Building an Edge Computing Strategy

Published 3 September 2021 - ID G00753920 - 20 min read

By Analyst(s): Thomas Bittman

Initiatives: Cloud and Edge Infrastructure

Edge computing projects are emerging in different parts of enterprises. To accelerate deployments, ensure extensibility, and drive efficiency and effectiveness, I&O leaders must create an edge computing strategy.

Additional Perspectives

- Summary Translation: Building an Edge Computing Strategy (26 September2021)

Overview

Key Findings

- Edge computing can be used to accelerate an enterprise's digital business strategies, but there is little consistency, understanding or expertise about how to do that.
- Disjointed edge computing projects are appearing in many different parts of enterprises, designed to fill a specific need, but without a guiding strategy to ensure enterprise efficacy, efficiency and leverage.
- Edge computing projects have common challenges that require invention to solve — but solutions are usually siloed and rarely shared across projects.
- Edge computing deployments are often designed to deliver a single use case or workload, but inevitably the requirements expand, and lack of extensibility is a key inhibitor.

Recommendations

I&O leaders responsible for cloud and edge infrastructure projects should:

- Create a vision for edge computing by linking up with business and other technology strategies, and communicating a measurable goal for edge computing.

- Identify existing and potential use cases for edge computing through proactive collaboration with business units and a clearinghouse to identify emerging use cases.
- Accelerate expanded deployments of edge computing by identifying common challenges and risks, and putting mitigation plans in place.
- Improve enterprise adoption of edge computing by building and communicating standards, best practices, guidelines and skills requirements.
- Ensure success of edge computing projects through a structured plan to manage unique edge computing requirements from proofs of concept (POCs) through production.

Strategic Planning Assumptions

By year-end 2023, 50% of large enterprises will have a documented edge computing strategy, compared to less than 5% in 2020.

Through 2025, 50% of edge computing solutions deployed without an enterprise edge computing strategy in place will fail to meet goals in deployment time, functionality and/or cost.

By year-end 2023, 20% of installed edge computing platforms will be delivered and managed by hyperscale cloud providers, compared to less than 1% in 2020.

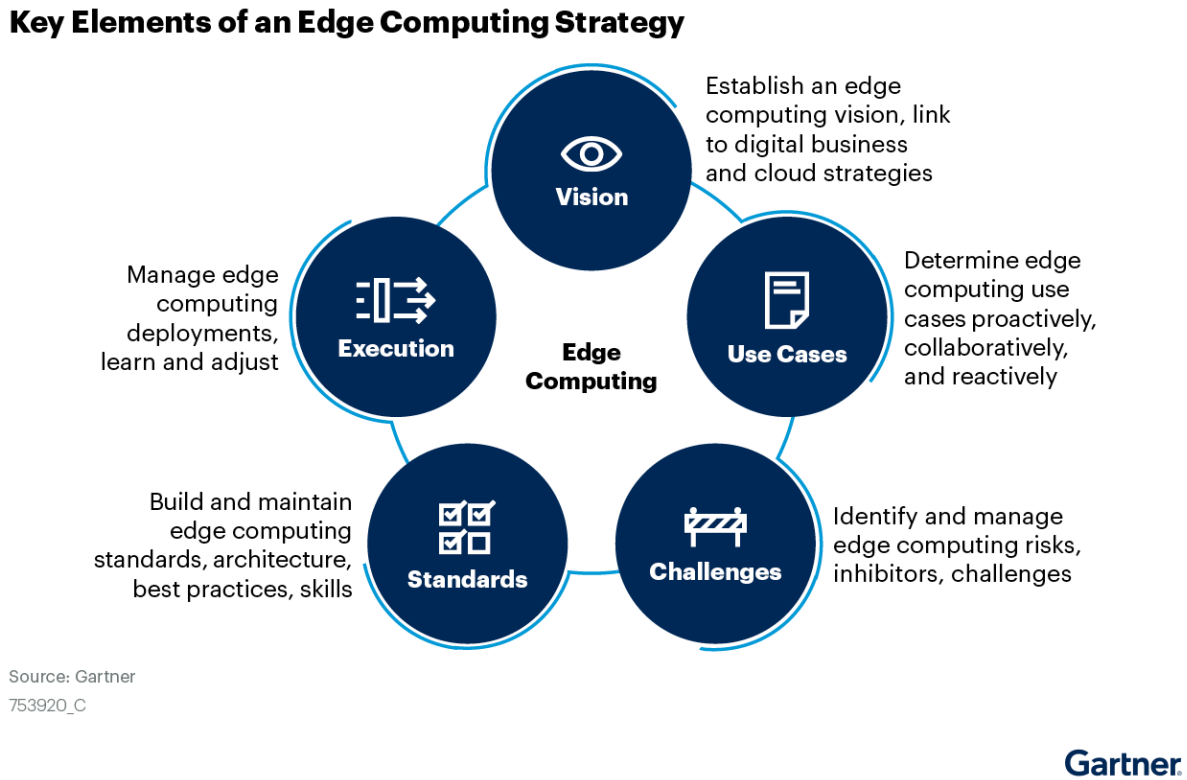
Introduction

Edge computing projects are increasing, based on inquiry discussions.

