Expanding Opportunities at the Edge with Private Cloud

Pathfinder

August 2021

Commissioned by

LUMEN®

vmware®

451 Research

S&P Global
Market Intelligence

About this paper

A Pathfinder paper navigates decision-makers through the issues surrounding a specific technology or business case, explores the business value of adoption, and recommends the range of considerations and concrete next steps in the decision-making process.

About the Authors

Rich Karpinski

Senior Research Analyst, Voice of the Enterprise: Internet of Things

Rich Karpinski is Senior Research Analyst for the Voice of the Enterprise: Internet of Things (IoT) survey and advisory offering at 451 Research, a part of S&P Global Market Intelligence. In this capacity, Rich tracks, analyzes and anticipates the pace and direction of IoT adoption, overseeing 451's quarterly survey of IoT adopters and twice-annual survey of operations technology (OT) professionals. As a member of 451's IoT team, Rich also is responsible for smart city and smart spaces coverage. In addition to that role, Rich closely tracks the evolution of mobile operator business models – including IoT and other digital service strategies – and works with 451's mobile operator customers to help them understand market trends and plot strategies.

Melanie Posey

Research Director, Cloud & Managed Services Transformation

Melanie Posey is the Research Director for the Cloud & Managed Services Transformation at 451 Research, a part of S&P Global Market Intelligence. In addition to managing the research team, she focuses on analyzing the evolution of enterprise IT through the lens of cloud and the associated transformation of IT consumption and delivery models. Melanie also manages 451 Research's Voice of the Enterprise: Cloud, Hosting & Managed Services offering.

Executive Summary

Many of the most critical industry workloads have requirements – including performance, low latency, security and data sovereignty – that current infrastructure options, both on-premises and cloud-delivered, are challenged to fully meet. On-prem options can deliver nearby compute and IT familiarity but lack the on-demand flexibility of the cloud. The cloud offers those benefits but can be challenged to meet high-performance, low-latency demands of modern organizations. In this 451 Research Pathfinder, we explore the emergence of an edge private cloud that can address those concerns and power mission-critical workloads, applications and use cases, at scale.

Key Findings

- **More organizations are taking advantage of edge compute.** Fifty-eight percent of enterprises said they are building new edge compute capabilities in 2021 to support their critical industry workloads.
- **Companies deploying IoT workloads are particularly edge-centric.** Such enterprises reported that an average of 51% of their IoT-supporting infrastructure is deployed at the edge (i.e., outside of core datacenters) and that 53% of their digital workloads are currently deployed in edge or near-edge locations.
- **Enterprises turn to edge compute to support a range of mission-critical requirements.** Enterprises consider cost (cited by 45% of enterprises), availability of supporting compute, storage and connectivity (44%), data sovereignty protection (42%) and location security (40%) as their top concerns when choosing an execution venue – requirements well-satisfied at the edge. Many digital workloads also require low-latency (28% of respondents) and highly resilient (27%) infrastructure that can often be best – if not exclusively – handled at an edge venue.
- **They also believe the edge has a major impact on their digital success.** Ninety-two percent of enterprises somewhat or strongly agreed that the location in which an application is deployed has a 'major impact' on their IoT project success.
- **The preferred infrastructure model for digital workloads is a hybrid approach – neither purely on-premises or cloud, nor totally edge or centralized.** Rather, enterprises are choosing the right venue for the job and increasingly orchestrating digital workloads across a multi-tier architecture. In all, 78% of organizations said they have or plan to have such a hybrid IT architecture in place in 2021.
- **An edge private cloud, combining key attributes of both an on-prem edge and managed cloud infrastructure, offers a strong option for digital workload execution.** Such an approach offers the dynamic, cost-advantaged characteristics of the cloud with the secure, low latency, high performance of an edge venue – key requirements for enterprises supporting critical industry edge use cases.

The Critical Workloads Challenge

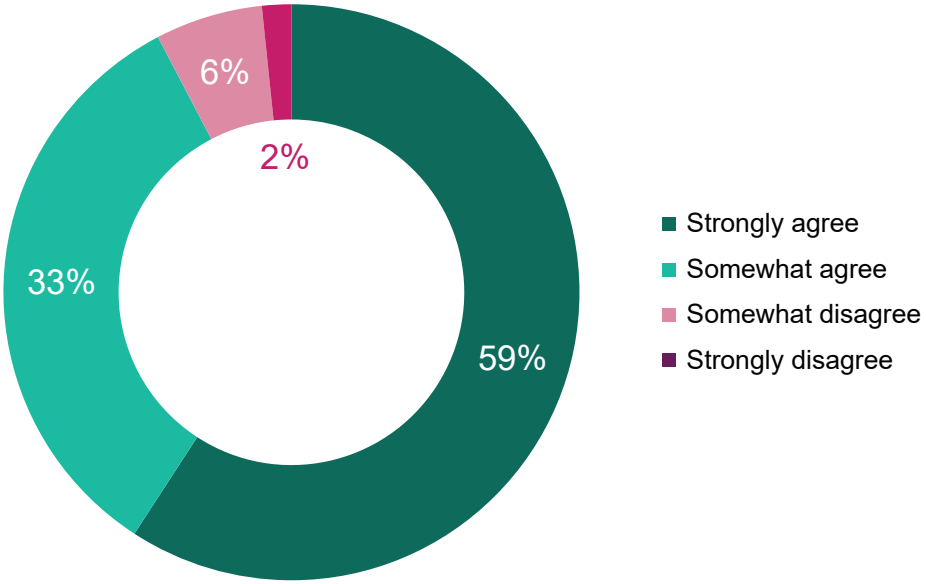
More than ever, industry and public sectors are relying on transformative technologies that acquire ever more financial, operational and customer data. Organizations use modern IT practices to house and manage this data on IT infrastructure, as well as deploy workloads and applications, and they analyze all of that data for both real-time action and support and long-term planning. All of this promises to reshape every industry and public sector, separating tomorrow's winners and losers. It changes how organizations know and interact with their customers and citizens, organize and monetize their assets and operations, and structure themselves via new and innovative customer-driven business models.

