Five IT Must-Haves for Modern Edge Computing

Pathfinder Report

November 2022

Commissioned by

CISCO

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 Author

David Immerman

Consulting Analyst

David Immerman is a Consulting Analyst for the TMT Consulting team based in Boston, MA. Prior to S&P Market Intelligence, David ran competitive intelligence for Interos, a supply chain risk management software startup. He spent nearly four years on PTC's Corporate Strategy Marketing team, providing thought leadership and market research on technologies and trends across manufacturing. Previously, David was an industry analyst in 451 Research's Internet of Things channel for three years, primarily covering the smart transportation and automotive technology markets. He holds a bachelor's degree in Business Administration from Marist College.

Introduction

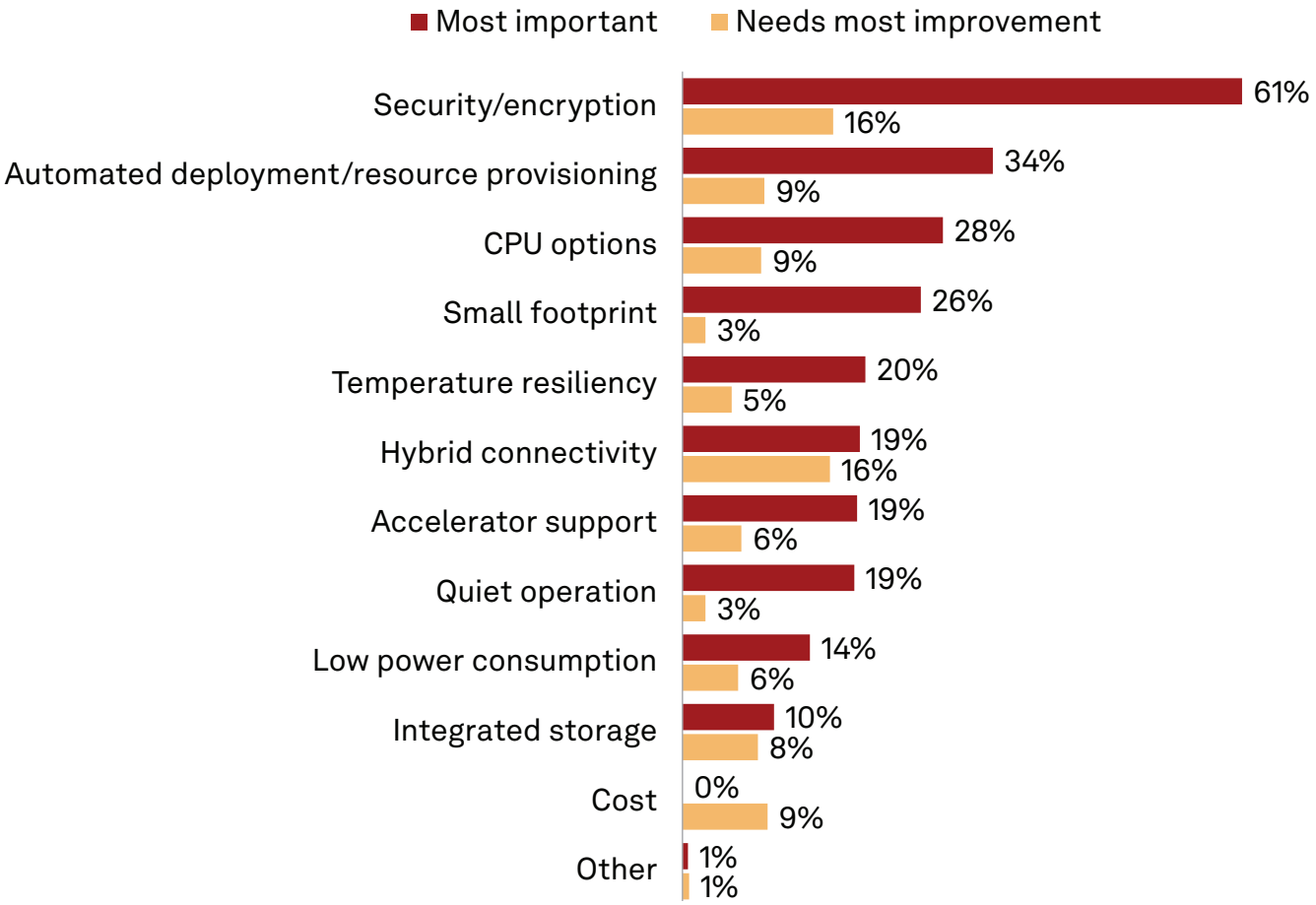
Edge computing has emerged as a key technological consideration enabling organizations to harness increasing numbers of data sources and analyze growing data volumes. According to our research, more than half (53%) of IoT workloads' data computation will be deployed at the edge and in near-edge environments by 2024.