Edge computing is part of a distributed computing topology where information processing is located close to the edge (the physical location where things and people connect with the networked digital world).

Often, edge computing projects are deployed independently, as custom solutions focused on a specific requirement, a specific use case or a specific part of the enterprise. However, enterprise experience has shown that edge computing projects multiply independently in different parts of the business, or expand from a single use case in a project to several use cases. Diversity in use cases is the norm in edge computing. But as the edge computing trend grows, synergy across projects and extensibility to enable new use cases is critical in overcoming challenges, creating standards, choosing technologies and managing costs. Edge computing should be pursued in a stepwise manner, but strategically (see Figure 1). Enterprises should be proactive about their strategies because, if they aren't, by default they're following the strategy determined by specific vendors, which will be optimized to vendor offerings, rather than enterprise requirements, cost and flexibility.

Figure 1. Key Elements of an Edge Computing Strategy



There are five important elements to an edge computing strategy, designed to simplify, synergize and systematize edge computing projects.

Analysis

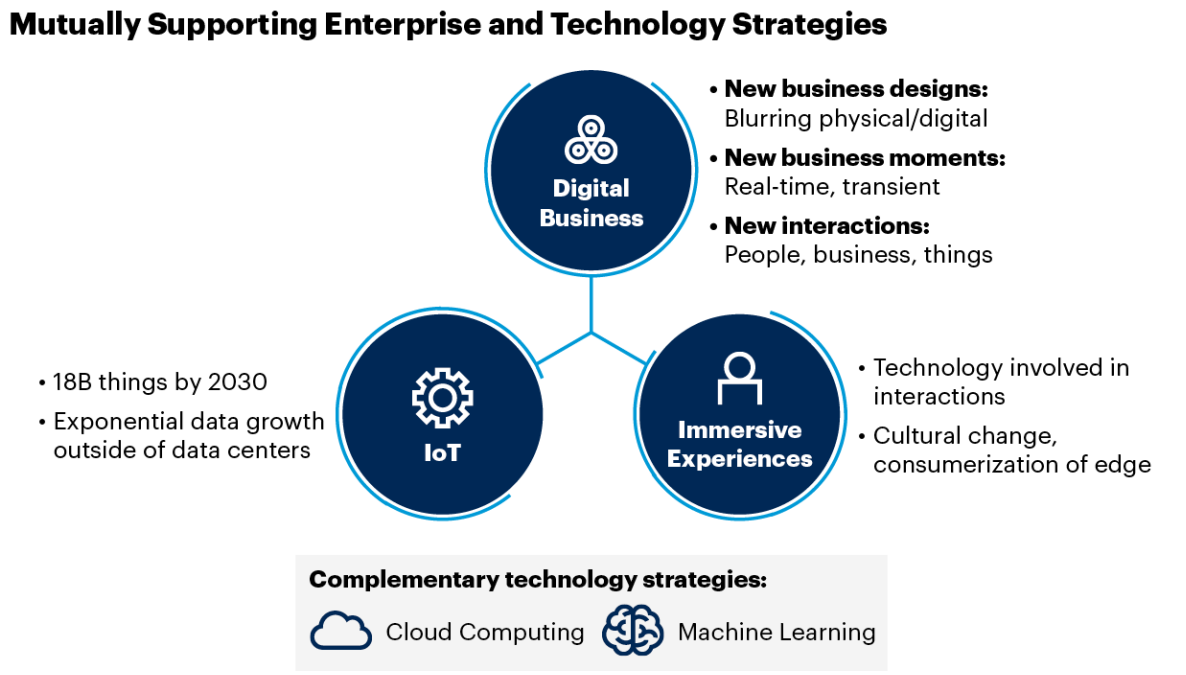
Create an Edge Computing Vision

What is edge computing, and why are you doing it? An edge computing vision states specific business outcomes the enterprise can enable. An edge computing strategy should be linked to other business and technology strategies (see Figure 2). It is critical to get stakeholder buy-in and engagement.

Enterprise Strategies and Objectives

Edge computing will be integral to digital business transformation by enhancing and expanding digital and physical interactions at the edge — where people and things are located and interacting — and expanding the use of digital information to make faster and better decisions. An edge computing strategy should be linked to the overarching enterprise digital transformation strategy.

Figure 2. Mutually Supporting Enterprise and Technology Strategies



Source: Gartner
753920_C

An edge computing strategy should also target specific business objectives. Edge computing can be used to deliver automation, improve efficiency or enhance quality, all of which help to improve an enterprise's bottom line. However, through the creation of net new business interactions and agile use of information, edge computing is also being used to improve customer experience and personalization or to create new business opportunities. Based on Gartner client inquiry, about half of edge computing use cases are targeting bottom-line efficiency, while the other half are focused on the top line and growth.

An edge computing strategy should be explicitly linked to business objectives (cost, quality, resilience, growth, agility, innovation, customer experience, etc.).

