

The Intel logo is displayed in white lowercase letters on a blue square background. The background of the entire slide features a complex, glowing blue circuit board pattern with various lines and components.

2025 OCP Global Summit

Scaling AI at the Speed of Openness: From Silicon to Systems

Sachin Katti

Senior Vice President
Chief Technology and AI Officer

Impact in collaboration with the Open Compute Project



Commitment

to sustainable, flexible, and scalable tech infrastructure.

Choice

Breaking free from proprietary constraints

~90 Contributions

From OCP-NIC to DC-MHS to Open Systems for AI

AI as a Strategic Priority

Massive transformation in AI

Redefining every layer of the stack deployed from **AIPC**, to **edge**, to **data center**

CPU relevance in AI ecosystem

Open heterogeneous strategy to deliver **systems, software & GPUs**

Agentic AI is the Means to Unlocking Enterprise Value in AI

~80% of AI compute cycles in inference by 2028*

Diverse workload requirements

Throughput, low latency, perf/\$/W and system-level tuning required

Physical AI

Real-time inferencing for physical world interaction

Agentic AI

Billions of daily inferences across thousands of models

Frontier AI

Model training - massive, infrequent, hyperscaler-scale workloads

2015

2020

2025

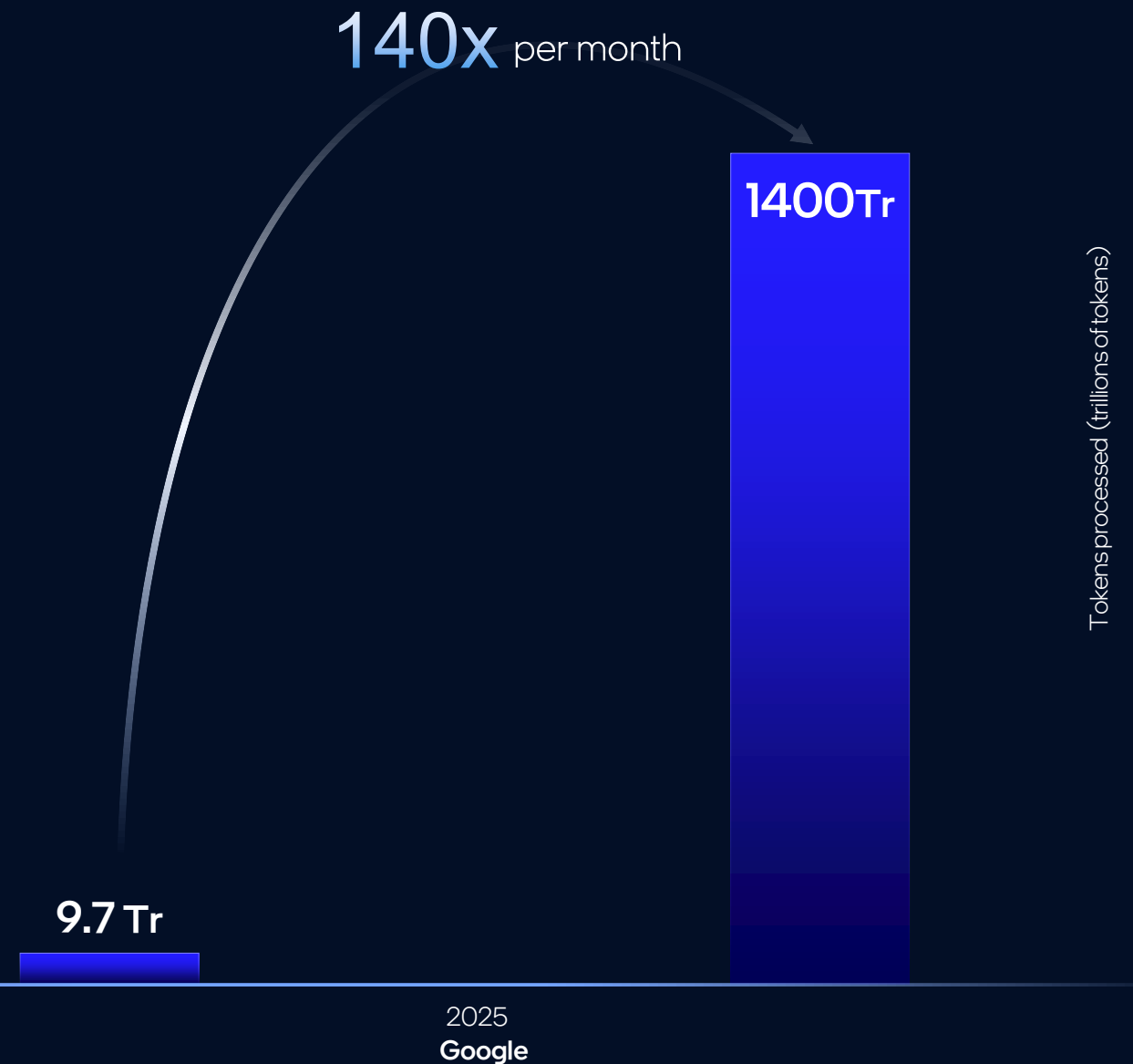
2030

2035

*Source: Gartner, published 08/2024

Explosive Token Growth

Critical to optimize system-level
tokens per dollar per watt



Source: 2025 AI Infra Summit

Agentic AI Today

Deployed on homogeneous systems

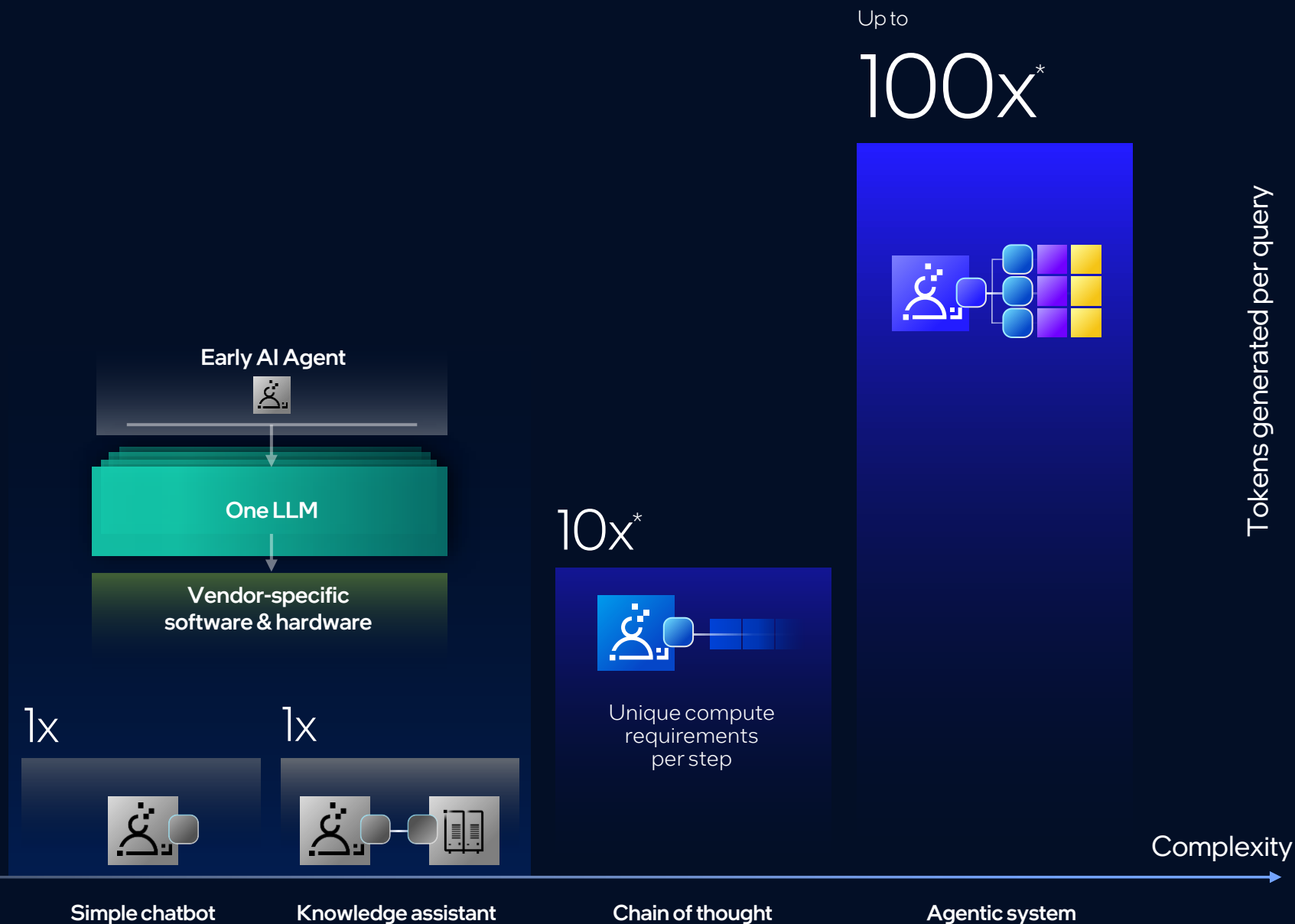
Vertically integrated networking & software

Poor performance per dollar

☐ Unique Model

☐ Unique Steps

*Not actual numbers, approximate tokens per query based upon external sources and Intel internal estimates

