

Edge infrastructure has become a strategic priority as AI workloads, data sovereignty requirements, and real-time processing demands push compute closer to where business happens.

The Unified Edge: Simplifying Secure, AI-Ready Infrastructure for the Distributed Enterprise

June 2026

Written by Brandon Butler, Senior Research Manager, Network Infrastructure and Services; and Max Pepper, Senior Research Analyst, Enterprise Workloads, Core Infrastructure

Introduction

Enterprise IT architectures are undergoing a fundamental transformation. The rapid proliferation of AI-driven workloads, real-time processing requirements, and increasingly distributed operations are pushing compute decisively toward the edge of the network, closer to where data is generated, where decisions must be made, and where business outcomes are determined.

The momentum behind this shift is substantial. IDC research shows that 27% of organizations have already deployed edge AI, rising to 38% among those at advanced stages of AI maturity, with an additional 54% planning implementation within the next two years (see Figure 1). The workloads driving this investment are telling: four of the five top enterprise workloads organizations say most often are running at the edge (content delivery, AI life-cycle management, text and media analytics, and unstructured databases) are AI-intensive or directly support AI edge use cases. Also, AI life-cycle management is expected to see the greatest growth in edge spending over the next two years.

The case for running AI at the edge rather than in centralized environments is increasingly clear. Latency-sensitive applications, such as fraud detection, computer vision–based quality control, and AI-assisted diagnostics, cannot tolerate round-trip delays to central datacenters. Data sovereignty requirements mandate local processing in many industries. The cost and bandwidth implications of transmitting growing volumes of raw data upstream are becoming untenable.

However, deploying and managing distributed edge infrastructure at scale remains a formidable challenge. Operational complexity, skills gaps, security requirements, and life-cycle management across hundreds of remote sites continue to constrain progress. A new generation of hyperconverged, AI-ready edge platforms integrating compute, storage, networking, security, and centralized SaaS-based management in a unified system is emerging to address these realities.

AT A GLANCE

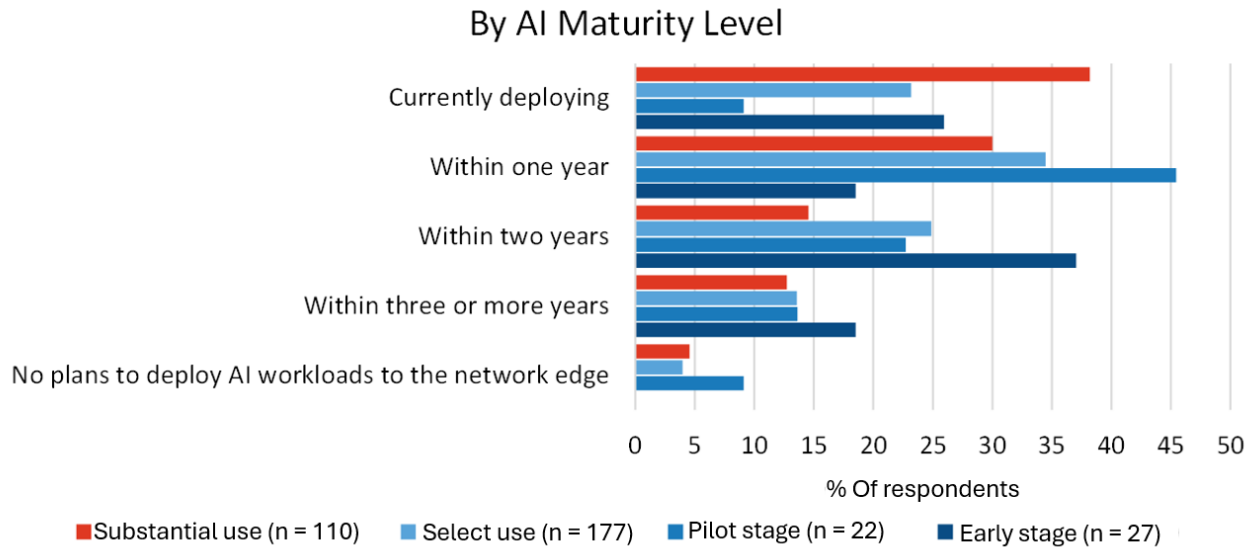
KEY STATS

- » About 27% of organizations have already deployed edge AI, and an additional 54% plan to do so within two years, signaling a rapid shift in enterprise AI strategy.
- » Worldwide edge enterprise infrastructure spending is projected to grow from \$50 billion in 2025 to \$110 billion in 2030, driven primarily by AI workload requirements.
- » Roughly 44% of organizations deploying AI workloads at the edge said they need to invest in more powerful servers to support them, highlighting the gap between legacy infrastructure and the demands of modern edge AI.

This paper examines the trends driving edge investment, challenges organizations face, and capabilities modern unified edge platforms must deliver.

Figure 1: Edge AI adoption is expected to grow significantly in next two years

Q What is your organization's time frame for extending deployment of AI workloads (including GenAI inferencing and applications) to the network edge?