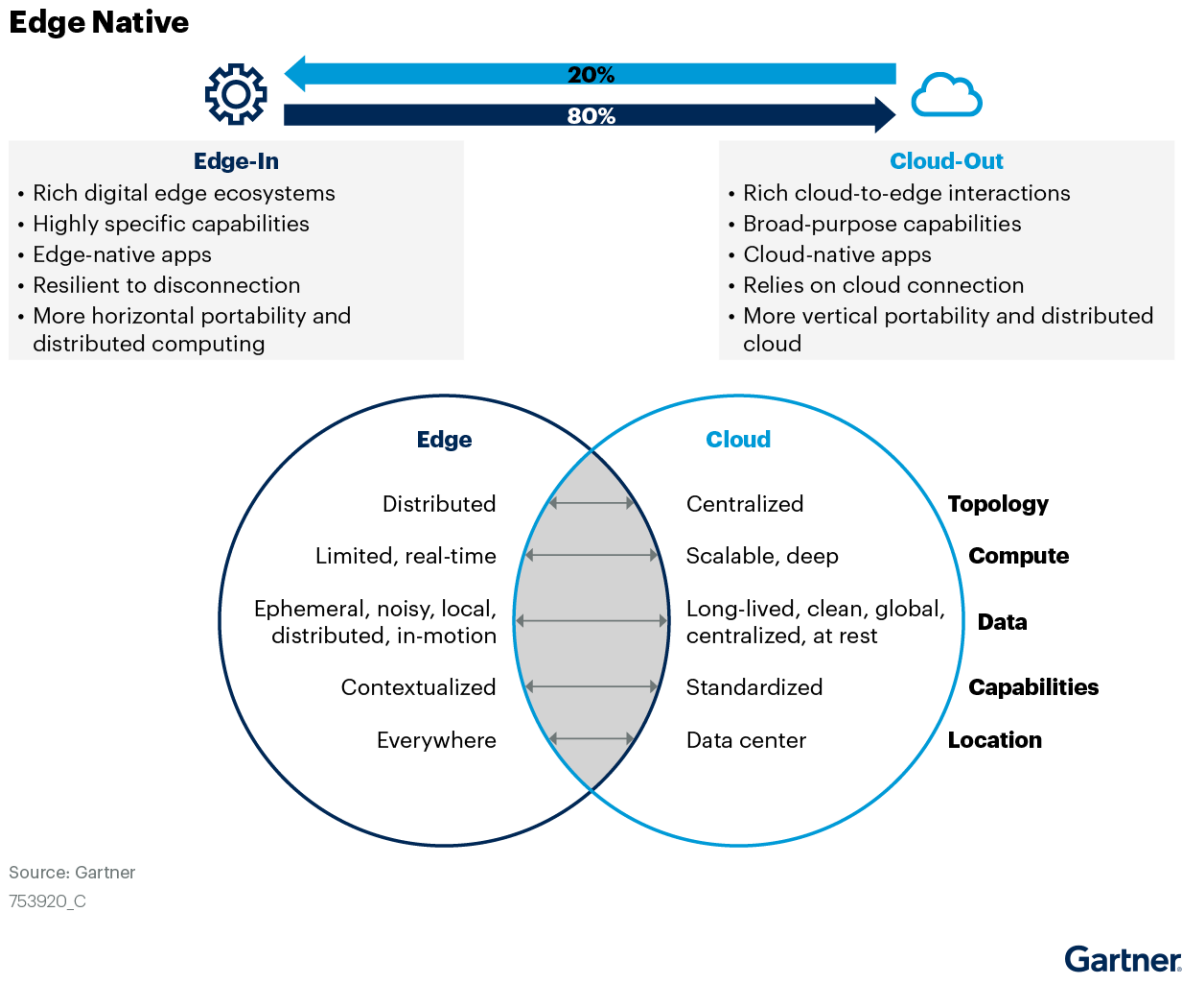
Technology Strategies

Edge computing is part of a distributed computing topology where information processing is located close to the edge (the physical location where things and people connect with the networked digital world). But its value is in complementing and enabling other technology strategies.

For example, cloud computing is a style of computing in which scalable and elastic IT-enabled capabilities are delivered as a service using internet technologies. In practice, cloud computing is about taking your IT requirements to providers' hyperscale data centers. Edge computing does the obverse — it brings computing capability to your IT (and OT) requirements, when that's necessary. Edge computing complements (and generally requires) cloud computing, integrating computing at or near the edge with back-end data center processing. While edge computing is location-specific, many cloud attributes can still be applied to edge computing solutions, such as continuous integration and continuous deployment, DevOps, "as a service" hardware management, and operating expense pricing. An edge computing strategy should be linked to the enterprise cloud computing strategy.

On the other hand, while consistency is important, edge computing and the requirements, technologies and capabilities at the edge are different than in hyperscale data centers (see Figure 3). "Cloud native" implies exactly that: native to cloud computing, typically in hyperscale and homogeneous data centers. "Edge native" is very real; where data is distributed and often ephemeral, workloads tend to be real time with limited scalability options, and processing location is not abstracted, it's a defining characteristic.

