Gartner®

Top Trends in Building a Digital Future: Next-Gen Computing

Published 15 August 2022 - ID G00761112 - 15 min read

By Analyst(s): Nick Jones, Arun Chandrasekaran, Arnold Gao, Alan Priestley

Initiatives: Technology Innovation

New computing technologies will deliver improved sustainability, enable new algorithms, provide improved price and performance, and deliver low-cost tagging and sensing. Enterprise architecture and technology innovation leaders will find these capabilities underlying many business innovations.

Additional Perspectives

- Summary Translation + Localization: Top Trends in Building a Digital Future: Next-Gen Computing
(21 December 2022)

Overview

Opportunities

- Given CEOs' increased focus on environmental sustainability, hardware and software will be subject to more sustainability scrutiny. Innovative hardware will provide new ways for organizations to improve sustainability without compromising functionality.
- Business innovation demands ever-greater computing power in new contexts such as edge computing, which will be delivered using new hardware architectures.
- The algorithms and capabilities needed for digital business leadership will outstrip current computing technologies, creating a demand for novel solutions.
- Digital business will demand dramatically enhanced real-time visibility into inventories, processes and supply chains, which will demand sensing and tracking at new size and cost points.

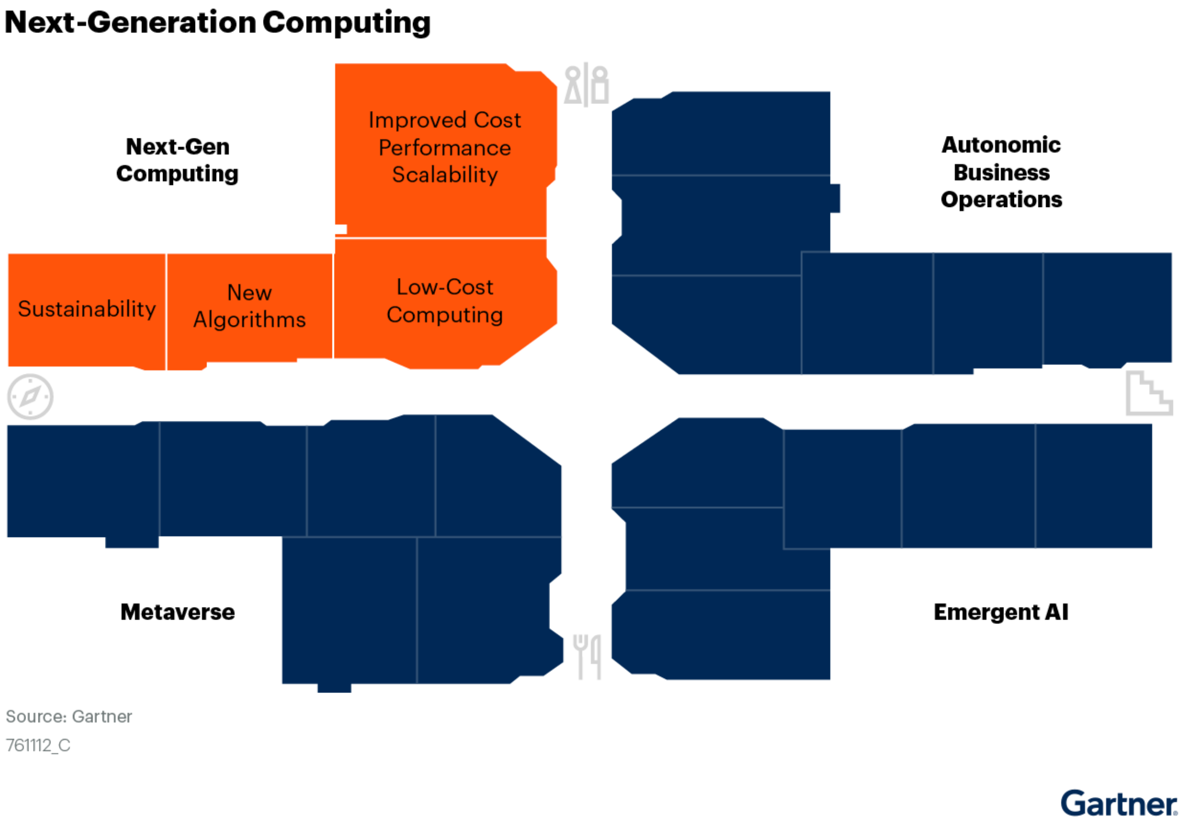
Recommendations

- Improve sustainability by adopting new energy-efficient hardware when the embodied carbon cost is outweighed by operational efficiency.
- Deploy special-purpose hardware accelerators such as GPUs as an alternative to general-purpose processors in selected compute-intensive applications.
- Initiate projects to identify opportunities using technologies such as edge supercomputing to implement new algorithms and capabilities, and consider pilot projects using cloud quantum systems to understand the technology.
- Identify key information shadows that could be illuminated using low-cost tagging and sensing, and pilot innovative tagging, sensing and manufacturing technologies to gain wider visibility into supply chains and processes and create new business value.

What You Need to Know

Next-generation hardware and software will be one of four key technology areas driving future digital business innovation, as illustrated in Figure 1.

Figure 1: Next-Generation Computing



The next generation of computing will be characterized by four areas of innovation:

- Increasingly resource-intensive tasks in areas such as artificial intelligence (AI) and simulation will require more computational power, more scalability, and better price and performance. Traditional hardware architectures won't be able to satisfy all the new requirements.
- A massive increase in Internet of Things (IoT) and smart devices will demand small, simple, low-cost devices such as sensors to operate under extreme power constraints.
- New business requirements will demand new types of computations, often impossible with traditional hardware. New algorithms will be required to exploit computing architectures such as quantum devices.
- Every organization's information and communication technology (ICT) activities will come under increasing scrutiny from a sustainability perspective, demanding attention to issues such as carbon footprint, e-waste and circularity.

The next generation of computing will involve a wide range of new technologies, algorithms and architectures. Some are available now, and some will emerge over the next five to 10 years to provide capabilities with new price or performance characteristics. The next generation of computing will deliver significant business benefits but will also pose challenges, as shown in Table 1.