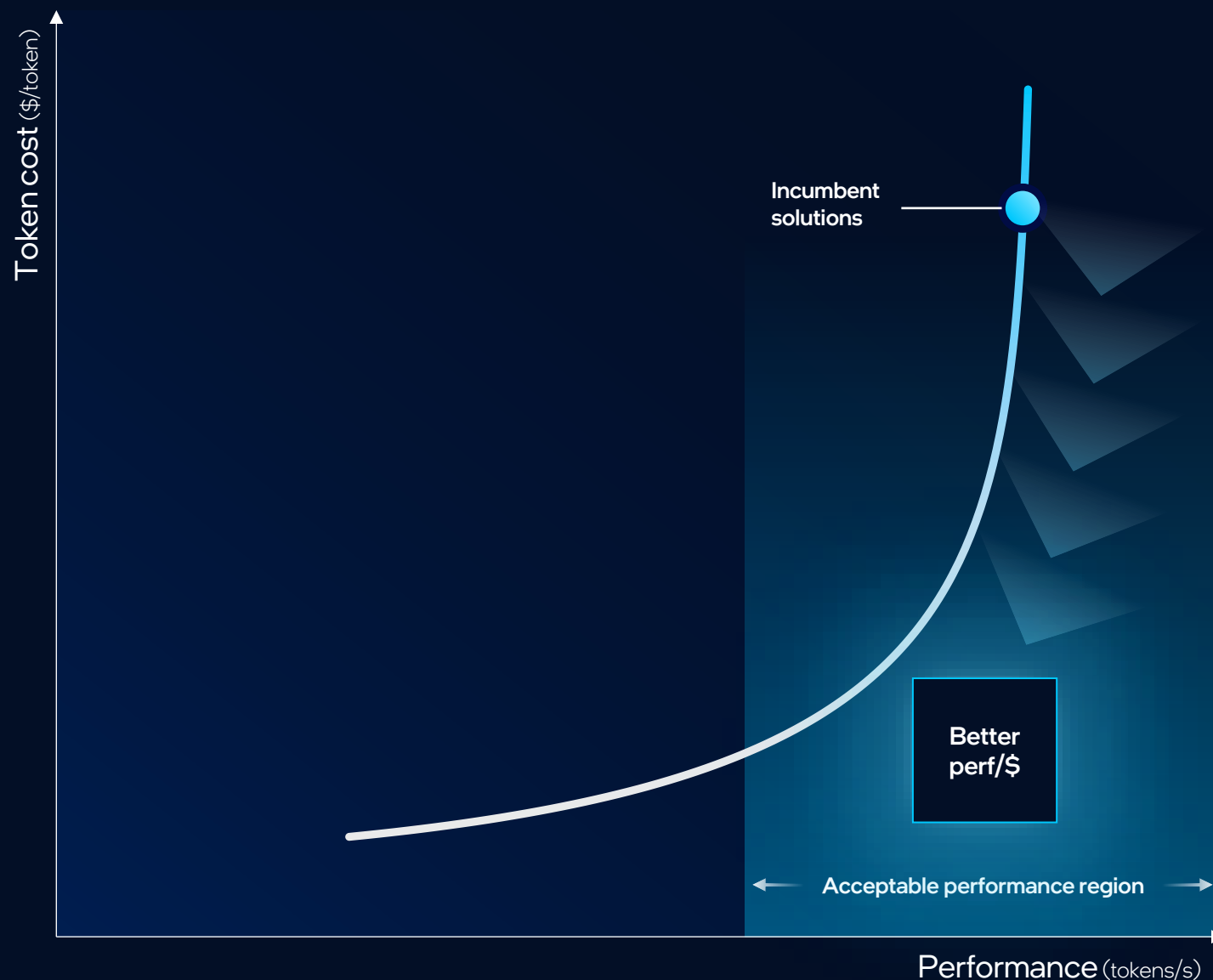


Based on Intel internal analysis of information available as of Sept. 2025. Visualization for illustrative purposes only.

Homogenous, vertically integrated → Hard to scale perf/dollar

Exponentially worse in Agentic
era due to diverse requirements

Based on Intel internal analysis of information available as of Sept. 2025.
Visualization for illustrative purposes only.



Agentic AI Future

Complex multi-agent workflow

Diverse models & APIs

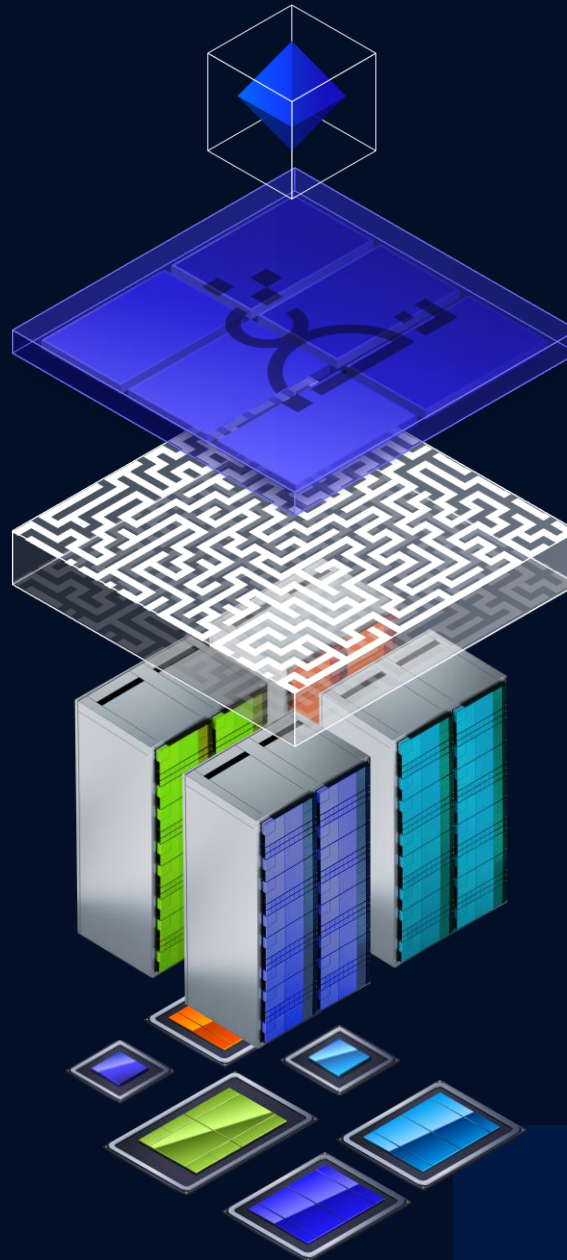
Different building blocks with diverse hardware req.

