/getdoc.jsp?containerId=US47115521>.

Availability of accelerators varies depending on SKU. Visit the [Intel Product Specifications page](#) for additional product details.

Performance varies by use, configuration, and other factors. Learn more at <https://www.intel.com/PerformanceIndex>.

Performance results are based on testing as of dates shown in configurations and may not reflect all publicly available updates. See configuration disclosure for configuration details. 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.

Your costs and results may vary.

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

You may not use or facilitate the use of this document in connection with any infringement or other legal analysis concerning Intel products described herein. You agree to grant Intel a nonexclusive, royalty-free license to any patent claim thereafter drafted which includes subject matter disclosed herein.

The products described may contain design defects or errors known as errata which may cause the product to deviate from published specifications. Current characterized errata are available on request.

© 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.

1122/MH/MESH/350495-002US