Table 1: Next-Gen Computing Example Outcomes, Opportunities and Obstacles

(Enlarged table in Appendix)

	ICT Sustainability	Improved Cost/Performance/Scalability	Low-Cost Computing/Sensing	New Algorithms/Capabilities
Outcomes	Better-align ICT with the organization's sustainability goals	Obtain greater computing power at lower cost to drive innovation	Better visibility into processes and supply chains	Enable currently unachievable algorithms and storage capabilities
Opportunities	Improve all aspects of ICT sustainability	Deploy more sophisticated solutions such as real-time simulations and AI	Make a wide range of objects trackable and able to sense, at low cost	Implement radically new algorithms
Obstacles	Cost of replatforming, challenges of defining and measuring ICT sustainability	Skills availability, software and platform fragmentation	Changes to business processes, vendor immaturity	Skills availability, technology and vendor immaturity

Source: Gartner (August 2022)

Sustainability

Analysis by Nick Jones

Description

ICT sustainability has many dimensions; in the midterm (through 2025), the primary focus will be on the full life cycle carbon footprint of hardware and computation/storage, and, increasingly, circularity and e-waste. ICT sustainability is a complex issue, often with many alternative paths to the same goal, such as:

- Adopting algorithms and architectures that involve less computation and data transmission, and so consume less energy.

- Switching to more-sustainable energy supplies such as renewables without necessarily changing hardware, architecture or algorithms.
- Replacing old computing hardware with newer (more efficient) hardware. Hardware energy efficiency tends to improve significantly every few years, and new hardware is more likely to be designed with recyclability in mind to enhance circularity. Some manufacturers will also consider factors such as ethically sourced materials, for example, avoiding conflict minerals. However, replacing hardware incurs an embodied carbon cost and will typically be part of a wider initiative involving software and architecture improvements.

ICT sustainability is ultimately a business investment decision, which must be aligned with the organization's wider sustainability goals. However, for those organizations seeking sustainability improvements from more-efficient hardware, next-generation computing will offer many opportunities to improve ICT sustainability.

Relevant technologies include:

- Hardware accelerators typically use far less power than general-purpose processors for tasks such as AI training and simulation. Technologies such as GPU and FPGA are mainstream, as are certain types of AI accelerators. Research groups are working on other accelerators (see Note 1).
- Postsilicon computing, in long-term technologies such as nanotube transistors, is predicted to use 10% of the energy of silicon equivalents.
- Early optical computing equipment has demonstrated certain types of computations such as evaluating neural networks at 10% of the energy consumption of GPUs.
- Battery-free computing technologies such as energy harvesting reduce e-waste from battery disposal for small IoT devices. Some printed electronics technologies are even biodegradable, therefore reducing e-waste.
- Quantum computing is expected to be able to implement algorithms that are unachievable with any form of classical computer architecture, regardless of how much energy is available.