Need for hardware diversity and openness is exacerbated



Heterogenous Infrastructure Required

To improve efficiency and performance per dollar today, and future ready for Agentic AI tomorrow

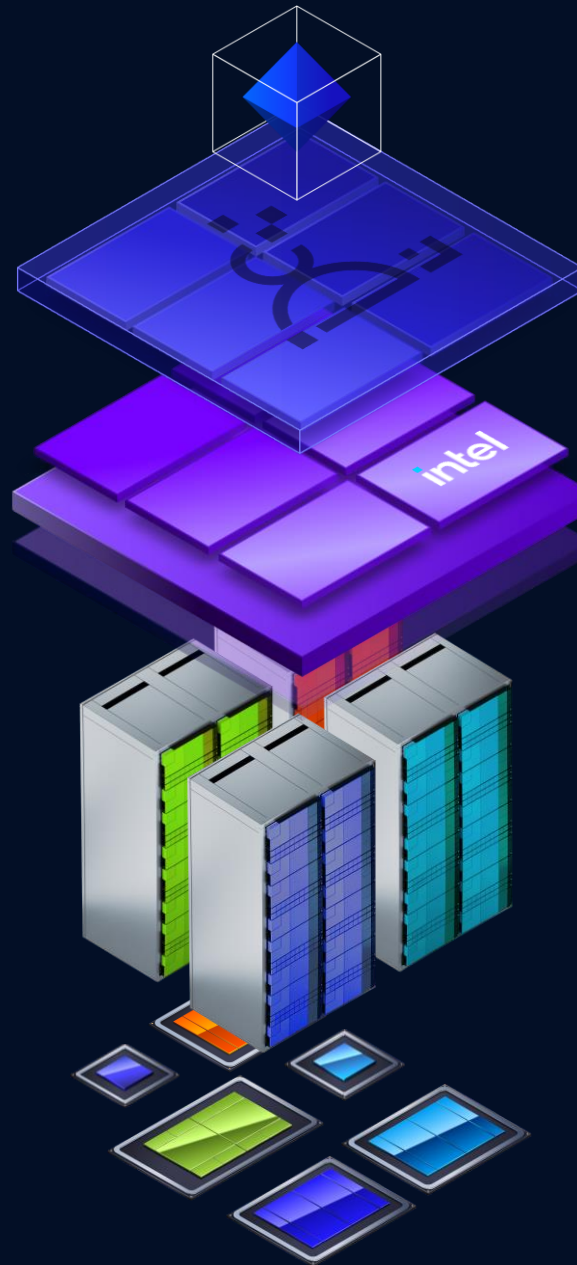


Need to reimagine the software and orchestration layers

Bring the benefits of open collaboration to hardware infrastructure

Need for Simplification & Abstraction

Delivering a unified software stack approach to enable multi-vendor and multi-architecture infrastructure that unlocks scalability across the enterprise



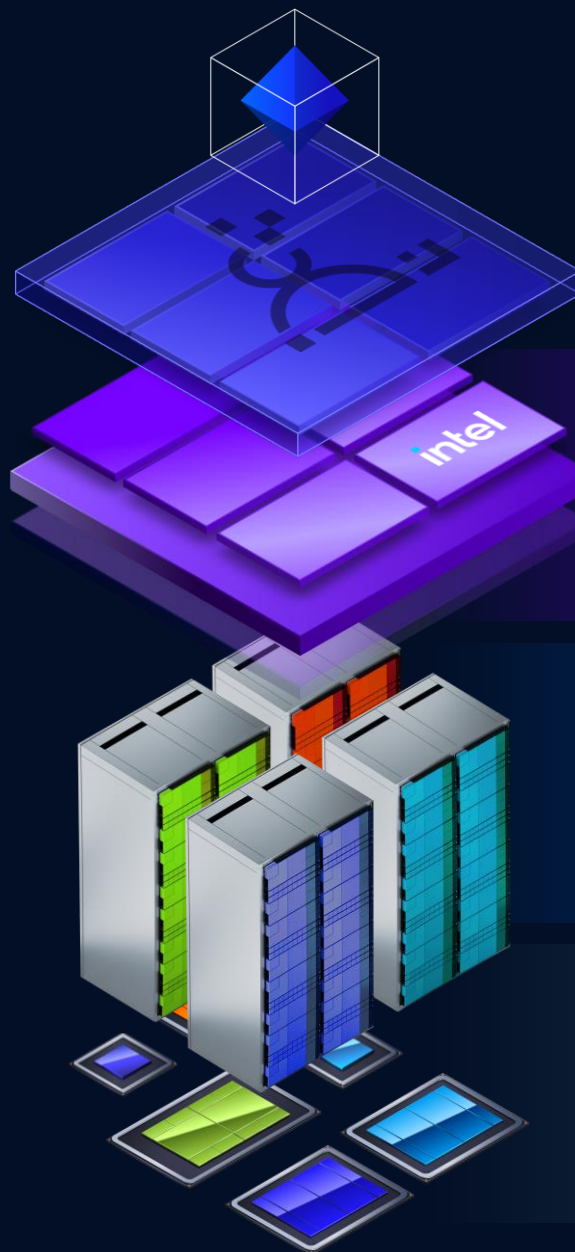
Abstract away heterogenous infrastructure complexity

