Solution Brief

Distributed Computing on Intel® Technologies

Unleash Growth with the Power of AI Through Distributed Computing

Intel technologies across distributed computing give you the freedom to choose when and where you process data.

The platform consistency of Intel technologies across the compute spectrum empowers users to choose where to place the right workload at the right time.

Authored by:

Jeff Jackson, Sr. Intel Principal Engineer

Tiffany Sargent, Sr. Intel Principal Engineer

Lokendra Uppuluri, Intel Cloud Systems and Solutions Architect

Grant Kelly, Intel Systems and Solutions Architect

Shawna Meyer-Ravelli, Intel Cloud and Data Center Product Marketing

Since the birth of computing, data has fueled our digital world, driving it forward relentlessly. With this data availability, unprecedented processing power, and computing capability across locations and systems, we find ourselves at the dawn of an era of endless opportunity and world-changing breakthroughs.

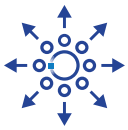
To harness the power of data, organizations require intelligent systems built on the latest edge-to-cloud architecture and using integrated silicon, software, and system components. Such a system should enable the distribution of workloads across compute resources in different locations, from the edge to the cloud and the networks that connect them. This flexibility allows organizations to process and utilize data where and when they need it, unlocking the full potential of insights at their disposal.

These modern computing needs are the impetus behind the rise of distributed computing. Distributed computing comprises various resources across multiple locations that collect, process, store, and transmit data while providing security measures to help prevent unauthorized access or manipulation. While it's not a new paradigm, distributed computing is on the rise due to advances in AI and the growth of data outside the data center. That's because distributed computing provides the flexibility to run inferencing on AI models at the most optimal location, often the edge, while also supporting the large datasets used in AI training to develop and optimize those models. Organizations now have the unprecedented freedom to convert data into useful information where it's most advantageous for the end user, all while enhancing cost efficiency, improving performance, and expanding network capacity.

This brief provides an overview of distributed computing, its tenets, and how Intel plays a key role in enabling it in the intelligence era. Through open standards and flexible architectures, Intel delivers the freedom to choose where data is processed. Intel technologies—from processors to networking to software to developer tools—are key enablers of distributed computing. They are the result of decades of purposeful design that enable open computing platforms and inspire collaborative partnerships within an ecosystem of companies and industry consortiums to realize this distributed computing framework today.

The platform consistency of Intel technologies across the compute spectrum empowers users to choose where to place the right workload at the right time. One example is training AI models in the data center while running AI inferencing workloads at the edge—all with the benefit of operational manageability, workload performance optimization, and hardware-based security features. These features include Intel® Software Guard Extensions (Intel® SGX) for confidential computing, which has helped multiple parties in healthcare collaborate seamlessly while helping ensure the confidentiality of both data and algorithms.

Distributed Computing Supports Multiple Computing Paradigms



Decentralized computing

Distributes computational resources, data storage, and decision-making across a network of interconnected nodes, spreading tasks and responsibilities among multiple devices and locations.

Parallel computing

Divides a computation task into smaller tasks and runs them concurrently, providing results faster than with sequential computing.

Cluster computing

Uses a group of computers working together as a single system. Clustered computers are often used for high-performance computing (HPC) applications, such as scientific research or weather forecasting.

Grid computing

Uses a network of computers to share resources. Grid computing is often used for large-scale projects, such as drug discovery or climate modeling.

Edge computing

Brings computing and data storage closer to the location where it is needed, which can help improve response times. This is particularly important for applications that require real-time processing or low latency, such as autonomous vehicles, industrial automation, and smart cities.

Cloud computing

Delivers computing services—including servers, storage, databases, networking, software, analytics, and intelligence—over the internet ("the cloud"). It also includes private clouds (on-premises servers running cloud services) and virtualization.