The shift toward hybrid and decentralized infrastructures encouraged through edge computing has driven key strategic initiatives, created operational efficiencies and generated valuable business outcomes. Edge infrastructure has a prominent role in digital transformation strategies; nearly half (49%) of digital leaders have already deployed edge infrastructure. These deployments create infrastructure and operational opportunities for optimization that span IT and physical-site-based operational technology (OT) employees, ranging from manufacturing process engineers to facilities managers, and the business.

While there is collaboration with OT, IT predominantly manages edge deployments and takes ownership of unlocking the opportunities the edge presents. IT is faced with novel technical challenges from the edge-driven decentralization of IT resources, which impact user experience, costs, operations, application performance, data security and more.

IT teams are wisely turning to modern technology solution providers to mitigate these challenges, yet there is a perceived gap for enterprises between the edge capabilities desired and viable solutions from vendors. Key findings from 451 Research's Voice of the Infrastructure: Edge Infrastructure & Services 2022 points to the most important features of edge servers, yet many require improvements.

Figure 1: Edge servers: most important features versus features that need improvement



Q: Thinking of your organization's edge computing requirements, what are the most important edge server features or capabilities? Please select up to three. Base: All respondents (n=331).

Q: What feature or capability needs the most improvement in your organization's edge servers? Base: All respondents (n=280).

Source: 451 Research's Voice of the Enterprise: Edge Infrastructure & Services, Use Cases 2022.

Edge servers need to support horizontal scaling from applications that require a few nodes to clusters that require dozens of nodes, while operating with a minimal footprint and with the resiliency to perform in a range of environments. To support the many different edge computing use cases, the growing list of required capabilities includes:

- **Cloud-like scale:** Use flexible infrastructure and storage to scale deployments and support fluctuations in demand and pre-processing to keep mission-critical data local for business, security, cost and compliance requirements.
- **Cloud operations:** Leverage automated provisioning and centralized management of global policies, seamless updates, common templates, configurations, patching and other on-demand resource rollouts across the edge deployment. Implement intelligent support to leverage and scale the collective intelligence of IT teams across all core and edge sites.
- **Flexible, high-performing and efficient infrastructure:** Achieving performance efficiencies that maximize the server's processing footprint in remote edge sites requires support of and integrations with different and increasingly powerful CPU/GPU chipsets and accelerators.
- **Data integrity and consistency:** Maintain consistency of data while ensuring high availability and fault tolerance across edge systems.
- **Multi-layered and resilient security:** Maintain system resiliency and security across edge nodes, devices, servers, data, networks, sites and personnel.