intel

Scalable Heterogenous Systems

Disrupting the current approach

Solving customer needs by
reducing friction to deploy
AI at scale



Open AI software stack

Programmability

Orchestration

Multi-framework/
model/API

System infrastructure

Enterprise-friendly
server form factors

Open ethernet-based
scale-out networking

Heterogenous hardware

Multi-architecture

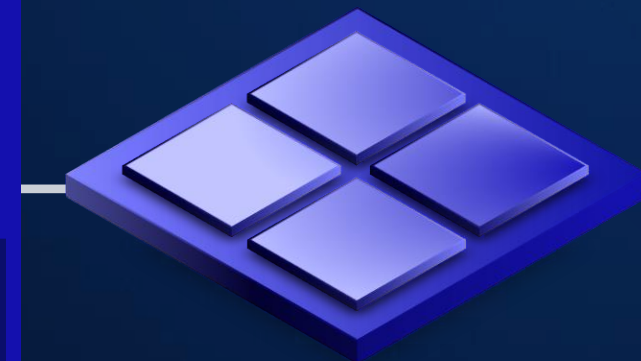
Multi-vendor

intel

Open AI Software Stack

Purpose-built to orchestrate
multiple agents across
hardware

Turnkey software offering to
simplify deployment and
scale of AI

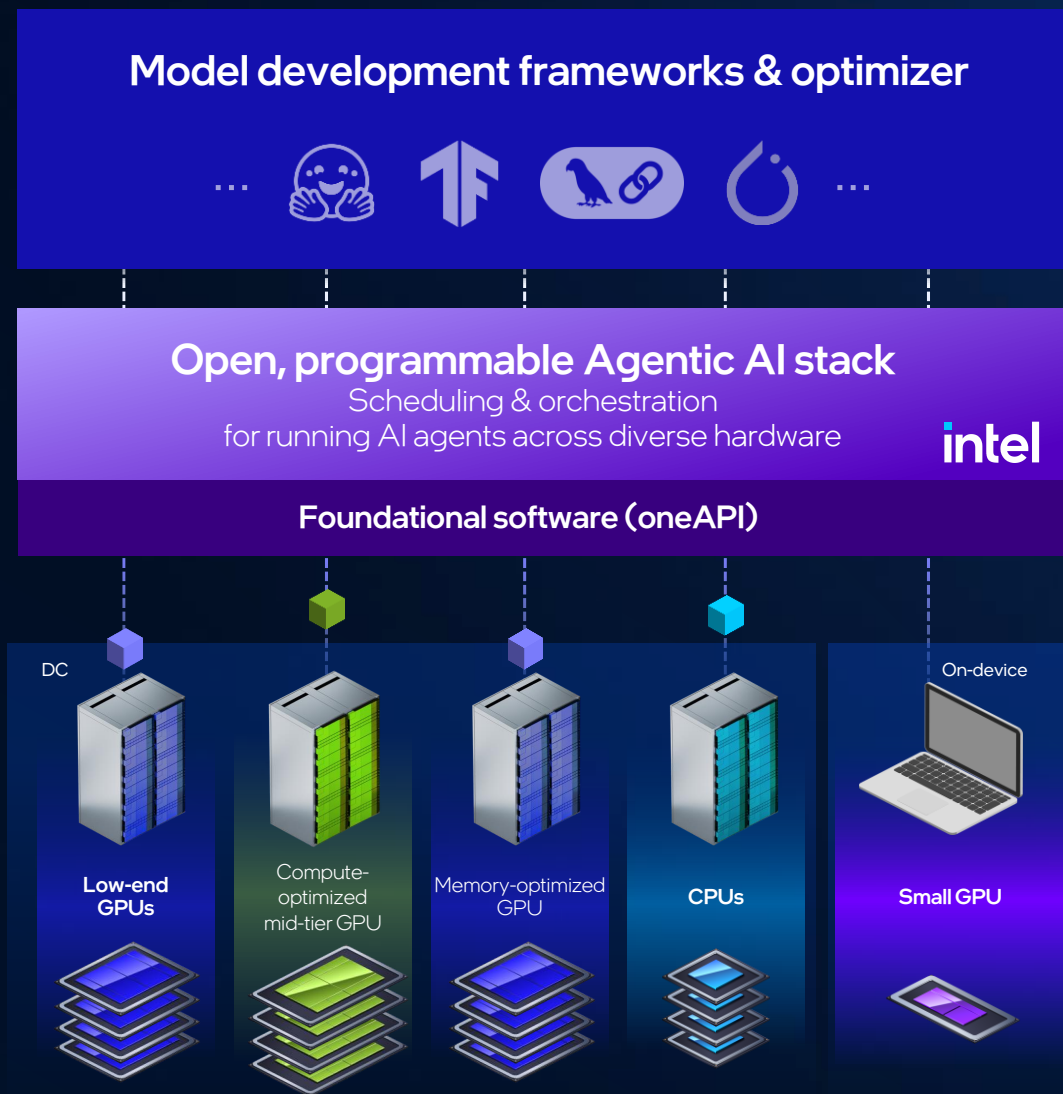