Enterprise IT infrastructure has evolved rapidly to keep up with this business change. Enterprise datacenters have become proprietary engines of innovation for many businesses – vital to operations but complex and expensive to own and manage. The public cloud brought new scale, availability and cost dynamics but also data security and privacy challenges due to their multi-tenant nature. Local, on-premises computing has evolved rapidly as well, moving from a focus on end-user desktops and devices to an endless – and industry-specific – variety of machine endpoints and IoT sensors at the enterprise operational edge that can be leveraged to deliver critical new business insights.

Choosing the Best Execution Venue for Critical Workloads

With this opportunity, choice and complexity, however, comes thorough decision-making. Where are these digital workloads best executed? What are their requirements, and where and how are those demands best met – most effectively but also most affordably? Such decisions are vitally important to digital success; according to 451 Research survey data, 92% of enterprises somewhat or strongly agree that the location in which an IoT application is deployed has a 'major impact' on their IoT project success (see Figure 1).

Figure 1: Where Digital Workloads Are Executed Correlates Strongly with Digital Success



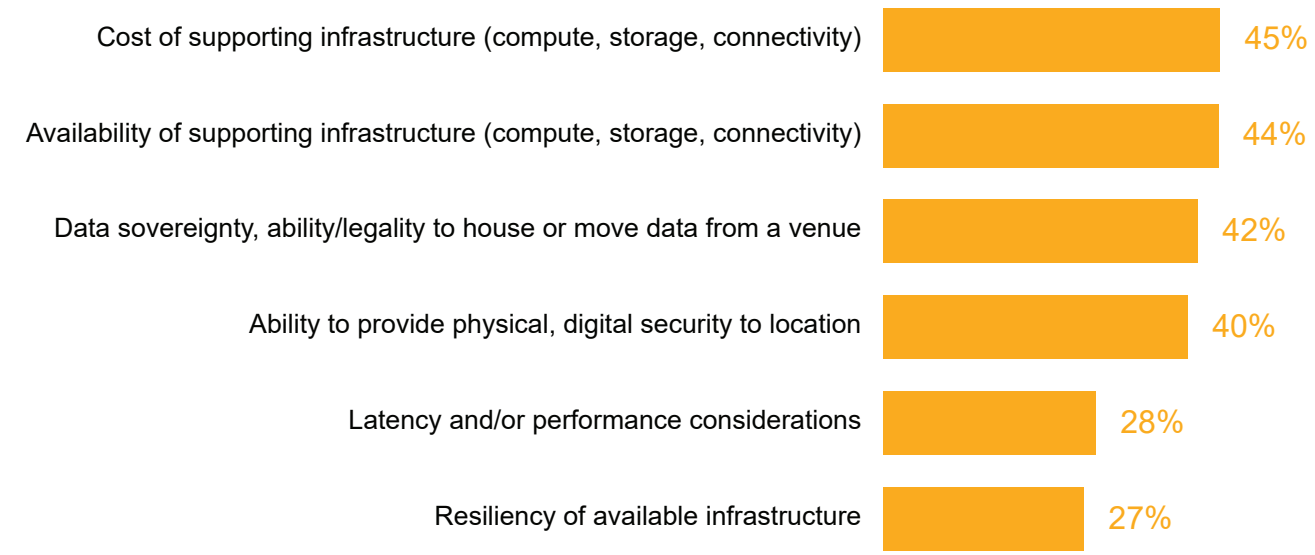
Q: For the following statement, please indicate whether you agree or disagree: My organization's IT department believes that the location in which our IoT applications are deployed – edge, near edge, cloud/core – has a major impact on our IoT project success.

Base: All respondents (n=365)

Source: 451 Research's Voice of the Enterprise: Internet of Things, Workloads and Key Projects 2021

Enterprises typically consider a range of IT venue execution requirements when thinking about how to best support their industry workloads – baseline capabilities like the cost and availability of supporting technologies as well as application and data security and sovereignty considerations. Beyond that, many digital workloads won't run optimally (and fail to deliver mission-critical business outcomes) without a high-performance, low-latency connectivity and compute environment that can support real-time transactions and insights (see Figure 2).