Why Trending

Sustainability is a board-level concern for many organizations. Gartner surveys show an increase in sustainability investments driven by customers, management, investors, regulators and other stakeholders (see Leading Sustainability Ambition, Goals and Technology in the 2020s). In addition, ICT vendors and service providers are being challenged to reimagine ICT for a net zero world/economy. Therefore, enterprise sustainability targets are trickling down to the IT organization where CIOs and CTOs are increasingly being asked to improve the sustainability of the organization's ICT activities.

Organizations such as the Green Software Foundation and the Responsible Computing consortium are working to raise awareness of software sustainability. The primary focus of most initiatives in 2022 is the energy consumption and greenhouse gas (GHG) footprint associated with ICT equipment and software. However, issues such as circularity, e-waste and ethically sourced materials are growing in importance along with broader sustainability considerations involving social and ethical considerations.

Implications

Implications include:

- Enterprise architecture and technology innovation leaders must understand the sustainability implications of the equipment they acquire, the architectures they deploy, the way systems are operated, the electricity they use, and the workloads they execute. New hardware incurs a cost in embodied carbon, which will influence when hardware replacement is the best sustainability tactic.
- Sustainability investment is a business decision based on its material impact on the organization's sustainability goals. Enterprise architecture and technology innovation leaders must make decisions in the context of the organization's overall sustainability targets to determine when investments in improvement are justified.

Actions

Enterprise architecture and technology innovation leaders should:

- Challenge ICT vendors to provide more sustainable solutions and update hardware purchasing processes and criteria to address areas such as energy consumption, circularity, sustainable materials and construction.
- Identify those ICT systems that consume significant energy in terms of having a material impact on the organization's sustainability targets.

- Consider replacing hardware when the embodied carbon acquired is outweighed by the reduced CO2 footprint of the applications.
- Favor new hardware that is designed to be upgradable or recyclable (which may not always be the case with novel devices from startups).

Further Reading

Leading Sustainability Ambition, Goals and Technology in the 2020s

Is Sustainable Software a Distraction or an Imperative?

2022 CEO Survey Research Collection

Cost/Performance/Scalability

Analysis by Arun Chandrasekaran

Description

Innovations in computing are altering the historical price/performance curves with significant gains in performance at lower costs. In addition, the advent of new computing paradigms such as quantum computing and neuromorphic computing represent a new era of computing architectures that will deliver significant performance benefits and enable the next set of innovative use cases.

Why Trending

The past few decades have seen exponential growth in processing power, mostly governed by Moore's Law, where computation became faster, more efficient and smaller. However, for emerging workloads, particularly in the field of artificial intelligence, there is a growing gap between system performance, heat emission challenges and increasingly expensive energy use. Organizations are increasingly seeking specialized compute that can bridge these gaps and yet deliver superior price/performance. Cloud providers have been innovating rapidly in this space along with a wide ecosystem of established chip manufacturers and startups. For example, AWS offers Inferentia, an inference chip, which delivers significant gains in price/performance when compared to GPU-based EC2 instances and AWS Trainium, a machine learning (ML) chip purpose-built for training deep-learning models. Google has been offering Tensor Processing Units (TPUs) with lower training costs for machine learning and a high degree of affinity to popular ML models such as TensorFlow.

Innovations in quantum computing and neuromorphic computing are noteworthy for the radical benefits they can deliver in the future. Quantum computing uses quantum mechanics to solve computational problems that are impossible to solve with current technologies. While it is still in early development, with progress depending on scalability of the number of quantum bits (qubits) and substantial decrease in error rate, it will be a viable technology for optimization problems before the end of this decade. Neuromorphic computing will engender purpose-built hardware needed to overcome the limits of CPUs and GPUs at potentially lower cost and with less energy — which may be particularly attractive in edge environments.

Implications

Next-generation computing technologies will enable further democratization of AI, allowing organizations to pursue high-impact and versatile use cases for AI. CTOs and innovation leaders will have a wider pool of compute technologies to choose from to balance their scalability, price and performance needs across a variety of distributed environments, including public clouds, data centers, edge and IoT devices. Building competencies to evaluate and pilot these solutions will be a primary bottleneck to leveraging emerging innovation. For this very reason, public cloud will continue to evolve as a preferred environment, given that many emerging compute innovations are from public cloud providers.

