



intel

Optimize your AI infrastructure for energy efficiency with Intel®

The trajectory of artificial intelligence (AI) within the enterprise can have huge impact on already challenging data center capacity constraints—including power and space issues—as well as your environmental footprint. A comprehensive approach to data center optimization is required.



Intel® Xeon® Scalable processors are the industry's most widely used CPUs, representing

65%

of the processor units (installed base) that are running AI inference workloads in the data center¹

Intel can help you deliver high-impact AI capabilities with a lower environmental footprint. Only Intel offers an end-to-end, heterogeneous portfolio of AI-optimized hardware, combined with a comprehensive, interoperable suite of AI software tools and framework optimizations to accelerate your AI workflows at every stage. With more acceleration capabilities and optimized performance, Intel hardware and software solutions can help execute AI functions more quickly and help your data center operate with more energy efficiency.

And with Intel's solutions and partner ecosystem, you can go beyond driving energy-effectiveness in AI training and inference workloads. Use AI to optimize the overall efficiency of your data center: executing carbon-aware computing based on renewable energy availability, optimizing cooling, and predicting hardware inefficiencies and server health.



Powerful results, with less power

Deliver resource-intensive AI workloads more efficiently with solutions that allow you to optimize performance-per-watt for both general purpose and AI workloads.



Efficiency with intelligent operations

Utilize AI insights to improve efficiency of your data center. Increase system performance and reduce power draw automatically with hardware-based telemetry and scalable software tools to make sure you're getting the most from every watt.



Business impact at a lower cost

Reduce the total cost of ownership (TCO) and environmental footprint of your data center with versatile solutions that can meet the needs of your data center today and scale into the future.

Intel® helps simplify right-sized AI



With the widest array of hardware architectures in the industry, built-in accelerators, software tools and frameworks, we can help manage your total cost of ownership (TCO) and ensure you're getting the most from your resources.

Use AI to reduce overall data center energy consumption

Intel technology offers real-time insights into power usage, thermals, resource utilization, and system health to help your data center operate with efficiency.

Continuously optimize cloud, hybrid and on-prem workloads in real-time with Intel® Granulate™

Granulate can increase data center performance by

up to 60%
and lower compute spend by up to 30 percent⁴

Build a more intelligent, carbon-aware data center, automatically

Increase reliability and carbon efficiency with a proactive approach to data center management using telemetry in Intel® Xeon® scalable processors and Intel's tools in Kubernetes



Find energy efficiency with optimized solutions for common AI workloads

Only Intel offers a full suite of software and hardware technologies for the entire data and AI pipeline, offering IT efficiency and energy savings optimizations.

Accelerate deep learning training, inference, and deployment with greater power- and carbon-efficiency than the competition

Habana® Gaudi2® servers deliver

2X higher

throughput per watt than the comparable NVIDIA A100, giving you more training with lower power consumption²

Intel's frameworks and libraries are designed to deliver drop-in 10-100X acceleration across a variety of applications, models, and use-cases

Intel® oneAPI can provide

16-53x

gain in image classification inference, a 10x gain for object detection, and nearly 5x gain for recommendation systems, giving you more insights for less power³

Expand your AI capabilities while managing TCO and environmental impact

Reduce unnecessary waste with leading Intel solutions that operate within your existing infrastructure AND scale to future needs.

Flex your hardware to the changing needs of your business

The latest Intel Xeon scalable processors include built-in acceleration capabilities and hardware-enhanced features, and with Intel® OnDemand, you can expand, or upgrade select CPU-based capabilities as needed

Ensure scalability of your AI solution architectures and free your data center from proprietary lock-in

Only Intel offers an end-to-end AI software ecosystem, built on an open, interoperable programming model, including plugins for NVIDIA and AMD hardware

See how Intel can help increase sustainability across your data center at intel.com/SustainableDataCenter

Notices & Disclaimers

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.

¹ Based on Intel market modeling of the worldwide installed base of data center servers running AI Inference workloads as of December 2022.

² Power performance measurements for power utilization on ResNet performed by Supermicro in their lab (April 2023). Configuration details available at Habana website: <https://habana.ai/habana-claims-validation/>.

³ For configuration details, see <https://venturebeat.com/ai/software-ai-accelerators-ai-performance-boost-for-free/>

⁴ For more information on this performance data, please go to the <https://edc.intel.com/content/www/us/en/products/performance/benchmarks/vision-2022/page>