Figure 2: Most Important Venue Requirements for Supporting Critical Industry Workloads



Base: All respondents (n=444)

Source: 451 Research's Voice of the Enterprise: Internet of Things, Workloads and Key Projects 2021

Several of those requirements stand out as particularly important for enterprises pursuing new opportunities:

- **Balancing critical workload requirements vs. cost.** Not surprisingly, cost came out as a central concern. It always is. But in this instance, cheapest doesn't always win; the game is about optimizing digital capabilities while keeping costs in check. Enterprises have benefited from the dynamic, pay-as-you-go pricing of cloud infrastructure, especially as needs increase sharply. Cloud makes it easier to budget and plan for digital growth and success, but cloud skills are also becoming a critical and increasingly available skill. The newest – and best – IT talent grew up on the cloud, and with cloud-native, microservices-driven computing becoming more widespread, it makes sense to leverage those skills not only where big compute is needed but at the edge as well. IT as a service is becoming the norm; it makes sense from both a technology and economic perspective. It fits new, digital applications particularly well while presenting challenges – worth addressing aggressively – for enterprises dealing with legacy, non-cloud applications.
- **Handling data-rich, high-performance workloads and applications.** As more enterprise data is unleashed, IT must adopt approaches that are best suited for utilizing and managing it, as well as controlling costs. For instance, not all IoT data – especially things like bandwidth-heavy video streams – is best sent directly to the cloud. Pre-processing and conducting initial analysis at the edge not only can speed insights, but also lower storage and bandwidth costs significantly. More data isn't better in its own right; it's just…*more*. And more costly. Dealing with data at each stage of a multi-tier IT architecture – doing some processing and analytics at the edge, doing a bit more in the enterprise datacenter or in an IaaS/SaaS environment, and storing the right amount of it for long-term analysis in a more centralized location – is a critical part of planning in a data-rich, high-performance application environment.

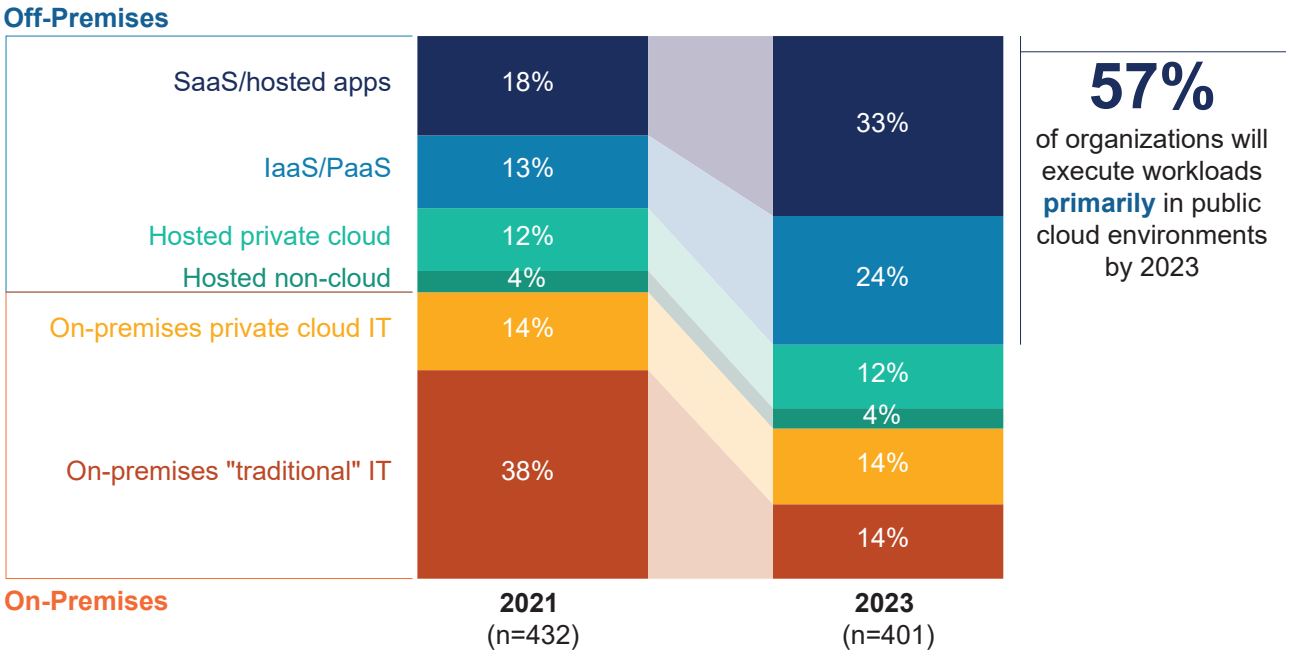
- **Enabling real-time operations.** One of the biggest enablers of critical industry applications is moving operations and decision-making into a more real-time, data-driven environment. The business value is obvious: increased agility and customer responsiveness. To enable it, enterprises must clearly understand the time-to-decision requirements of their new and emerging applications and use cases – and the infrastructure best suited to meet those demands. Workloads requiring sub-millisecond responses will need low-latency connectivity and quick-turnaround processing; others centered on the insights of complex machine learning models will require high-performance compute. Such applications and use cases would benefit from a managed edge venue capable of supporting those needs, while other applications with fewer real-time operational requirements might be best pushed out to the cloud or an enterprise datacenter for execution.
- **Supporting larger digital business goals.** Digital transformation is all about business reinvention – digital technologies and use cases can help businesses reinvent their business models, enter adjacent markets and defeat competitors in altogether new ways. In such an environment, however, risk also increases. Security attacks become more widespread and potentially damaging. Managing all that data brings new compliance and regulatory requirements. Staffing and talent management becomes both a major challenge and a potential differentiator. The decisions regarding how and where best to support mission-critical digital workloads aren't just technology decisions; they are business decisions as well, with significant implications beyond IT. For instance, 92% of enterprises said they somewhat or strongly agree that the location in which an application is deployed has a 'major impact' on their IoT project success (Figure 1).