Actions

Enterprise architecture and technology innovation leaders should:

- Identify use cases where there are current constraints around price/performance, and pilot emerging AI training and inferencing services in the cloud and on-premises that can solve them.
- Develop a plan to hire the right talent to assess and pilot these technologies.
- Create a methodical approach for leveraging quantum computing with a timeline for assessment, piloting and mainstream adoption.
- Assess potential use cases for neuromorphic computing, particularly for edge ML inference, and plan for early pilot deployments in 2024 and beyond.

Further Reading

4 Advanced Computing Algorithms That Lead to Next-Generation Profits

2022 Strategic Roadmap for Compute Infrastructure

Emerging Technologies: Quantum Computing Planning for Product Leaders

Emerging Technologies: Neuromorphic Computing Impacts Artificial Intelligence Solutions

Tiny IT Enables Low-Cost Computing and Sensing

Analysis by Nick Jones

Description

Many innovative applications that could exploit sensing, location tracking or tagging are cost-sensitive or demand physically small devices with long battery life. A new generation of chips, wireless technologies and energy harvesting will drive innovation by enabling low-cost sensing, computing and communications with extreme resource constraints. Key technologies include ultra-low-power processors, energy harvesting, battery-free devices, and ultra-low-power wireless technologies, all delivered at increasingly affordable cost points. This combination of trends is sometimes known as “Tiny IT.”

Examples of products and academic research illustrating this trend include:

- Battery-free sensing and tagging technologies from companies such as Wiliot and Atmosic.

- The University of Michigan “Micro Mote,” which has been used to demonstrate computing and sensing attached to snails and a migrating butterfly.
- Northwestern University “flying microchip” sensors the size of a sand grain deployed using the principles of winged seeds.
- Backscatter technologies from companies such as Jeeva Wireless enabling battery-free sensing.
- Tiny rechargeable temperature sensors from companies such as CubiSens.
- Special-purpose chips for tasks such as low-power AI. These allow sophisticated processing such as neural networks to be integrated into sensors or small devices with long battery life.

Manufacturing techniques such as printed electronics will offer new ways to combine and deliver such technologies in application areas like sensing and smart packaging.

Why Trending

The key technologies underlying Tiny IT — low-power processing, wireless and sensing — have now matured to the point where they can be combined and packaged in ways that allow sensing in situations that were until recently unachievable. This aligns well with edge computing, which provides a convenient architecture to collect and consolidate information from large numbers of nearby sensors.

Implications

Tiny IT will create many opportunities in areas where low-cost, unobtrusive tagging and sensing can enhance currently invisible business processes or when a small amount of local computing enables new capabilities. Opportunities include better analytics, innovative products and the ability to collect information in currently impossible contexts.

For example:

- Smart packaging, integrating sensing (e.g., of temperature or acceleration) with supply chain monitoring for products such as food or pharmaceuticals.
- Large-scale, real-time inventory and asset monitoring, as an alternative to RFID.

- Sensing in currently challenging situations, such as in biological implants or in cases where previous sensors were too large or too power hungry (e. g., for research purposes).
- Disposable tactical sensing, such as sensors dropped from drones or aircraft for military or agricultural purposes.
- Consumer-facing products that need very-low-cost tagging and communications capable of simple Bluetooth interactions, such as for provenance tracking or to link physical products with apps and websites.
- Integrating low-cost chips with novel manufacturing techniques such as printable electronics to create new types of products like printed sensors.

Actions

Enterprise architecture and technology innovation leaders should:

- Educate business peers in the emerging technologies that can enable new types of sensing and tagging and new capabilities such as provenance-tracked products and brainstorm opportunities.
- Use brainstorming and ideation workshops to identify opportunities where real-time sensing, tagging and tracking could provide significant business benefits but, to date, are unachievable for reasons of sensor cost, size or battery life.
- Design internal wireless network deployments to enable them to be used as power sources in addition to communications.