Figure 3. Edge-Native



While cloud computing complements edge computing, edge use cases are either approached from a cloud-out perspective (i.e., extending cloud capabilities and services closer to the edge) or edge-in (i.e., focusing on rich edge ecosystems and integrating to cloud services). As highly diverse edge systems evolve, the role of hyperscale providers will focus on centralized value. By year-end 2023, 20% of installed edge computing platforms will be delivered and managed by hyperscale cloud providers, compared to less than 1% in 2020.

Edge computing also enhances IoT. In many ways, Phase 1 of the IoT was simply about digital connection: digitizing data, getting it to a data center and sending actions from the data center. As the amount of data from things continues to grow, and as requirements for lower-latency digital processing grow, relying on back-end data centers is insufficient. Edge computing provides the preprocessing or low-latency local processing for growing IoT deployments. Further, as the interactions between people and things at the edge multiply and become complex local systems, more of the digital processing needed is local and latency is defined as the latency of complex systems, not individual transactions.

Immersive technologies are being used in a variety of ways to create the future of experience, for employees and for customers. Immersive technologies will range in capability from headsets to smartphones (doing augmented reality [AR]/virtual reality [VR]) to intelligent speakers (providing a more natural digital interface). As systems of interaction and collaboration at the edge continue to grow, computing for immersive experiences will be distributed between immersive technologies themselves, the cloud and edge computing.

Enhanced automation and intelligence at the edge will rapidly expand the use of machine learning. While inference is usually implemented at the edge, training is often done by sending massive datasets to central processing. However, as the cost of compute continues to decline compared to the cost of bandwidth, training may also be done closer to the edge where the inference models are deployed. In the future, more training might be distributed in a federated or hierarchical way. Edge computing will be a part of deploying inference models and a part of training. And edge computing requirements will push the requirement for machine learning applications to be easy to deploy and easy to use.

An edge computing strategy should be linked to a variety of technology strategies an enterprise is pursuing, including cloud computing, IoT, immersive experiences and machine learning, among others.

Edge Computing Vision and Leadership

The purpose of a vision is to provide an understandable target state that can help motivate and direct the team internally, and also that can be used to present measurable results to the enterprise. How will the organization operate differently with edge computing in five years? What new capabilities will be enabled? The edge computing strategy is linked to enterprise and technology strategies — what will they look like in five years?

A vision for edge computing could include:

- An objective business impact, such as a percentage of digital business initiatives that include edge computing, net new business transactions enabled by edge computing, amount of money saved through edge computing initiatives, etc.
- Specific goals for edge computing use in the office, the factory, the store or branch.
- Percentage of customer interactions leveraging edge computing.
- Number of automation projects completed.
- Range of types of use cases deployed.
- Deployment agility, i.e., the number of POCs or time to deployment.

Key to success of an edge computing strategy will be executive buy-in and sponsorship. For a strategy to be meaningful, there needs to be executive and staff support, and the vision needs to be well communicated. As the strategy changes — and an edge computing strategy will evolve as technologies and use cases emerge — updates to the strategy and vision also need to be well communicated.

Recommendations

I&O leaders should:

- Create a vision statement for edge computing, communicate it, and ensure executive buy-in, sponsorship and alignment with staff.
- Create linkages between the edge computing strategy and the overall enterprise strategy and objectives, especially those related to digital business.
- Link the edge computing strategy to other technology strategies, specifically cloud computing, IoT, immersive technologies and machine learning.