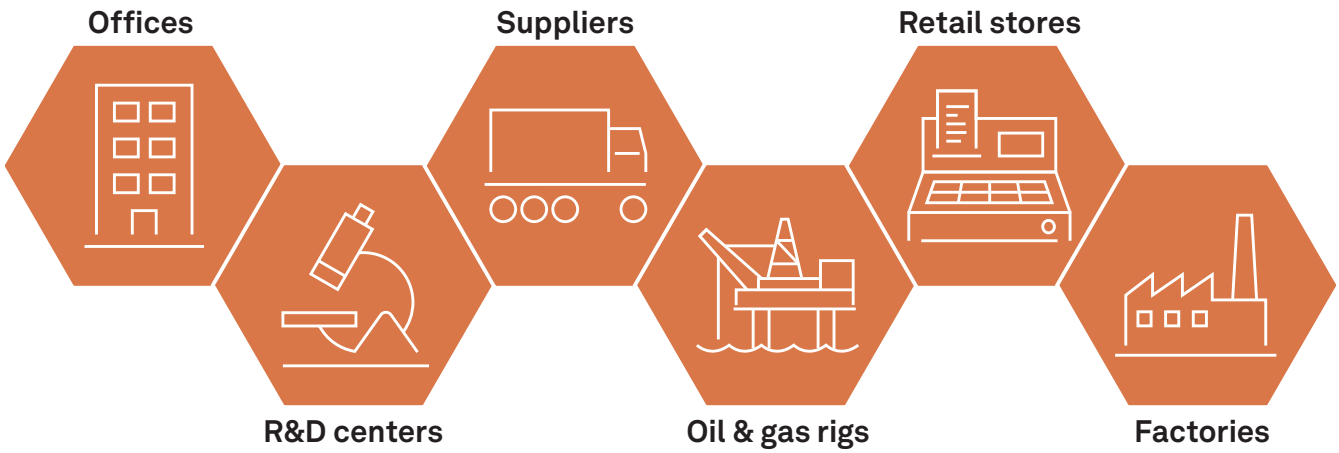
Providers with differentiating features in these categories that can check multiple or all boxes will rise to the top in edge purchasing decisions. 451 Research has outlined these five IT "must-haves" as key considerations for evaluating solution providers, creating a modern edge computing environment and obtaining the full value of efficiently scaling the edge across the enterprise.

Attain Agility at the Edge With Cloud-Like Scale

Scaling IT to Meet Expectations of the Ever-Expanding Edge

Edge infrastructure and sites are unquestionably on the rise; within two years, nearly a quarter of organizations will have more than 100 sites with dedicated edge infrastructure, and each site is estimated to have on average six server nodes, according to 451 Research's Voice of the Enterprise (VotE): Edge Infrastructure & Services, Use Cases 2022.

Figure 2: Edge site types



Source: 451 Research.

These edge sites include facilities ranging from remote offices and oil and gas rigs to brick-and-mortar retail stores and factories. Edge infrastructure deployments typically include brownfield integrations that bootstrap edge capabilities into existing sites. This process has largely been inefficient and capital-intensive where purchases are based on best guesses of required capacity. According to 451 Research's VotE: Hyperconverged Infrastructure, HCI at the Edge 2022, 40% of organizations increased investments in existing edge deployments to address technical challenges. The unplanned costs of capital expenditures and over-provisioning of physical infrastructure can inhibit the value that edge unlocks. These unbudgeted investments stress IT resources and miss the opportunity to leverage emerging consumption models that centralized infrastructure has benefited from, such as cloud computing "scale-on-demand" purchasing.

As edge deployments continue to grow, IT departments must consider flexible infrastructure that efficiently scales to support the myriad use cases, applications, data formats and bandwidth demands. Consider "follow-the-sun" applications where edge infrastructure manages a significant influx in traffic during pre-determined peak hours in a specific region or a time- or event-related use case. An edge site must operate with cloud-like scale to manage demand-driven compute instances in an efficient and dynamic manner with the edge infrastructure allocating virtual resources to support it.

Build a Modern Foundation With Flexibility for Growth

As with any digital transformation, edge deployments often start small with one or two high-impact use cases supported by a few nodes. When benefits are realized, the project — and the resources it demands — can rapidly expand. Organizations need to invest in technologies that provide a flexible foundation to deliver cloud-like scale with features supporting new and existing edge deployments, use cases and infrastructure, and an operational model that can connect these sites.

To further alleviate the strain on IT from managing widely dispersed edge clusters, teams must continuously look for automation opportunities. This includes timely automated provisioning and optimization of new workloads based on geographic location, power and facility costs, equipment configurations, regional regulations, network latency and other critical operational parameters. Mandating global policies and common templates across edge sites enables continuity across IT teams, further expediting provisioning.

Some local storage is required and must be flexible to support fluctuations in demand as required for follow-the-sun applications. Initial pre-processing capabilities can greatly lessen the need to constantly send, store and retrieve large volumes of data in cloud environments.

IT teams will aim to channel these cloud-like capabilities to form a flexible modern foundation and recognize cost optimizations.

Unlock New Levels of Operational Productivity With Cloud Operations

IT Is Constrained With New Service Processes, Workforce Shortages and Costs

While edge computing has largely decentralized infrastructure, centralizing IT operations to run remote management and provisioning is required to maximize the edge infrastructure at scale and respond to possible issues with unprecedented speed.