Workloads: AI inferencing (edge), AI training (cloud), data storage, data compression/encryption (encode/decode), network functions, graphics and visualization, and user applications

Distributed computing overview

Intentional design and optimized engineering are the foundation of the five tenets of distributed computing, which support computing across a diverse set of industries and usage models. These tenets are explained in this section.

Connectivity is the pipeline for compute resources to communicate effectively with users, data sources, and other infrastructure components across an open, standards-based network. Distributed computation solutions must consider bandwidth, latency, and network reliability to determine where workloads are placed. Intel® Ethernet technologies and standards for the industry support interoperability and integration across the distributed computing spectrum. Intel Ethernet controllers, adapters, and accessories—which are extensively tested for network interoperability, reliability, and performance—can deliver speeds from 1 to 100 gigabits per second (Gbps) and include capabilities to help optimize workload performance.

Manageability enables discovery, provisioning, and administrative tasks from edge to data center to cloud. Intel edge manageability solutions enhance overall efficiency and productivity by supporting updates and monitoring from remote locations. The Intel vPro® platform's in-band and out-of-band management capabilities allow users to remotely monitor and manage client and edge devices, update firmware and operating systems, and perform many other administrative operations. Intel technologies support container orchestration that automatically schedules and distributes computing workloads, storage, and network bandwidth across resources from the data center to the network to the edge.

Security needs to extend from hardware to software. A silicon root of trust helps ensure that data across distributed computing environments is protected from unauthorized access, breaches, and compliance violations. Compute infrastructure should also be resilient to failures and disruptions to help ensure the continuous availability of applications and services. Intel hardware-based security technologies are designed to help customers defend against modern cyber threats.

Interoperability enables applications and services to seamlessly work, integrate, and scale across diverse platforms. Compute resources should be extensible to accommodate fluctuating workloads, customer demand, and future growth. Through co-engineering partnerships with software providers like Microsoft, Amazon Web Services (AWS), Google, SAP, Red Hat, and VMware, Intel jointly designs and builds new solutions with the software industry to accelerate advancements through integrated and optimized software and hardware. This collaboration is critical to Intel's inherent approach to designing open architectures instead of proprietary technologies.

Performance manages both optimized workloads and costs across the distributed computing spectrum and is especially important in the AI era. Platform performance depends on standard hardware platforms optimized at the right location for each workload. Intel® Xeon® processors featuring built-in hardware accelerators boost performance for demanding workloads, such as AI/machine learning (ML), encryption, and data compression.

Across these five tenets of distributed computing, complex factors come into play when creating a solution. These factors are situational and cross multiple industries and usage models that drive the placement of workloads encompassing network connectivity, scalability, security, cost, resilience, regulatory compliance, workload performance, and environmental considerations. Intel offers technologies across the distributed computing spectrum that are designed to optimize performance, improve the value of technology investments, and successfully navigate through these complexities.

Distributed Computing: The Freedom to Choose Where You Process Your Data

Distributed computing is an interconnected system built on edge-to-cloud architectures using integrated silicon, software, and system components.

The five core tenets of distributed computing

- Connectivity
- Manageability
- Security
- Interoperability
- Performance

Software and Workloads

- oneAPI
- OpenVINO™ toolkit
- Intel® Geti™ software
- Intel® Tiber™ portfolio

Cloud and Data Center

Data centers and cloud service providers (CSPs)

- Intel® Xeon® processors
- Intel® Accelerator Engines
- Intel® Gaudi® AI accelerator
- Intel® Data Center GPUs

Edge

Devices and "things"

- Intel Xeon processors
- Intel® Core™ processors
- Intel Atom® processors
- Intel® GPUs and accelerators
- Intel® Movidius™ Vision Processing Units (VPUs)

Network

Wired and wireless (Wi-Fi, 5G, LAN, WLAN)

- Intel® Ethernet products
- Intel® Infrastructure Processing Units (Intel® IPUs)

Solutions available today

Unlocking the power of data through distributed computing can transform your organization. Intel and our trusted partner ecosystem offer solutions that allow workloads to be moved across different locations throughout the computing spectrum. Through open standards, a software platform across heterogeneous architectures, and security-first engineering—all built for scale—Intel and its partners deliver a full range of distributed computing solutions.