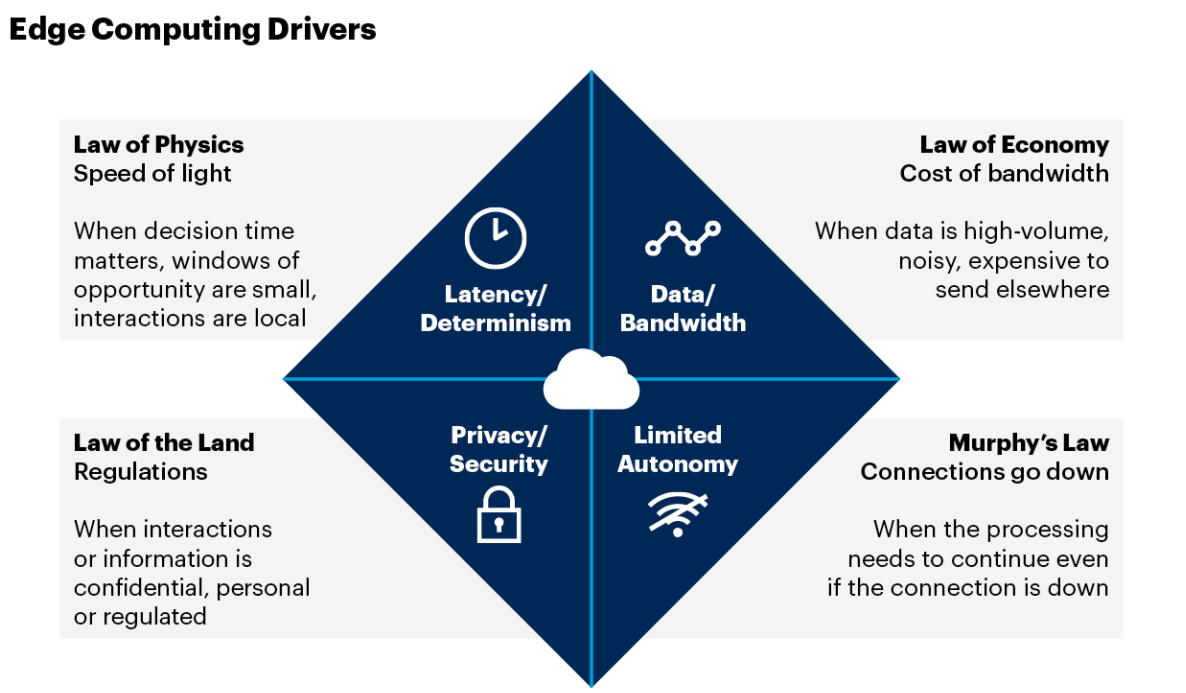
Identify Edge Computing Use Cases

Edge computing use cases will be highly diverse, in diverse parts of the enterprise, with diverse objectives. An edge computing strategy should include an understanding of edge computing drivers, requirements and existing deployments. It should also have a process for proactively finding new use cases and correctly identifying use cases as they emerge.

Edge Computing Drivers

Edge computing fills a gap that other computing architectures can't fill. If a business problem can be solved efficiently and effectively with a cloud computing or stand-alone solution that should be the default. An edge computing strategy should define when an edge computing solution should be pursued. There are four main drivers (see Figure 4), and every use case may require one or more of them.

Figure 4. Edge Computing Drivers



Source: Gartner
753920_C

Gartner

- **Latency/Determinism:** This refers to when the response time needed for processing requires that the compute be located physically close. In addition to low latency, the use case might require deterministic, predictable response time (for example, at the speed of an assembly line). Beyond individual transactions, there could be a number of people or things interacting in sequence, and the latency of that overall system might be the issue. Latency requirements can be addressed at different points in an edge computing topology using:
  - Specific cloud regions
  - Local data centers or content delivery networks (CDNs)
  - 5G connections and multiaccess edge computing (MEC) capability
  - Micro data centers
  - Local edge servers
  - Intelligent gateways
  - Embedded processing
  - Or some combination
- **Data/Bandwidth:** As data at the edge grows in volume, the cost of sending noisy, ephemeral data elsewhere to be filtered or processed might be more than moving the compute to the data. Edge computing will be used for preprocessing, for filtering or to completely handle certain edge transactions. Most of the data produced at the edge will never leave the edge.
- **Limited Autonomy:** Edge computing might be required to maintain a working environment even when the connection to the central data center or cloud service goes down or is unavailable for some time. This could include queuing transactions for later processing, or it could be essentially a subset of the data center or cloud capability that runs at the edge.
- **Privacy/Security:** As more and more intimate raw data about interactions and the things and people in a location are digitized, that data might become regulated (e.g., facial recognition data). Regulatory requirements might be different for different edge locations. In many cases, enterprises will self-regulate to keep certain data and interaction data on-premises or on-device.

An edge computing strategy should explicitly discuss the requirements that will be addressed by edge computing specific to enterprise requirements.