IT systems frequently require software updates for managing deployments and configurations across their infrastructure, as well as physical intervention for replacing hardware components, assessing disks and fans, monitoring temperatures, on-site troubleshooting and other server maintenance activities. Given that these systems are traditionally located in centralized datacenters with dedicated IT staff, digital and physical support processes are more defined, and costs are better known, manageable and planned. Executing on these processes through continuous optimization of the IT infrastructure generates costs savings and achievement of other business goals.

However, given the decentralized nature of geographically dispersed and growing edge deployments, servicing these IT assets is less defined, and costs can be unknown, higher and unexpected. Dispatching IT staff to hundreds or thousands of edge sites can quickly compound operating costs including travel, vehicles, fuel and staff. Common service metrics are quickly extending into IT departments, such as time to resolution, truck rolls and downtime.

For companies with constrained resources, this physical service intervention is often limited to business-impacting outages. According to 451 Research's VotE: Edge Infrastructure & Services, Use Cases 2022, 41% of organizations' IT teams can only physically access servers in edge sites for maintenance or upgrades monthly. Multiply costly restricted maintenance activities with a competitive IT labor market where challenges with recruiting, onboarding and training new staff continue to rise, and the opportunities and benefits that edge is supposed to deliver can quickly diminish.

Adopt Cloud Operations to Scale IT Expertise

Rectifying these costly factors requires appropriately leveraging current resources and investing in technologies that enable remote deployment and resolution of operational issues. IT teams need to be able to efficiently manage a one-to-many ratio of IT staff to edge deployments (e.g., one IT administrator managing 50 edge sites). The ratio will vary, but the goals of continuously improving edge operational efficiency and identifying areas for ongoing optimization will be universally critical. To accomplish this, IT must democratize, standardize and scale the collective intelligence and domain expertise of the entire department across the edge deployment base and invest in technologies and tools that empower them to do so.

Remote resolution is likely the biggest cost saver in this scenario; truck rolls (dispatching an employee or contractor) are a major expense. Software that can remotely monitor, manage and autonomously recover edge systems can decrease and even eliminate the need for a costly IT truck roll. Equipping IT to issue remote software updates and patches across edge sites in a few seconds from a centralized operations center provides a cost-effective solution to mitigate issues.

Leveraging the cloud can enable broad geographical coverage and remote monitoring of edge systems for hardware and software performance metrics, CPU/RAM utilization, network usage, security and other key performance indicators. Providing this single pane of glass and operational plane across the deployed edge systems gives IT a comprehensive source of truth to simplify and streamline deployments, manage life cycles across hardware in hybrid environments (datacenter, edge sites, etc.), understand bottlenecks, take rectifying actions, and provide ongoing optimization and support. IT must leverage the speed and scale of the cloud to fuel operations and optimize productivity for edge computing to succeed.

Gain Efficiencies With Flexible and High-Performing Edge Infrastructure

A Shifting Compute Paradigm, but Similar Performance Priorities

With Moore's law in effect, processors continue to deliver higher levels of performance with a smaller physical footprint. However, most of these silicon innovations are designed to operate in more centralized infrastructure and cloud datacenters versus edge sites, which pose unique challenges. Without the benefits of centralized infrastructure and access to IT staff, operational challenges arise due to geographically dispersed, hazardous or atypical climates of edge locations. Enterprises must maximize performance efficiencies across these disparate deployments with considerations for price/performance, software licensing costs and performance per watt/rack unit. To do this, organizations are investing in technologies that simplify the management of edge systems and optimize the existing power and IT infrastructures. Additional "bolt-on" capabilities based on edge sites' unique needs (e.g., SD-WAN to ensure network performance and AI/ML for pre-processing to lower cloud storage costs) provide another layer of support and cost optimization.

Enterprises with edge deployments are adjusting to this shift in computing resources; 61% of organizations are planning to increase the amount of hyperconverged infrastructure (HCI) in their edge environments over the next two years. However, the performance — or at least the perception of performance — at the edge is currently lackluster; compute performance was the most frequently cited HCI capability in need of improvement in organizations' edge deployments. There is a clear capabilities gap forming for edge solution providers to fill as enterprises lean on edge computing for performance-oriented features in mission-critical and latency-sensitive applications requiring significant real-time computation, data ingestion and analysis.