The Evolution of Digital-Enabling IT Infrastructure

The future of business requires IT architectures that can accommodate the ever-expanding 'extended enterprise' that shifts organizational boundaries far beyond the traditional perimeter of enterprise datacenters and centralized public clouds. The edge is not simply one location but everywhere distributed business operations and customers are or are likely to be, including physical and virtual storefronts, hospitals, factories, warehouses, airports, customer touchpoints (i.e., mobile apps) and smart IoT-enabled endpoints.

IaaS/PaaS public cloud (and off-premises infrastructure more generally) continues its march into the mainstream of enterprise IT. As illustrated in Figure 3, nearly one-quarter of organizations will execute workloads primarily in IaaS/PaaS environments by 2023, and nearly three-quarters of organizations will leverage off-premises environments for the bulk of workload deployment.

Figure 3: More Organizations Are Using Off-Premises Infrastructure to Execute Workloads



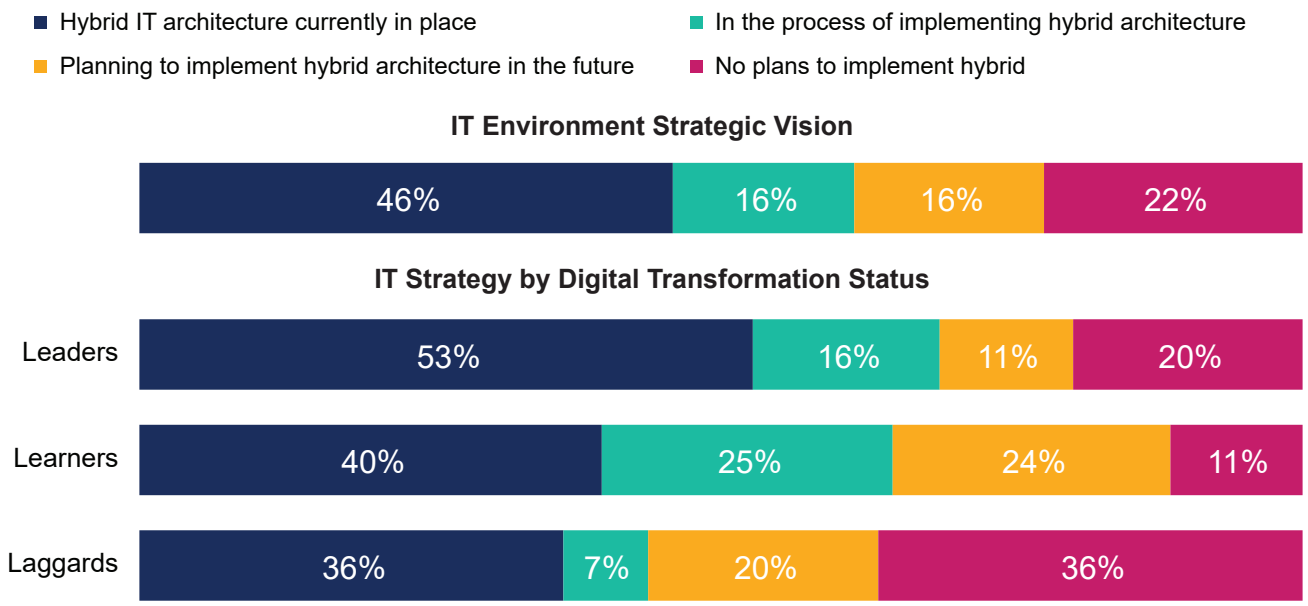
Q: Which of the following best describes the primary environment in which your organization's workloads/applications operate today? Two years from now?

Base: All respondents

Source: 451 Research's Voice of the Enterprise: Cloud Hosting & Managed Services: Workloads & Key Projects 2021

The data suggests that organizations' IT estates will be heterogeneous, with primary environments to house the bulk of workloads and other environments designated for specialized or use-case-specific deployments. Hybrid/multicloud IT architectures were once the unintended (yet in retrospect, totally predictable) result of unchecked technology sprawl, but they are now the predominant strategic posture for enabling and executing on digital-era IT and business transformation (see Figure 4). Edge computing locations are part of this mix.

Figure 4: Hybrid and Multicloud IT Architectures Are Increasingly Predominant for Enterprises



Q: A hybrid IT environment leverages both on-premises systems and off-premises cloud/hosted resources in an integrated fashion. Which of the following best describes the current state of your organization's IT environment?

Base: All respondents (n=423)

Source: 451 Research's Voice of the Enterprise: Cloud Hosting & Managed Services: Workloads & Key Projects 2021

Enterprises naturally want to keep their IT options open so they can choose the most appropriate venue for each workload based on cost, technology preferences, performance, proximity of customers, end users and partners, or other criteria. The hyperscalers' cloud-to-ground services (e.g., AWS Outposts and Azure Stack) extend the public cloud to enterprises' preferred locations, mimicking the centralized public cloud experience further out at the edge with standardized hardware and pre-installed (and remotely managed) software to ensure a consistent experience.

Generally, cloud-to-ground services are highly packaged, resulting in a fairly simple installation process with an integrated and tested stack. However, there may be some trade-offs involved. For example, there could be a limited range of hardware support, as well as limitations on where the cloud-to-ground stack can be located (resulting in a solution that might be more 'core' than 'edge'). There is also the potential for difficulties in operating cloud-to-ground stacks untethered from the centralized public cloud, resulting in roundtrip-induced latency for transaction-oriented applications and/or workflow delays for distributed data-intensive processes. The learning curve for these services may also be steep, resulting in unanticipated deployment issues and delays in realizing time to value.