Note:

n = 336

Small samples sizes for early/pilot stage respondents

Source: IDC's Worldwide AI in Networking Special Report, 2025

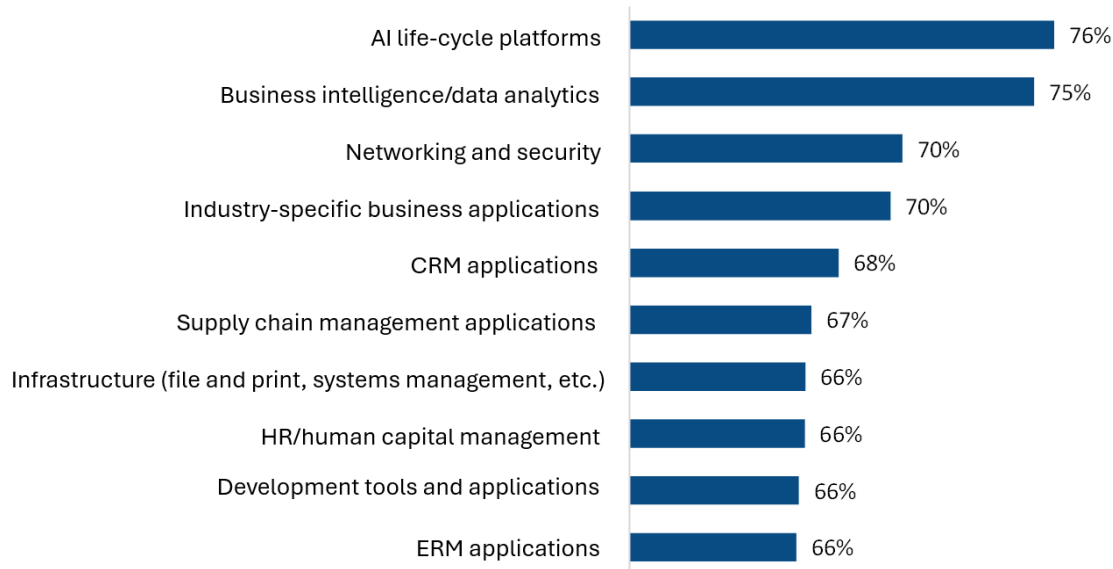
Trends driving the edge infrastructure market

IDC defines edge infrastructure as distributed compute, network, and storage systems located outside centralized cloud datacenters, positioned as close as necessary to where data is generated and consumed. Once a peripheral extension of the enterprise IT stack, the edge is rapidly becoming a primary locus of AI execution, operational intelligence, and business-critical workloads.

Enterprise workload data confirms this evolution. In IDC's 2H25 Enterprise Workloads Infrastructure Survey, AI life-cycle platforms and text and media analytics ranked among the workloads most likely to be deployed at the edge, trailing only content delivery software. The top 3 workloads organizations expect to increase edge spending on over the next two years are AI life-cycle platforms, business intelligence and data analytics, and networking and security, areas in which AI plays a considerable and growing role (see Figure 2).

Figure 2: Top workloads expected to see increased edge investment

Q For workloads that are currently running at the edge, do you expect that investments in the edge infrastructure supporting them will increase, decrease, or remain about the same over the next 2 years? (increase only)



Source: IDC's *Enterprise Workloads Infrastructure Survey for 2H25*, February 2026

IDC forecasts that by 2028, 45% of enterprises will deploy inferencing at the edge, and worldwide edge enterprise infrastructure spending is projected to grow from \$50 billion in 2025 to \$110 billion in 2030 at a CAGR of 17.1%. For IT decision-makers, the implication is clear that a distributed infrastructure strategy is no longer optional.

Driving this investment are use cases across retail, manufacturing, healthcare, and financial services — computer vision, predictive maintenance, AI-assisted diagnostics, and fraud detection — that share a common requirement. They cannot tolerate the latency of routing data to a central datacenter. Real-time responsiveness is a functional necessity, not a performance preference. Data sovereignty mandates and the prohibitive cost of transmitting large raw data volumes upstream further reinforce the case for local processing.

Managing distributed infrastructure at scale remains a defining challenge. About 69% of the respondents consider AI capabilities as a top selection criterion for edge networking solutions, but only 31% of edge network management tasks are currently augmented by AI. Reflecting the urgency to close that gap, 58% of the respondents favor AI-powered, cloud-based management for their core and edge networks, signaling an accelerating shift toward SaaS-delivered operations that reduce dependence on scarce onsite IT expertise.

Challenges in adopting edge infrastructure

Edge environments differ fundamentally from the controlled conditions of a centralized datacenter, and organizations that underestimate this gap do so at considerable cost. Remote and branch sites frequently lack guaranteed power and

cooling, controlled temperatures, protection from dust and vibration, and reliable physical security. Most operate with little or no local IT expertise, making remote management an operational necessity. When something goes wrong and a technician must be dispatched, the costs and disruptions of that truck roll are substantial and, at scale, unsustainable.