With environmental, social and governance (ESG) initiatives skyrocketing up the list of C-suite and board member priorities, sustainability has become an enterprise initiative, and IT and the edge are not exempt. Respondents cited energy and power efficiency/transition goals (44%) and supporting ESG programs (43%) as key digital transformation drivers in a custom 451 Research survey of industrial companies. A growing edge footprint adds ESG, power consumption and scope 1 emission calculations as questions that IT must be prepared to answer.

Achieve Efficiencies at the Edge

Whether by investment in new IT systems or leveraging existing compute resources, enterprises need to drive performance efficiencies in the expanding edge infrastructure. This can include optimizing the current number of cores or reducing the need for new ones with more powerful CPUs delivering the same performance at lower cost from fewer licenses and better energy efficiency. Multiple CPU cores and sockets may not provide a better performance-to-cost ratio over single-socket systems with more powerful CPUs. The flexibility to rapidly scale up for modern compute and core-intensive workloads such as video surveillance and scale down for intermittent demand is an attractive value proposition because companies across industries typically have a mix of edge applications with varying performance requirements.

Maximizing the physical and data footprint of these edge deployments with high-capacity storage and performance density is key to reduce reliance on network connectivity to centralized datacenters and cloud environments, where data ingress and egress costs can run rampant. Native integrations between vendors, silicon, network, software and hardware common at the edge can simplify configurations and push edge solutions to higher performance by optimizing capacity utilization, and by provisioning of edge systems' workloads with resource usage.

Achieving and measuring these collective performance efficiencies can be tied back to ESG goals and KPIs such as improving performance per watt of electricity consumed. Multiplying even marginal improvements in energy usage across thousands of edge devices could save millions of dollars in operating costs while improving ESG metrics for carbon footprint or greenhouse emissions.

Maintain Data Integrity and Consistency at the Edge

The Edge Generates a Lucrative but Challenging Data Source

Edge computing provides a more efficient way to process and analyze data, which is the foundation for the innovative use cases and business outcomes it enables. Edge sites contain critical data that organizations are keen to leverage. However, harnessing and obtaining value from edge-generated data is hardly straightforward; data and system inconsistencies are common and must rely on a limited edge infrastructure.

In typical edge scenarios, IoT-enabled devices may be generating petabytes of data and requiring sub-millisecond response times, which creates a noisy and unmanageable data environment. Edge site infrastructure may not have the IT resources to resolve these data dilemmas and could be in areas with poor network connectivity, which jeopardizes data backup processes and remote administration capabilities.

In edge environments, the need for resilient systems that maintain data integrity and provide a means for data recovery is paramount. Solutions must be able to maintain consistency across these data sources and file systems while maintaining high availability and fault tolerance. Downtime for a mission-critical edge application can cost an organization thousands of dollars in operational costs or damage brand reputation in customer-facing edge use cases.

Gain the Full Value of Data by Ensuring its Integrity

Much of the innovation happening at the edge leverages capabilities initially born in the cloud. For example, “cloud witnesses” provide a data integrity check by using “voting” mechanisms between them to check on data inconsistencies and prevent downtime.

Applying this logic to two edge node systems, a third-party or witness can act as the final vote when there is a disagreement or “split brain” condition between the two nodes due to one being unavailable or having out-of-sync data.

Operating this witness locally (instead of through WAN) by using edge routers, switches, access points or other devices can reduce network connectivity and availability concerns, while potentially reducing costs from traditional methods that operate witnesses on centralized virtual machines. Whether locally or in the cloud, leveraging these lightweight witnesses can better maximize limited edge resources and costs as edge deployments scale.