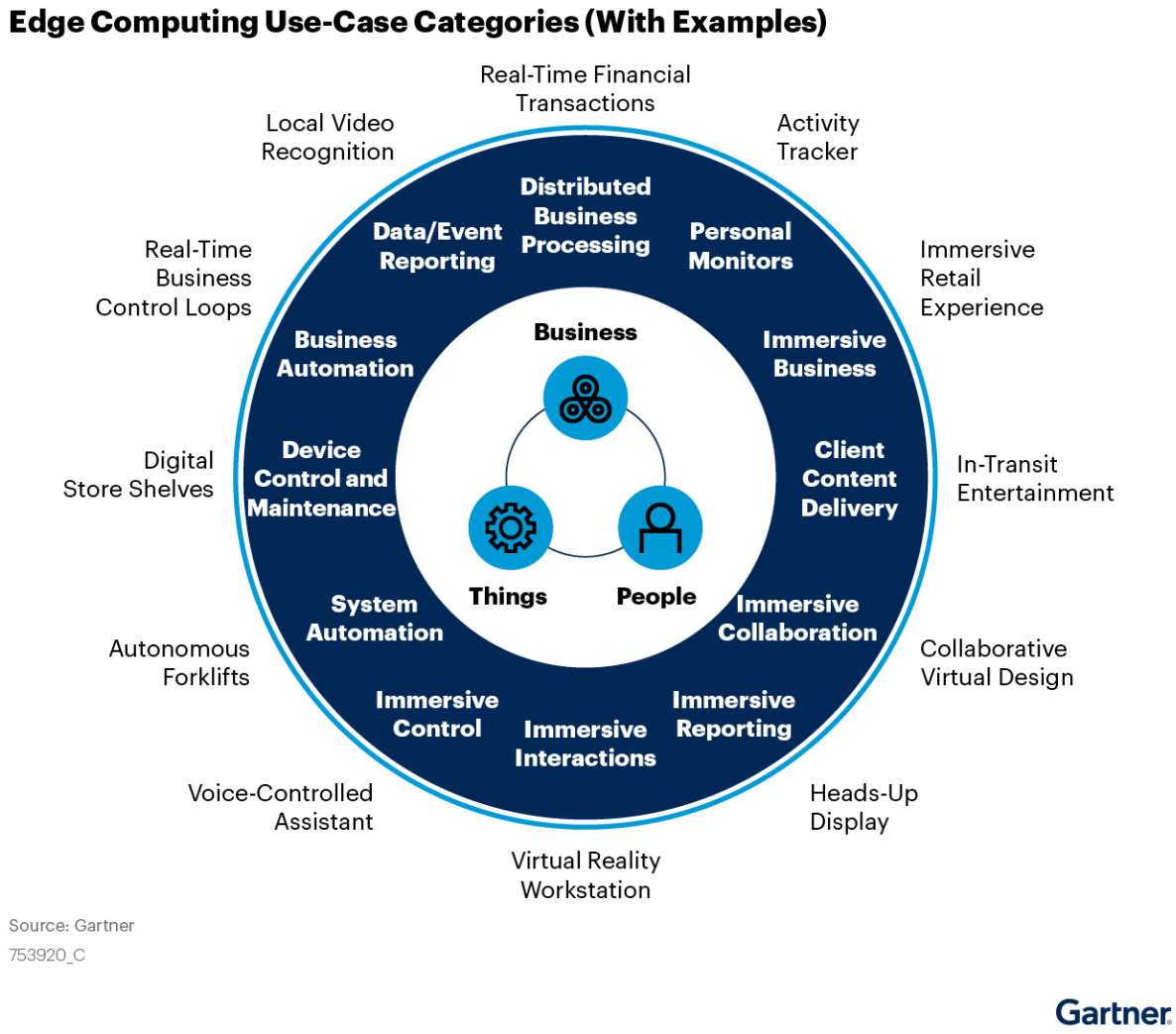
Existing Edge Computing Use-Case Landscape

Before there was an edge computing strategy, the enterprise encountered — and dealt with — some edge computing use cases. On the manufacturing floor, this would include existing OT (operational technology), programmable logic controllers (PLCs), and robotics. These are often known as embedded systems. In the store, it could include smart check-out. An edge computing strategy should identify all of the existing deployments that fit under its purview. The strategy should include how changes and new investments in those deployments should be done to move them closer to chosen edge computing standards, guidelines, technologies and best practices.

Potential Edge Computing Use Cases

An edge computing strategy should establish a proactive process to identify potential use cases. This needs to be collaborative between IT and the various business units. There should be well-defined linkages between the edge computing and various enterprise strategies, a clear vision and well-understood drivers. IT and business units should work together to find the art of the possible; but the goal would be to find at least a single prospect use case that directly supports the outcomes expressed in the edge computing vision. There are 12 different categories of edge computing use cases, based on enhancing interactions between business, things and people (see Figure 5).

Figure 5. Edge Computing Use-Case Categories (With Examples)



Edge Computing Use-Case Clearinghouse

An edge computing strategy should include creating a clearinghouse for new use-case ideas, that is, structured processes for how new use cases and edge workloads are identified, vetted and prioritized. Linkages to enterprise digital business strategies and technology strategies for cloud computing, IoT and others will facilitate the identification of edge computing candidates. Edge computing should be considered whenever latency, edge data, semiautonomy or location privacy requirements cannot be solved by traditional enterprise data center or cloud-based solutions.