intel

Open AI Software Stack

Purpose-built to orchestrate
multiple agents across hardware

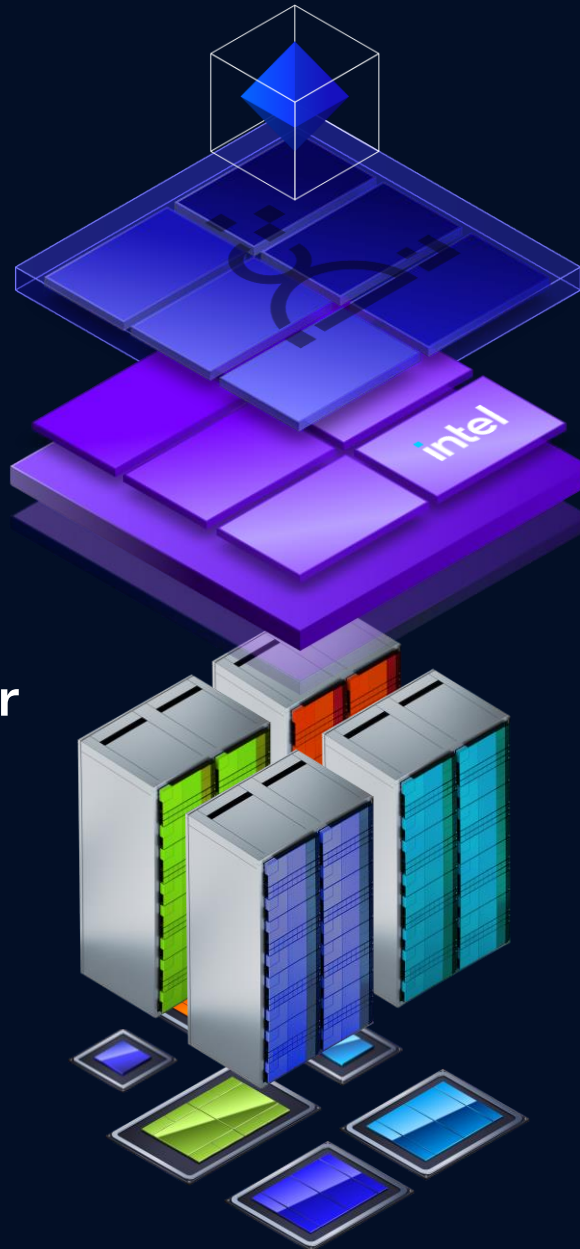
Turnkey software offering to
simplify deployment and
scale of AI



Performance & TCO Benefits

Scalable, heterogeneous,
multi-vendor systems, coupled
with an open enterprise stack

**Sustainable solutions that deliver
better TCO** without performance
compromise



[Details at intel.com/performance/index](https://www.intel.com/performance/index)

Up to

1.7x

perf/TCO benefits

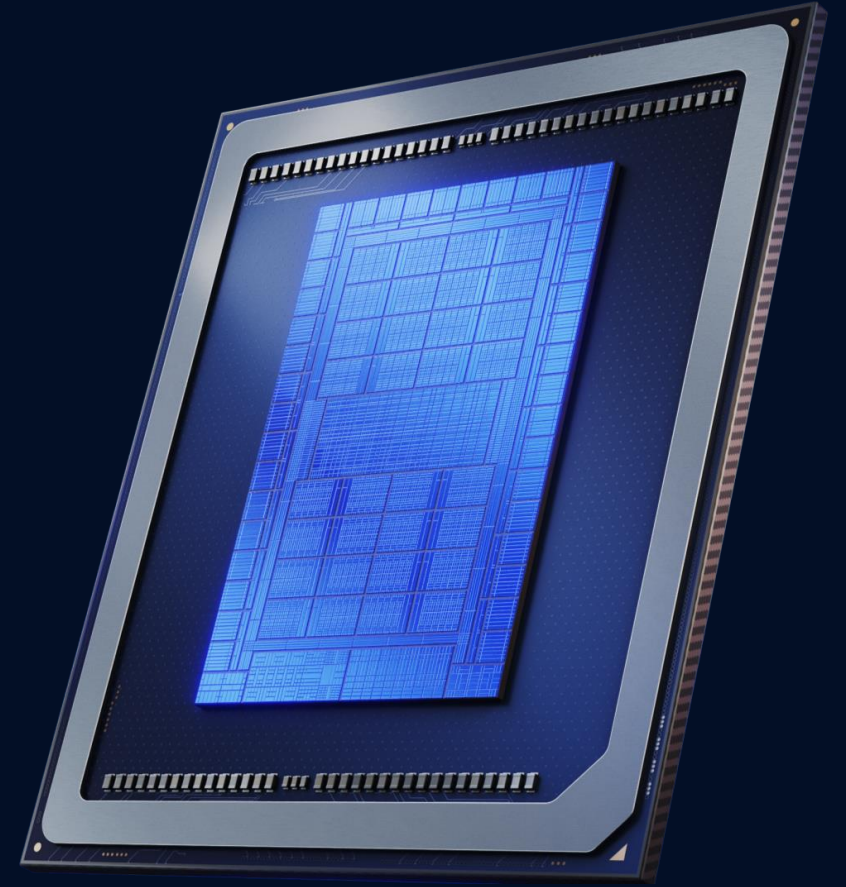
Gaudi3 + B200 heterogeneous system
vs. B200 homogeneous systems