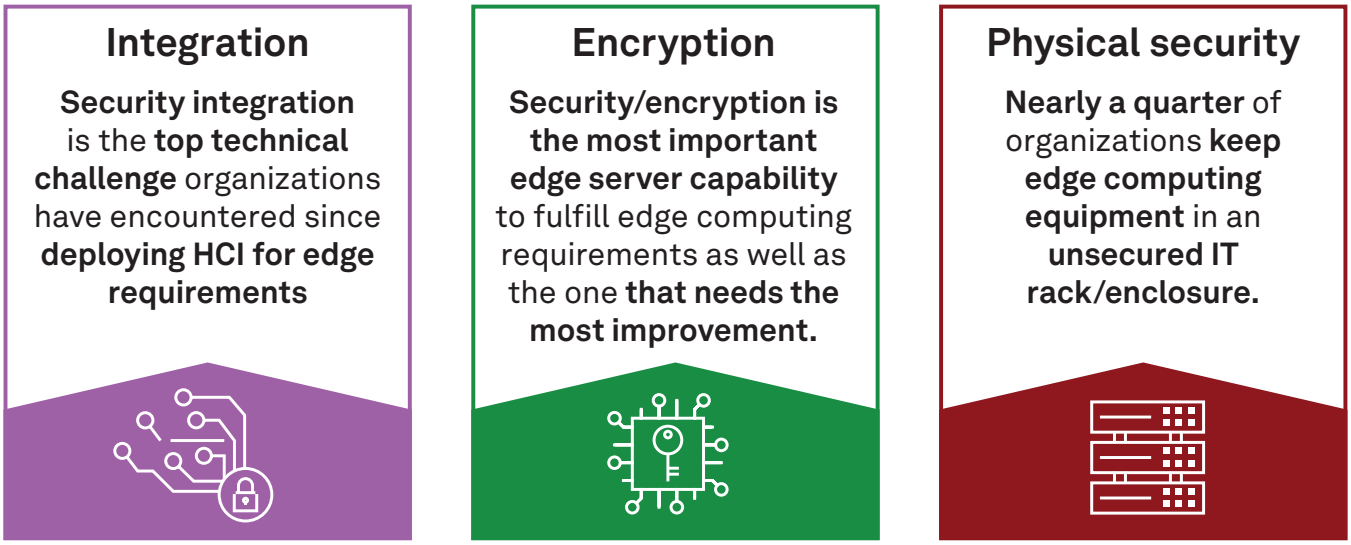
While the cloud-witness approach helps diminish data recoverability concerns, IT teams will need to address additional data management, protection, backup, security and recovery hurdles with modern techniques and technologies across on-premises, hybrid and cloud environments.

Create a Resilient Edge With Physical and Digital Security

Vulnerabilities Persist With Expanding Edge Infrastructure

As the infrastructure grows and decentralizes, IT must maintain a resilient security posture to address the unique physical and digital challenges. IT professionals cited security/encryption as the most important edge server capability required to fulfill their organizations' edge computing requirements, as well as the capability most in need of improvement.

Figure 3: Top technical challenge with HCI for edge; most important edge server feature; most common edge computing deployment location



Q. What has been the most difficult technical challenge your organization has encountered since deploying HCI for its edge requirements?

Base: Organization has deployed HCI in edge/remote/branch office (n=52).

Source: 451 Research's Voice of the Enterprise: Hyperconverged Infrastructure, HCI at the Edge 2022.

Q. Thinking of your organization's edge computing requirements, what are the most important edge server features or capabilities? Please select up to three.

Base: All respondents (n=329).

Source: 451 Research's Voice of the Enterprise: Edge Infrastructure & Services, Use Cases 2022.

Q. In which of the following locations has your organization deployed (or does it plan to deploy) edge computing infrastructure? Please select all that apply.

Base: All respondents (n=329).

Source: 451 Research's Voice of the Enterprise: Edge Infrastructure & Services, Use Cases 2022.

As the number of edge locations grows, it becomes increasingly difficult to maintain governance, policies and physical security with limited resources (people, budget, equipment, etc.); nearly a quarter (23%) of organizations with edge computing equipment have some of it located in an unsecured IT rack or enclosure. Many have suffered the consequences of the "it could never happen to me" mindset, having to endure the massive reputational and financial blowback from a data breach or system hack. The long list of documented thefts range from a disgruntled hospital worker downloading private data via USB drives to the well-known Colonial pipeline ransomware incident where attackers stole 100 gigabytes of data, generating ripple effects that shut down gas supplies across the U.S. and disrupted global supply chains.