The ability to deploy cloud-native capabilities at sites other than centralized hyperscaler cloud availability zones is fueling interest in edge computing. While cloud-to-ground offerings are one option, the cloud-native approach to application development and deployment makes workloads portable, enabling them to operate with more independence with regard to packaging model (bare metal, containers inside VMs, cloud instances) and best execution venue. The 'run anywhere' abstraction provided by cloud-native constructs enables workloads to be run at the right place at the right time – closer to end users, machines, sensors, cameras and other devices. Lower latency is one motivation for the edge; another is the opportunity to save on transit costs and egress fees by eliminating the need to send data back and forth from cloud to ground.

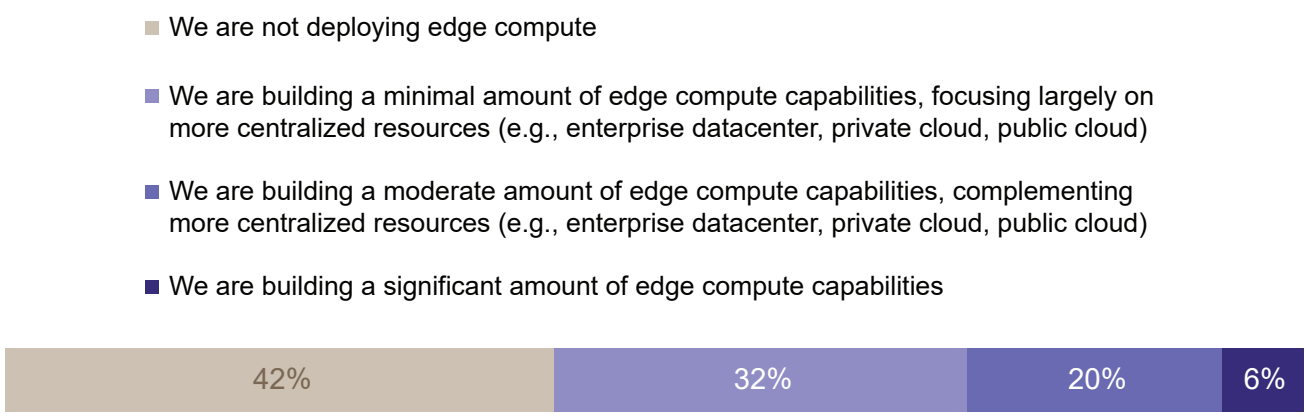
While there is no categorically wrong place to run enterprise workloads, there may be locations that are better optimized for specific workloads, use cases or business processes. When it comes to determining the best execution venue, organizations need to account for a wide range of variables, including latency tolerance, data volume, high-speed networking availability, security requirements, regulatory compliance, data sovereignty/locality considerations, and requirements around operational control. With this in mind, there are emerging options that bridge the gap between the as-a-service functionality of public cloud and the operational control and single-tenancy of private cloud.

The Role of an Edge Private Cloud

Private clouds at the edge offer a distributed extended enterprise solution that can be integrated into hybrid environments, which are rapidly becoming the norm in enterprise IT. Edge private cloud managed infrastructure offerings are tailored to serve as edge private clouds that combine the agile and scalable aspects of public cloud (along with the 'opex not capex' cost element) with the low latency, 'close to the digital action' benefit of an edge venue. They offer the operational control and security of a private cloud as well as reliable networking and dynamic, multi-access connectivity. Edge private clouds offer the best of both worlds by delivering both technology and business benefits for increasingly mission-critical industry edge workload scenarios.

As enterprises embark on the task of building the infrastructure needed to support digital business operations, edge computing factors into the equation to some extent for a majority of organizations, as illustrated in Figure 5. Nearly 60% of enterprises have plans in place to incorporate edge infrastructure capabilities into their IT estates to augment/extend the more centralized IT resources located in datacenters/colocation facilities or public clouds. As with adoption of hybrid IT environments, digital transformation leaders are also leading the charge in edge computing. Furthermore, 451 Research's Voice of the Enterprise: Digital Pulse, Budgets & Outlook 2021 survey found that nearly 20% of organizations plan to increase their spending on edge IT locations to support and extend both horizontal and vertical digital business processes to meet the ongoing need for improved customer experience and more efficient business operations.