Llama 8B FP16 / Llama 70B FP16

Next Gen

Intel® Data Center GPU

Code-name – Crescent Island



Inference optimized & Performance per TCO optimized

Next Gen

Intel® Data Center GPU

Code-name – Crescent Island

Designed for tokens/watt built on a reliable open software stack

160GB LPDDR5x

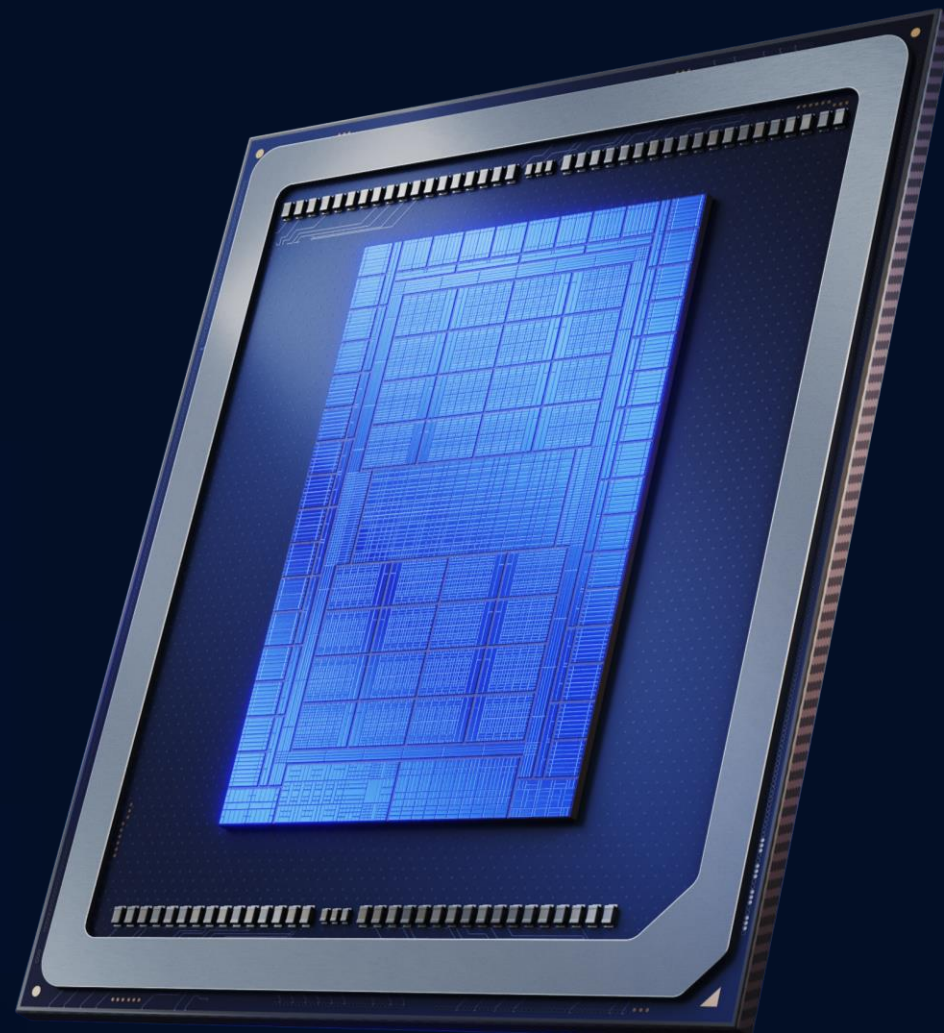
to meet the needs of memory intensive AI workloads

Latest Generation Xe 3P IP

General purpose GPU architecture for competitive perf/watt

Widest Range AI datatypes

Support from FP4/MXP4 to FP32 and FP64



The x86 Scale Advantage

One year of the x86
Ecosystem Advisory
Group



Choice

Diversity and Breadth of Ecosystem

Software Scale

Application Scale

Virtualization
Support

Accelerated time to market

Compatibility across platforms

Legacy Support

Interfaces

SW Consistency

Trusted Results

Predictable performance

Workload
Breadth

Continuous
Optimization

Real-world
Reliability

Open System Development

Foundation for heterogeneous AI infrastructure

Open system design



Open rack design
(OCP ORv3)