"Man-in-the-middle" attacks occur when a cybercriminal intercepts systems that disconnect and reconnect to the network or gains access via insecure nodes with exposed network cabling. Given the inconsistent network connectivity and thousands of edge devices, this is an avenue for cybercriminals to compromise a connection, infiltrate a network and create a serious security threat for IT.

Hardening and embedding security into edge devices is necessary, yet it is hardly guaranteed, particularly across older edge systems. Some manufacturers of connected devices make cost tradeoffs between security, connectivity, application features and digital experiences, where security is not prioritized. The impact of not securing these connected devices can be sizable, laying the foundation for massive distributed denial-of-service attacks where cybercriminals have used unsecured devices as part of a botnet that overloads a web server with requests.

Build Modern, Multi-Layered, Resilient Edge Security

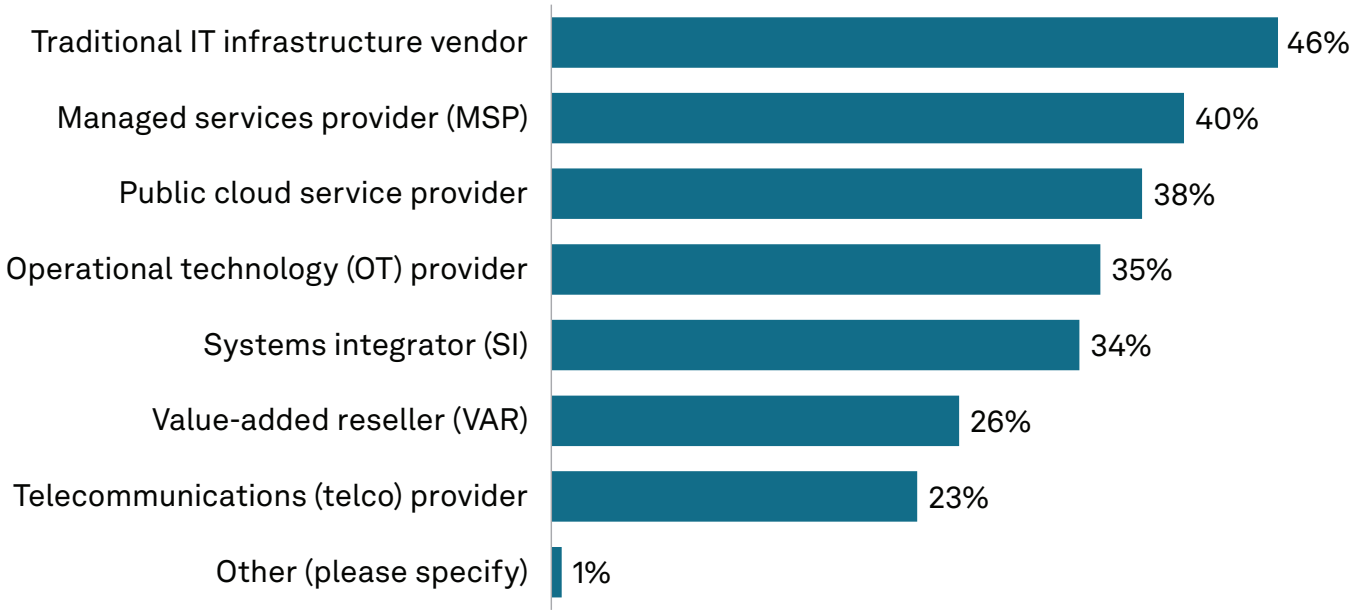
Organizations must provide IT with the resources, technologies and tools to create, maintain and manage the multi-layered security architecture required to enforce a resilient edge infrastructure. IT and security teams cannot create or enforce piecemeal processes and point solutions to cover the growing gaps between constantly expanding attack surfaces, unpatched vulnerabilities and multi-layered security requirements. To thwart the various attack vectors that cybercriminals may seek to exploit across the expanding edge, organizations need a multi-pronged approach to security with robust encryption service as the first line of defense.

Encrypting mission-critical data both at rest and in transit reduces the likelihood that cybercriminals can leverage the data, such as holding it for ransom or selling it. As mentioned, remote servers aren't always physically secured, and thefts can occur. In combination with investing in stronger physical security, enterprises should consider encrypting data both in disk and in memory. "Always on" encryption on these devices should be an absolute requirement. Using cloud-based management systems helps create consistency across environments. A single and unified management portal permits centralized IT administration, governance and provisioning of security policy updates to edge sites.