Figure 5: Enterprise Approaches to Edge Computing



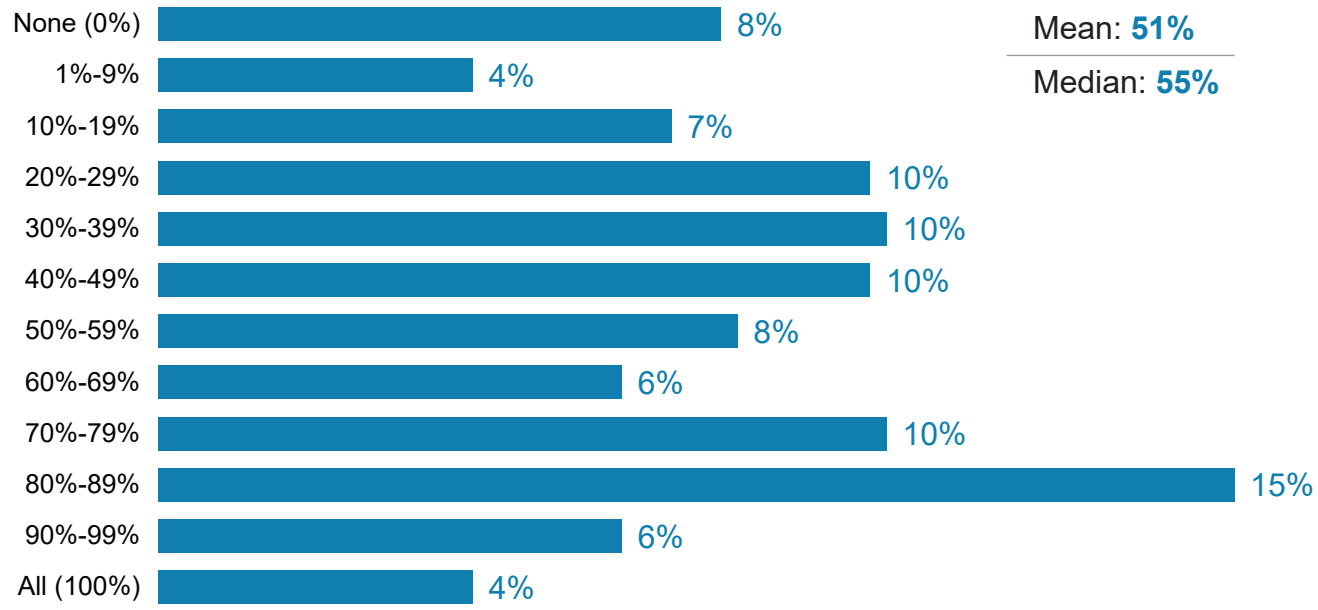
Q: Which of the following best describes your organization's approach to 'edge computing' – the deployment of storage, compute and application functionality close to where enterprise data is generated and analyzed?

Base: All respondents, abbreviated fielding (n=239)

Source: 451 Research's Voice of the Enterprise: Digital Pulse, Budgets & Outlook 2021

Organizations deploying IoT workloads are a particularly edge-centric subset of the overall edge computing universe. As Figure 6 shows, organizations with IoT workloads in place or planned in the near future reported (on average) that 51% of their IoT-enabling infrastructure is deployed beyond the perimeter of their core enterprise datacenters or public cloud providers' availability zones. Figure 7 further illustrates the configuration of organizations' IoT workloads, with 53% currently being deployed beyond 'core' edge locations in near-edge or extended edge venues.

Figure 6: Percentage of IoT-supporting infrastructure deployed at the edge

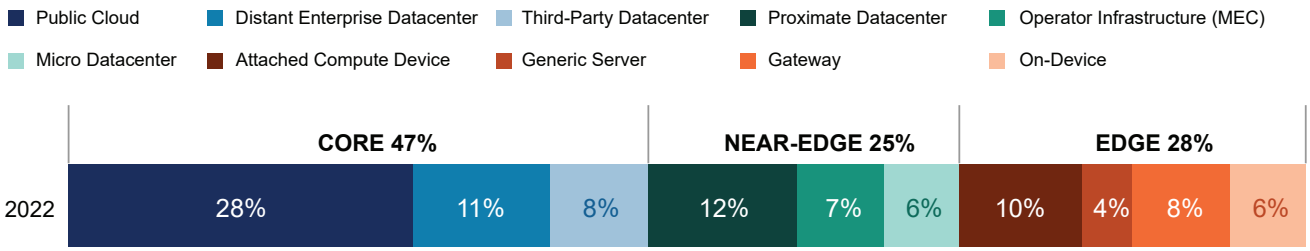


Q: Approximately how much of your organization's IoT-supporting infrastructure is currently deployed at the edge (e.g., compute, storage, analytics)? By 'at the edge,' we mean 'outside of your core datacenter or cloud provider locations.

Base: All respondents (n=433)

Source: 451 Research's Voice of the Enterprise: Internet of Things, Workloads & Key Projects 2021

Figure 7: IoT Edge Workloads



Q: Where is your IoT data initially stored and analyzed?

Base: IoT respondents

Source: 451 Research's Voice of the Enterprise: IoT, Workloads & Key Projects 2021

Given that the exact location of the enterprise edge can vary just as much as the best execution venue for the IT environment as a whole, organizations need to ensure that their edge deployments can integrate and interoperate with other critical elements of the IT estate – covering the waterfront from centralized core to core edge, near edge and extended edge.