Run multicloud analytics on virtual machines

VMware and Intel collaborated to deliver a comprehensive solution to manage virtual machines (VMs) and orchestrating containers. This reference architecture brings compute capacity and AI models to where enterprise data is created, processed, and consumed, whether in a public cloud, in an enterprise data center, or at the edge. This verified, end-to-end solution also accelerates analytics deployment for a wide range of workloads; delivers agile, scalable, and security-enabled infrastructure with excellent analytics performance; and increases throughput with Intel® architecture–optimized deep learning (DL) frameworks.[1]

Use federated learning for disease detection and treatment

Intel and Penn Medicine used federated learning to unlock multi-site data silos to improve disease detection and treatment. Research using federated learning represents a significant step forward in medical AI, improving brain tumor detection by 33 percent.[2] Intel SGX and other Intel technologies decentralized data processing while maintaining data privacy, enabling research without compromising patient information.

Protect confidentiality with distributed AI

Intel helped develop Open Federated Learning (OpenFL) to enable data scientists to address the challenge of maintaining data privacy while bringing together insights from disparate, confidential, or regulated datasets. This library was developed to promote collaboration and standardization in distributed AI. OpenFL enables privacy-preserving AI by allowing data analysis without sharing raw data using technologies like Intel SGX and the Gramine project.

Fend off phishing attacks with AI at the edge

BUFFERZONE has collaborated with Intel to pioneer the first endpoint-based anti-phishing solution that harnesses the computational power of Intel AI technologies. By running at the edge, the BUFFERZONE NoCloud AI anti-phishing detection system provides a 91-percent reduction in anti-phishing operational costs, inference speed improvements of more than 40 percent compared to CPU inference, and a 70-percent decrease in detection latency compared to cloud inference while protecting user privacy.[3] The Intel® Core™ Ultra processor integrates the CPU, graphics processing unit (GPU), and neural processing unit (NPU) into a single package and helps accelerate AI at the edge instead of in the cloud, where DL workloads are traditionally run.

Enable AI and security with Red Hat–validated designs

Red Hat–validated patterns from Intel make it easy to enable AI, security features, and accelerations provided by the latest Intel Xeon Scalable processors running Red Hat OpenShift 4.14. These patterns simplify the deployment process, eliminating the need for extensive manual configurations and setting up comprehensive, fully operational AI workflows with ease. The solution utilizes Intel® Advanced Matrix Extensions (Intel® AMX) to boost the speed and efficiency of AI workloads. In terms of security, Intel SGX helps provide robust support for confidential computing, improving data protection. Additionally, application performance is improved by offloading encryption and compression to a dedicated accelerator using Intel® QuickAssist Technology (Intel® QAT).

Optimize image inferencing with Intel AMX

Using Red Hat–validated patterns, Intel quantized a medical diagnosis pipeline to use lower precision to accelerate image inference at the edge. Quantization is a technique in AI to reduce the computational memory costs of running inference. Using the Red Hat Node Feature Discovery (NFD) operator to identify nodes equipped with Intel AMX and an open source scaling tool, the extended, validated pattern unleashed the power of AI to assess chest X-rays for the risk of pneumonia.

Accelerate containerized workloads

Kubernetes is a powerful tool that provides a robust framework for running distributed systems, allowing users to deploy, scale, and manage containerized applications anywhere. The Intel device plugins framework for Kubernetes makes it easy to accelerate complex containerized workloads by using advanced hardware resources such as GPUs, high-performance network interface controllers (NICs), and field-programmable gate arrays (FPGAs). Intel provides a container experience kit that's a simplified mechanism for installing and configuring Kubernetes clusters on Intel architecture using Ansible.