Recommendations

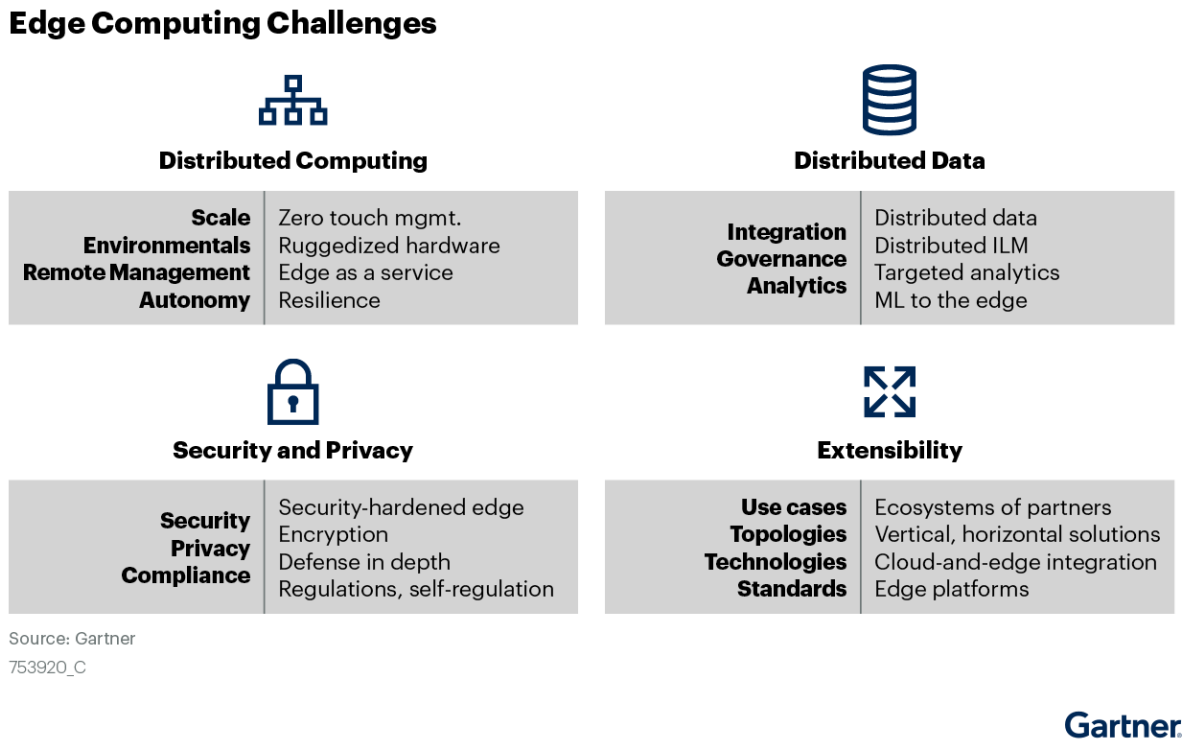
I&O leaders should:

- Identify the specific requirements that edge computing can address for the enterprise in the areas of latency, data, semiautonomy and privacy.
- Identify existing technologies and deployments that should be included within the purview of the edge computing strategy.
- Identify potential use cases that could be addressed with edge computing, proactively and collaboratively with business units.
- Create a use-case clearinghouse to identify edge computing candidates as they emerge, with structured processes for how new use cases and edge workloads are identified, vetted and prioritized.

Focus on Edge Computing Challenges

Edge computing creates risks that need to be mitigated and new challenges that need to be overcome. An edge computing strategy needs to maintain a focus on them. Different industry verticals may have unique edge challenges, or risks that need to be mitigated. However, there are four challenges with edge computing that are applicable to the vast majority of enterprises (see Figure 6).

Figure 6. Edge Computing Challenges



Distributed Computing

Edge computing can be highly distributed in locations that have heat, air quality and power challenges, with limited to zero IT skills. The number of nodes can be extreme, for example, every store, throughout factories, in automobiles and homes. Different geographies may have different environmental issues, varying network access, or challenges in local skills and services availability. And different locations may have different regulatory requirements. Resilience requirements may demand hardware, software and networking that can ensure continuous operation, or processes to rapidly replace bad equipment. Remote management and monitoring and zero touch management are often essential, and a variety of edge-as-a-service offerings might be important. As requirements change — software changes, new hardware requirements — the enterprise needs to be able to adapt.

Security and Privacy

The distributed scale and lack of traditional physical data center security in edge computing put a greater burden on ensuring that edge computing nodes are hardened and that the enterprise is protected through defense in depth. Edge computing architectures need to be protected from tampering, data rerouting, hijacking or denial-of-service attacks. In addition, the data being captured may introduce new legal, regulatory and ethical considerations that determine how (and if) data is stored and encrypted and where data can be sent. As time goes on, workloads may change, capturing and processing new kinds of data and creating new challenges.

Distributed Data

As more things produce data at the edge, and as more interactions at the edge become digital, data growth outside of data centers will become significant. Edge computing creates a massive distributed data challenge — distributed governance, distributed data integration, and distributed data and analytics architectures. The nature of data at the edge is quite different from data in data centers. Data tends to be more ephemeral, noisy and locally specific. Unlike data centers, where data management is primarily about data preservation, a challenge at the edge is to destroy useless data as quickly as possible (to avoid expensive storage and minimize governance risks). But determining data value dynamically and not destroying data that has potential value are real challenges.

Extensibility

In order to meet specific requirements but keep costs down, edge computing nodes will typically be very purpose-specific to their environment and workloads. However, the workloads in a location will change, and new ones will emerge. Edge computing platforms need to be both special-purpose and extensible. An edge computing strategy should attempt to predict future workloads and enable others beyond the planning horizon. A typical enterprise may have a number of "edges," in offices, equipment and factories. At the same time, when it makes sense, common challenges should be solved with common technologies and processes, rather than having entirely distinct management tools across heterogeneous edge nodes, for example. In an emerging space like edge computing, new technologies and vendors will arise, and dependence on specific vendors drives lock-in. An edge computing strategy will need to manage partners and include exit plans.