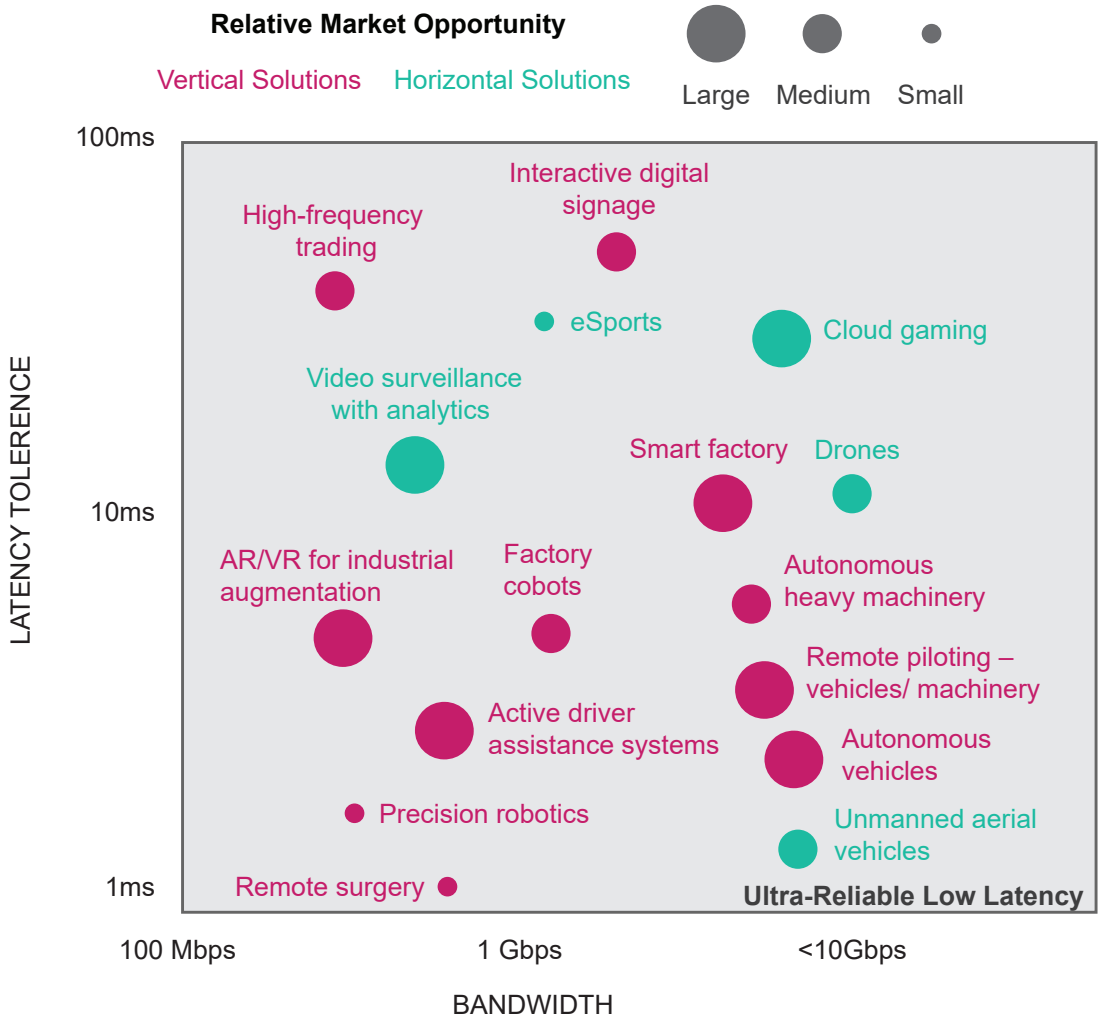
An edge private cloud offering that is managed infrastructure as a service can yield important business and technology benefits for organizations looking to build or enhance the IT foundation of digital business architectures:

- **A fit-for-purpose IT environment that delivers the resource scalability, resiliency and security that modern digital enterprises need.** Productized as-a-service offerings can eliminate the performance and governance risks of informal 'skunkworks' edge compute deployments enabled via stand-alone disconnected boxes in random locations. Mission-critical applications operating at the edge deserve tailored offerings managed by experts in IT (not moonlighting OT or line-of-business personnel). This type of standardized product meets growing demands in a more efficient manner, providing data and transactional protection through dedicated compute resources and on-net connectivity.
- **Resources ready to support computing when and where required**. The ability to spin up compute and storage resources on demand is now a mainstream function of modern IT. Organizations receive this capability via hyperscale public clouds, and now it has come down to earth at the enterprise edge. Business processes that take place at the industry and market edge can now tap into consumption-based, pay-as-you go infrastructure – with the added benefit that the enterprise enjoys broader control of operation and workload placement, thus mitigating the impact of vendor lock-in.
- **Delay is not an option – placing mission-critical workloads at the edge accelerates digital business operations.** In the digital era, enterprises must keep up with (and stay ahead of) customer and employee expectations. As key constituencies become ever more intolerant of digital delivery delays, the proximity of the necessary workload components is an increasingly urgent prerequisite for delivering real-time (or at worst, near-real-time), low-latency performance at the edge. Edge private cloud offerings featuring pre-packaged and integrated infrastructure speed up time to value and overall return on digital business investment.
- **Nearby compute with integrated connectivity for performance and data sovereignty.** An edge private cloud is situated 'on-net' for fast access and supporting large-scale data backhaul with adjacent compute resources capable of sub-millisecond response times. That delivers the performance required for operational workloads and keeps data local – rather than in a remote cloud – when security, compliance or other enterprise requirements demand it.

Supporting Digital Industry Use Cases with Edge Private Cloud

A variety of industries and use cases can benefit from the attributes of an edge private cloud managed infrastructure as a service. Of particular focus are applications – both horizontal and industry-specific – that require high-performance, low-latency and highly secure edge compute (see Figure 8).

Figure 8: Industry Use Cases with Unique Requirements Fulfilled by Edge Private Cloud



Source: 451 Research

Several sectors in particular – e.g., manufacturing, government, retail and financial services – have use cases particularly suited to such infrastructure. Understanding their unique requirements and how a private cloud edge can serve them helps to highlight the value of an edge-private-cloud approach.