Further Reading

Hype Cycle for Edge Computing, 2022

Hype Cycle for the Internet of Things, 2021

Emerging Technologies: The Future of Sensing

New Algorithms and Capabilities

Analysis by Arnold Gao

Description

Next-gen computing introduces a broader range of hardware architectures and technologies that enable new algorithms and capabilities, and can “make the impossible possible;” for example:

- Quantum computing can break the bottleneck of combinatorics and other calculations that are technically or commercially impossible by existing classical computing architecture.
- DNA storage will theoretically enable thousandfold increases in data density compared to hard-disk drive (HDD), solid-state drive (SSD), tape, or optical enterprise-grade media, allowing 10 zettabytes (ZB) of data to be stored in a shoebox. Data in DNA storage can endure thousands of years and remain unchanged, free from degradation or drive failure compared to current technologies.
- Extreme parallelism makes supercomputing possible by orchestrating hundreds of thousands of processors or cores operating in parallel.
- Emerging optical computing can speed up neural-network-related calculations (e.g., matrix multiplication) to empower low-latency AI computations for time-critical use cases that must respond to real-time changes to their environment.
- Application-specific chips are the best alternatives for some tasks such as machine learning and privacy-preserving homomorphic computing, which are not well-suited for general-purpose chips (e.g., CPU).

Why Trending

Most computing and storage solutions today based on general-purpose processors and traditional media are increasingly constraining the evolving needs of innovations. Tasks, like complex optimizations and massive big data storage, demand algorithms and capabilities that are not available from current architectures. Next-gen computing technologies address this and include new computing architectures, new computing methods and storage technologies needed for technology innovation in the next five to 10 years.

Implications

The new hardware architectures in next-gen computing will transform the current computation paradigm by enabling new digital capabilities and algorithms, including:

- New quantum optimization algorithms in the areas of complex manufacturing, supply chains and financial portfolios management will be available in the next five years. Other disciplines such as cybersecurity, bioengineering and materials science will be disrupted by quantum computing in the longer term.
- Life-long or permanent retention of hyperscale big data with virtually no maintenance or cost, which can be accessed without the need for data migration, will be possible.
- Resource-intensive tasks such as simulation and AI training will not only be processed in the cloud, but also at the edge to overcome bandwidth and latency challenges.
- Autonomous driving, drone operations and robotic surgery empowered by low-latency neural network computations in the cloud, at the edge and in the local data centers or local server rooms in buildings will emerge.
- Competitive collaboration among multiple participants by processing and analyzing sensitive data that needs privacy protection (e.g., healthcare data, financial transactions) will be possible.

Actions

Enterprise architecture and technology innovation leaders must:

- Investigate the maturity of each technology that will bring breakthroughs to the current computation paradigm.
- Identify challenging innovations that cannot be accomplished by current architectures and try new emerging techniques with proof of concept projects.
- Anticipate difficulties in the utility of new hardware and solutions, but prioritize early use cases to focus near-term success.
- Prepare for the long-term disruptions by implementing innovation labs/hackathons that enable skills development and to discover opportunities.

Further Reading

Predicts 2022: 4 Technology Bets for Building the Digital Future

Emerging Technologies and Trends Impact Radar: Compute and Storage, 2021

Emerging Technologies: Critical Insights on AI Semiconductors for Endpoint and Edge Computing

4 Advanced Computing Algorithms That Lead to Next-Generation Profits

Evidence

The information in this research has been obtained from discussions with peers and clients, academic papers, and Gartner surveys shown in the research listed in the Further Reading sections.

Note 1. Application-Specific Accelerators

Several silicon manufacturers have produced chips optimized for AI tasks. Examples include Intel's neuromorphic computing and Syntiant's low-power neural network chips. Other types of special-purpose chips are also under development, for example, to accelerate extremely large decision tables. Some of these devices combine analog and digital technologies, as the former can be much more power-efficient for certain types of processing.

Table 1: Next-Gen Computing Example Outcomes, Opportunities and Obstacles

	ICT Sustainability	Improved Cost/Performance/Scalability	Low-Cost Computing/Sensing	New Algorithms/Capabilities
Outcomes	Better-align ICT with the organization's sustainability goals	Obtain greater computing power at lower cost to drive innovation	Better visibility into processes and supply chains	Enable currently unachievable algorithms and storage capabilities
Opportunities	Improve all aspects of ICT sustainability	Deploy more sophisticated solutions such as real-time simulations and AI	Make a wide range of objects trackable and able to sense, at low cost	Implement radically new algorithms
Obstacles	Cost of replatforming, challenges of defining and measuring ICT sustainability	Skills availability, software and platform fragmentation	Changes to business processes, vendor immaturity	Skills availability, technology and vendor immaturity

Source: Gartner (August 2022)