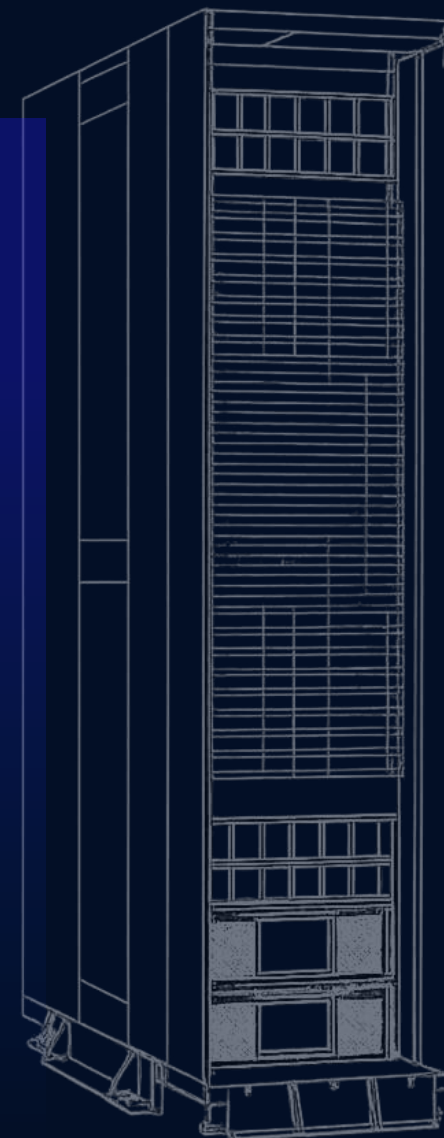
Open software



Open scale-up
connectivity



Open scale-out
connectivity



The Intel logo is centered on a dark blue background. It features the word "intel" in a white, lowercase, sans-serif font. A small blue square is positioned above the letter "i". To the right of the word "intel" is a small registered trademark symbol (®).

intel®

Forward Looking Statements Disclaimer

Statements in this presentation that refer to business outlook, plans, and expectations are forward-looking statements that involve risks and uncertainties. Such statements may include, but not be limited to, those regarding: our business plans, strategy and leadership and anticipated benefits therefrom; future products, services and technologies, and the expected goals, timeline, ramps, progress, availability, production, regulation and benefits of such products, services and technologies, including future process nodes and packaging technology, product roadmaps, schedules, future product architectures, expectations regarding process performance, per-watt parity, and metrics and expectations regarding product and process competitiveness; internal and external manufacturing plans; future production capacity and product supply; supply expectations; plans and goals related to Intel's foundry business, including with respect to anticipated customers, future manufacturing capacity and service, technology and IP offerings; and other characterizations of future events or circumstances.

Such statements involve many risks and uncertainties that could cause our actual results to differ materially from those expressed or implied, including: the high level of competition and rapid technological change in our industry; the significant long-term and inherently risky investments we are making in R&D and manufacturing facilities; the complexities and uncertainties in developing and implementing new semiconductor products and manufacturing process technologies; implementing new business strategies and investing in new business and technologies; our ability to time and scale our capital investments appropriately and successfully secure favorable alternative financing arrangements and government grants; changes in demand for and margins on our products; macroeconomic conditions and geopolitical tensions and conflicts, including geopolitical and trade tensions between the US and China, tensions and conflict affecting Israel and the Middle East, rising tensions between mainland China and Taiwan, and the impacts of Russia's war on Ukraine; the evolving market for products with AI capabilities; our complex global supply chain, including from disruptions, delays, trade tensions and conflicts, or shortages; recently elevated geopolitical tensions, volatility and uncertainty with respect to international trade policies, including tariffs and export controls, impacting our business, the markets in which we compete and the world economy; product defects, errata and other product issues; potential security vulnerabilities in our products; increasing and evolving cybersecurity threats and privacy risks; IP risks including related litigation and regulatory proceedings; the need to attract, retain, and motivate key talent; strategic transactions and investments; sales-related risks, including customer concentration and the use of distributors and other third parties; our debt obligations and our ability to access sources of capital; our having ceased to return capital to stockholders; complex and evolving laws and regulations across many jurisdictions; fluctuations in currency exchange rates; changes in our effective tax rate; catastrophic events; environmental, health, safety, and product regulations; corporate responsibility initiatives and requirements; and other risks and uncertainties described in this presentation, our most recent Annual Report on Form 10-K and Quarterly Reports on Form 10-Q, and our other filings with the SEC.

Unless specifically indicated otherwise, the forward-looking statements in this presentation do not reflect the potential impact of any divestitures, mergers, acquisitions, or other business combinations that have not been completed as of the date of this presentation. All information in this presentation reflects management's views as of September 28, 2025, unless an earlier date is specified. We do not undertake, and expressly disclaim any duty, to update such statements, whether as a result of new information, new developments, or otherwise, except to the extent that disclosure may be required by law.

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.