Intel also developed a purpose-built framework, Cloud Native Data Plane (CNDP), optimized for packet processing in cloud applications. By integrating with AF_XDP plugins for Kubernetes, CNDP applications can run in security-enabled, unprivileged pods.

Sustainability in distributed computing

As part of the global community, Intel is focused on the environmental goals shared by many organizations today, such as energy efficiency, a small carbon footprint, and renewable energy sources. Intel technologies are engineered to help organizations achieve their sustainability goals, with innovations such as built-in AI accelerators on Intel Xeon Scalable processors that provide better performance per watt—up to 10x better on select AI workloads.[4]

These performance benefits deliver real-world results. Nature Fresh Farms, a family-owned non-GMO greenhouse farm supplying retailers in North America, wanted an innovative edge solution to support its goal of reimagining the food supply chain from the ground up. Using an AI-based solution at the edge with Intel Xeon Scalable processors and the Intel OpenVINO™ toolkit, the farm optimized its operations, including improvements in heating, cooling, humidity, and irrigation. The solution optimized the path from seed to supermarket shelf, leading to better food for people, less waste in the food supply chain, and up to 10x more plant yield per acre compared to traditional farming.[5]

Another recent innovation is Intel Xeon 6 processors with Efficient-cores (E-cores) that deliver performance and power efficiency for high-density, scale-out workloads in the data center. These processors enable a 3:1 rack consolidation for overall power savings, rack-level performance gains of up to 4.2x, and performance-per-watt gains of up to 2.6x.[6]

Empower your organization with distributed computing

As distributed computing continues to rise in importance with the pervasive use of AI, Intel technologies can be mobilized to drive optimized and relevant solutions. The right choice of compute locations can offer customers several benefits. Distributed computing can speed up the process of getting information and improve the quality of information; reduce IT costs related to data transmission, storage, and infrastructure/operations; and enhance security. New types of innovation can be unleashed by allowing computing to happen at the earliest point of data ingestion and separating the compute nodes where AI inference and training are performed. Distributed computing solutions from Intel are designed to help businesses stay ahead in the technology landscape, increase efficiency, and unlock new opportunities.

Join us at an upcoming event to gain relevant business insights, collaborate with others, and learn about the latest advancements, best practices, and how Intel can help you overcome today's business challenges. These webinars are free and available on demand.

- Webinar: GenAI Webinar Series
- Webinar: Growing Revenue with Artificial Intelligence
- Webinar: Reduce Costs by Bringing Your Technology into the Future
- Webinar: Mitigating Security Risks with Confidential Computing

Stay informed at the Intel hub for technology leaders.

intel.

[1] VMware. "News & Stories: VMware Collaborates with Intel to Unlock Private AI Everywhere." November 2023. https://news.vmware.com/releases/vmware-explore-2023-barcelona-intel-private-ai.

[2] Nature Communications. "Federated learning enables big data for rare cancer boundary detection." December 2022. nature.com/articles/s41467-022-33407-5.

[3] BUFFERZONE. "BUFFERZONE® and Intel® AI Anti-Phishing Solution presented at Mobile World Congress." February 2024. https://bufferzonesecurity.com/news-events/bufferzone-and-intel-ai-anti-phishing-solution-presented-at-mobile-world-congress/.

[4] See [A16] and [A17] at intel.com/processorclaims: 4th Gen Intel Xeon Scalable processors. Results may vary.

[5] Intel. Organic Intelligence fact sheet. 2024. https://download.intel.com/newsroom/2024/artificial-intelligence/newsroom-organic-intelligence-fact-sheet.pdf.

[6] Intel. "Intel Accelerates 'AI Everywhere' at Computex 2024." June 2024. intel.com/content/www/us/en/newsroom/resources/computex-2024.html.

Printed in USA 0824/SMR/PRW/PDF Please Recycle 360571-001US