Recommendations

I&O leaders should:

- Identify the risks, challenges and inhibitors that need to be overcome and mitigated, and put special ongoing focus on those challenges in terms of management, investment and skills.

Build and Communicate Edge Computing Standards

Diversity is a defining characteristic of edge computing use cases, but that makes it even more important to find and maintain synergies by leveraging technologies, platforms, best practices, standards, processes and skills across disparate deployments. Many enterprises have a cloud center of excellence (CCOE). A CCOE is a centralized enterprise architecture function that leads and governs cloud computing adoption within an organization. The CCOE could be expanded to include edge computing, or a separate center of excellence could be created for edge computing. There are three main pillars of an edge center of excellence.

Edge Computing Technologies

An edge computing strategy should lay out and maintain a set of technology and architecture standards, frameworks, topologies and delivery models that will be used across edge computing deployments. An enterprise should maintain a list of trusted vendors (and a process for analyzing vendors).

The technology and architecture areas covered should include at least:

- **Cloud integration:** Applications, data, networking, etc.

- **Delivery models:** Edge-as-a-service options for hardware, software, service providers, etc.
- **Hardware:** Ruggedized, fit to purpose, composable, extensible, etc.
- **Application/software platform:** Application architecture, development tools, hypervisor, container architecture, etc.
- **Management tools:** For hardware, software platform, applications, etc.
- **Data management:** Storage, governance, metadata, etc.
- **Data analytics:** Traditional reporting and business intelligence (BI), advanced analytics, artificial intelligence (AI), etc.
- **Security technologies:** All layers including hardware, software, data, encryption, etc.
- **Networking:** Connectivity to devices, to the cloud, etc.
- **IoT technologies:** IoT endpoints, platforms, device management, stream processing, security, etc.

Edge Computing Governance

An edge computing strategy needs to lay out policies for governance and security of edge computing deployments, software and data, and measure the value and return on investment. This includes standards for leveraging as-a-service providers (which could be for any or all layers of the edge computing stack). Policies, guidelines and guardrails should be put in place that can be used to guide all future edge computing deployments and how the edge computing technologies will be used (and when they won't be used). An edge computing strategy must also factor in compliance requirements that might be different in different locations, but also potential legal and ethical considerations (for example, how facial recognition data is stored and used).

Edge Computing Best Practices and Skills

An edge computing strategy should include how best practices both outside the enterprise and from enterprise deployments will be captured and leveraged. Since edge computing is so diverse and so new, lessons learned will be extremely valuable to accelerate successful deployments and avoid duplicate efforts.

The strategy should also include edge computing skills identification, development, roles and responsibilities, and organization and matrix organization structure.

Recommendations

I&O leaders should:

- Create an architecture function focused on edge computing (e.g., “edge center of excellence”).
- Build, maintain and communicate technology and architecture standards, frameworks and topologies that will be used across edge computing deployments.
- Capture and maintain best practices both outside the enterprise and from enterprise deployments.
- Identify and develop edge computing skills, roles and responsibilities, and organization structure.

Ensure Success of Edge Computing Execution

Because edge computing is an emerging concept where “firsts of a kind” predominate, enterprises will often be pioneering new technologies and new uses of technologies, and learning along the way. There are two unique aspects of edge computing that need to be considered in an edge computing strategy.

Edge Computing POCs

Edge computing deployments will need to be piloted for efficacy, manageability, autonomy and scale. The most common reason that edge computing POCs tend to fail in production is inability to efficiently scale (in terms of number of nodes and number of locations). A pilot should attempt to evaluate how a project will operate and how it can be monitored and managed at full scale (using vendor edge computing labs, simulations, or evaluating similar reference customers). In addition, the ability to handle disconnection needs to be tested in a variety of ways. The edge computing strategy should lay out defined steps for deployments from POCs through production, including centralized coordination, measures of success and the capture of lessons learned.

Evolution Management

Deployments of edge computing evolve during development, and often evolve and greatly expand after they’re in production. An edge computing strategy should have guidelines and guardrails in place for how to monitor and manage changes in a deployment’s requirements, workloads, data, scale and uses over time. The strategy should guide those evolutions — and in some cases, the evolutions should cause the strategy to change.

Recommendations

I&O leaders should:

- Require POCs to test (or simulate) unique edge computing requirements, such as scale and tolerance to disconnection.
- Build a process to manage the evolution of an edge computing deployment in production, especially in terms of how workloads, data, scale and uses change over time.

Evidence

This research is drawn from hundreds of inquiries with Gartner clients, including those beginning edge computing deployments, those in the middle of the process, and those that have completed a rollout.

Recommended by the Author

Some documents may not be available as part of your current Gartner subscription.

Infographic: Understanding Edge Computing

Hype Cycle for Edge Computing, 2021

2021 Strategic Roadmap for Edge Computing

Predicts 2021: Cloud and Edge Infrastructure

Why and How I&O Should Lead Edge Computing

Emerging Technologies: Emerging Edge AI Use Cases in Retail