Form an Edge Ecosystem

As with any digital transformation, an ecosystem approach that leverages internal teams, their skill sets and tools combined with technology solutions from external vendors and partners provides the fastest time to value. According to 451 Research's VotE: Edge Infrastructure & Services, Strategy & Implementation 2022, digital laggards are more likely to prefer a "DIY" approach to developing and deploying their edge infrastructure. Digital leaders are more likely to form partnerships with IT infrastructure vendors, managed service providers and several others to form the technological backbone of modern edge deployments.

Figure 4: Top sources of edge computing infrastructure

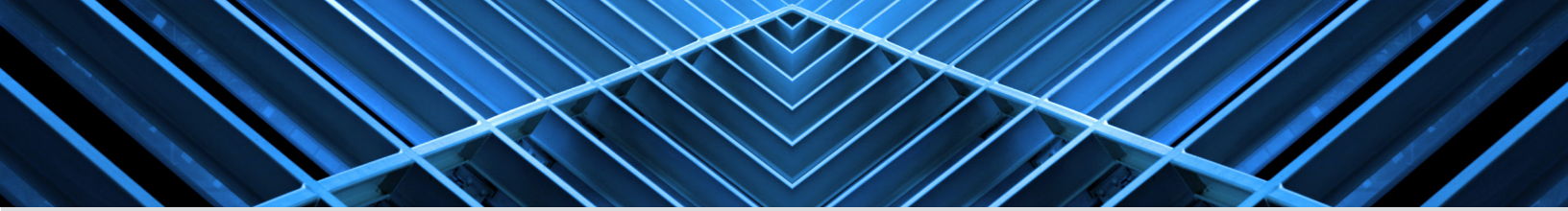


Q. Which of the following sources provides/will provide edge computing infrastructure (e.g., servers, storage devices, networking devices) for your organization's edge sites? Please select all that apply.

Base: Respondents whose organization has edge computing infrastructure in use, in discovery/proof of concept or plan to implement in the next two years (n=300).

Source: 451 Research's Voice of the Enterprise: Edge Infrastructure & Services, Strategy & Implementation 2022.

Enterprises will need to consider these external vendors to drive successful edge projects. The impact of these partnerships amplifies as the edge projects scale and increase the number of endpoints, nodes, use cases and sites. Leveraging this ecosystem approach to unlock the five IT must-haves for modern edge computing will establish a flexible foundation that can achieve the desired scalability. Enterprises and IT teams coordinating these internal and external ecosystems will recognize the transformational value of edge at scale and drive competitive advantages in their respective market spaces.



content provided by: CISCO AMD

Modernize Edge Infrastructure – Cisco HyperFlex with AMD EPYC Processors

Cisco and AMD have collaborated to deliver solutions for the edge that meet all the must-haves noted above. Cisco's HyperFlex Edge systems powered by AMD EPYC processors provide solutions that scale from two nodes up to 64-node clusters. Organizations can manage and optimize their edge locations using Cisco Intersight, leveraging the power of SaaS to improve performance and efficiency while also scaling their IT teams expertise across an ever-expanding landscape. Intersight provides one place to manage on-prem, edge and hybrid cloud environments, facilitating global deployment of servers and apps, policy configuration, monitoring of system health, and issue identification and resolution.

HyperFlex Edge systems deliver unique, simple, cost effective and flexible data integrity and resiliency capabilities that eliminate the need for additional edge infrastructure or connectivity to ensure resiliency and stability. Cisco also partners with leading data backup vendors to quickly protect and recover data across edge environments. HyperFlex offers a range of encryption options to secure sensitive information at edge locations, such as self-encrypting drives and software-based encryption. AMD EPYC processor's secure virtualized environments take it a step further and help protect against a malicious hypervisor.

HyperFlex combines a purpose-built data platform, Cisco-UCS servers, and lifecycle management from the Cisco Intersight™ cloud operations platform. Combined with AMD EPYC processors, HyperFlex is engineered to meet the needs of today's edge infrastructure, enabling organizations to deliver better end-user experiences by simplifying infrastructure and operations.

For more information, please go to: www.cisco.com/go/hx-amd

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