Manufacturing Edge Use Cases

Industry 4.0 concepts like automation, real-time data and networked systems are prime drivers of the enterprise edge within manufacturing. It is best to think of manufacturing use cases as part of a continuum as digital transformation has been underway for decades. Top use cases such as production manufacturing, quality assurance and inventory management/logistics benefit from more – and more *real-time* – endpoint data and advanced analytics that yield actionable insights. Edge compute and IoT endpoints also enable entirely new mission-critical industry use cases – such as fully autonomous robots and connected/augmented workers – that weren't fully possible without the high-performance, low-latency capabilities of a connected edge.

Why Edge Private Cloud: Mature industrial IoT manufacturing environments already include heavy doses of technology, increasingly running in IT standard environments. After sending many IoT 1.0 applications to the cloud, manufacturers have become among the strongest proponents of edge compute because of the need for high performance, the challenge of scaling from pilots to full production and the desire to minimize costs associated with such full-production deployments. A heavy emphasis on artificial intelligence/machine learning for deep data insights – particularly real-time operational insights – also calls for strong edge compute capabilities at tier one manufacturers while the long tail of tier two/three manufacturers benefit from the more managed nature of edge private cloud versus trying to deploy and run their own edge infrastructure.

Government/Public Sector Edge Use Cases

The government sector is ripe for digital transformation, enabled by data-driven 'smart city' concepts and use cases. Near-term digital use-case wins are often executed at a departmental level, led by relatively simple sensor-driven applications such as smart parking and resource monitoring. Larger, more encompassing use cases – like smart traffic or street lighting – can pay significant dividends, but they require a much larger investment to be rolled out at significant scale.

Why Edge Private Cloud: Edge compute can help cities execute digital use cases without having to backhaul and store large amounts of data to centralized locations, while a private cloud maintains the privacy and sovereignty of sensitive citizen data. Video analytics will play a key role across smart city applications, with edge compute providing on-device or near-device processing of video metadata and real-time analytics to generate insights, from the simple (e.g., open parking spaces or the number of people on a city bus) to the endlessly complex (e.g., coordination of city roads and transportation systems with increasingly autonomous vehicles). As 5G, SD-WAN and other types of advanced connectivity roll out across cities, municipal governments will not only be partners to those deployments but frequent consumers of those high-bandwidth, near-edge compute resources.

Retail Edge Use Cases

Retail is characterized by a mix of critical edge endpoints – payment registers, merchandised store environments and inventory systems that manage data flows from the supply chain to inventory warehouses and onto the store floor. To that end, we see a broad mix of use cases that take advantage of the nearby compute of the edge and in-store managed private cloud instances – from shelf-level shopper analytics to cold chain food safety. The emphasis and timing of the deployment will depend on the type, size and digital ambition of the retailer; the largest retailers that serve constantly evolving brick-and-mortar environments are among the most IT-savvy and edge-driven of enterprises. In short, data-driven and edge-enabled digital innovation will be a key disruptor in the sector over the coming years as retailers enhance the customer experience and find new ways of competing with e-commerce-only rivals.

Why Edge Private Cloud: The retail edge – i.e., store locations – must typically support a bevy of legacy applications, such as VoIP, point of sale and HR/scheduling. On top of that, a range of new data-endpoint-driven use cases require in-store sensors and cameras to track products, shoppers and workers. Use cases requiring low-latency execution – for instance, real-time merchandising – require real-time edge compute, while all of that data must (or can) also be shuttled off to more centralized locations to support use cases such as store planning and post-sales analysis.

Financial Services Edge Use Cases

The financial services sector is one of the most data-intensive and IT-forward of all industry verticals. Digital use cases vary, but almost all are built on speed of execution – e.g., payment processing, back-end reconciliation and low-latency trading. The sector's focus is on executing those financial transactions more quickly and securely while leveraging new endpoint data and insights to improve critical processes like fraud detection and approval/underwriting. On top of that, many financial services firms have an array of brick-and-mortar and digital customer endpoints, such as ATM kiosks, that could be modernized to better leverage advanced connectivity, private edge compute and real-time data-enabled analytics.

Why Edge Private Cloud: Financial services firms could benefit from a higher-performing yet still private and secure enterprise edge in several ways: further digitizing branch offices and machines like ATMs and registers; better leveraging edge resources to reduce the cost and management complexity at centralized datacenters; and fully embracing a modern edge-to-cloud approach including cloud-native/microservices applications to enhance agility. Infrastructure that delivers not only edge performance but also the security of a private cloud instance is critical for a sector built on such highly sensitive and highly regulated customer transactions.

Conclusions

A growing landscape of industry-specific workloads and use cases has fundamentally shifted the conversation regarding IT deployment and best execution venues. Analytics and business processes rely on data from multiple sources – traditional business applications typically hosted at the core ***and*** real-time operational data generated from mission-critical endpoints at the edge. While there are multiple ways to enable the edge component of the edge-to-core compute/storage/analytics continuum, enterprises generally want flexibility and familiarity. Edge private clouds provide dedicated, secure and network-connected outposts of IT resource capacity where needed to power and connect the increasingly extended ecosystems of people, machines, devices, locations and platforms that drive the data processing, transactions and automated business processes characteristic of digital enterprises.

CONTACTS

The Americas
+1 877 863 1306
market.intelligence@spglobal.com

Europe, Middle East & Africa
+44 20 7176 1234
market.intelligence@spglobal.com

Asia-Pacific
+852 2533 3565
market.intelligence@spglobal.com

www.spglobal.com/marketintelligence