Operational complexity grows as deployments span hundreds or thousands of sites. IDC research shows that one in three

About 43% of organizations that deployed AI workloads cited difficulty integrating systems as a top deployment challenge.

edge projects costs more than originally planned, with inconsistent management practices leading to configuration drift, missed security patches, and service disruptions that are difficult to detect and costly to remediate. Distributed infrastructure also expands the attack surface, and inconsistent policy enforcement creates elevated and persistent security vulnerability across remote environments.

Legacy platforms add further friction as AI-intensive workloads arrive at the edge. About 44% of organizations deploying AI-intensive workloads at the edge reported needing to invest in more powerful servers to support them, and 43% cited difficulty integrating systems as a top deployment challenge. Finally, the rapidly evolving application landscape demands platform flexibility that legacy

infrastructure often cannot provide. Organizations need hardware that can adapt to changing software stacks at scale, with many organizations undergoing transformations of traditional virtualization today to containerized or alternative hypervisor environments tomorrow, without forcing full infrastructure replacement.

Key capabilities of edge infrastructure

As organizations scale edge deployments to support AI inferencing, real-time operational workloads, and distributed business applications, the capabilities demanded of edge infrastructure platforms have grown considerably more sophisticated. Purpose-built edge platforms must go well beyond simply placing a server in a remote location. They must deliver an integrated, operationally resilient system that can be deployed, managed, and secured at scale without dependence on local IT expertise. The following capabilities define what modern edge infrastructure must deliver:

- » **Modular, purpose-built hardware design.** Edge platforms should support flexible configurations, including CPU-optimized, storage-optimized, and GPU-enabled nodes, within a compact, low-noise chassis suitable for non-datacenter environments. The ability to mix and match node types and upgrade individual components without replacing the entire system is essential as workload requirements evolve.
- » **Hyperconverged software stack.** Integrating compute, storage, and networking into a unified software-defined layer reduces hardware footprint, simplifies operations, and enables consistent management across distributed sites. VM-based licensing models tied to VM count rather than CPU cores further optimize cost for edge deployments.
- » **Zero-touch provisioning and remote deployment.** The factory pre-installation of hypervisor and OS software allows hardware to be shipped directly to remote sites and initialized without onsite IT expertise, dramatically reducing truck rolls, accelerating rollout, and enabling organizations to scale deployments across large numbers of geographically distributed locations.
- » **Centralized, SaaS-based life-cycle management.** Full-stack life-cycle management covering day-0 provisioning, day-1 deployment, and ongoing day-2 monitoring, patching, and firmware upgrades should be executable remotely from a single management plane without disrupting cluster operations.

- » **Unified visibility across hardware and software layers.** A single management dashboard spanning both hardware telemetry (server health, power consumption, temperature) and software metrics, including hypervisor status, VM performance, and cluster alerts, gives centralized IT teams the situational awareness needed to manage distributed infrastructure efficiently and resolve issues proactively.
- » **Physical and digital security.** Edge deployments require security at multiple layers, from lockable chassis bezels and front-accessible physical controls to centralized policy enforcement, role-based access management, and proactive security advisory monitoring, working in concert to reduce risk across environments in which physical security cannot always be guaranteed.
- » **Support for heterogeneous workloads.** Platforms must accommodate both traditional VM-based applications and modern containerized workloads, including Kubernetes-based AI inferencing pipelines and network virtualization functions, on a single integrated system, eliminating the need to maintain separate infrastructure stacks for different application types.
- » **Proactive support and automated return merchandise authorization.** Integrated telemetry-driven support models, including automatic tech support bundle generation, synchronized case management between hardware and software vendors, and proactive hardware replacement, reduce mean time to resolution and minimize operational disruption at remote sites where onsite intervention is not a viable option.

Benefits of Advanced Edge Infrastructure

Investing in a modern, purpose-built edge platform delivers operational, financial, and strategic returns that are tangible and measurable:

- » **Reduced operational burden and faster deployment.** Automated provisioning, centralized management, and zero-touch deployment models reduce dependence on onsite IT expertise, eliminate staging steps, and allow organizations to bring new edge locations online more rapidly and at lower cost, a critical advantage when scaling across hundreds of sites simultaneously.
- » **Mission-critical performance.** Purpose-built edge platforms with local AI inferencing capabilities deliver the near-real-time processing required for latency-sensitive use cases, from fraud detection and computer vision in manufacturing to clinical patient monitoring in healthcare.
- » **Stronger security posture.** Integrated physical and digital security, centralized policy management, and automated patch delivery reduce exposure across distributed environments where inconsistent manual practices would otherwise create persistent vulnerability.
- » **Lower total cost of ownership and long-term flexibility.** Reduced hardware footprint, VM-based licensing, and the elimination of routine truck rolls lower both capex and opex. Modular, upgradeable platforms and software stack flexibility supporting transitions between virtualization environments, container platforms, and hypervisor stacks protect investments across multiple technology generations without forcing full infrastructure replacement.

Profile: Cisco Unified Edge with Nutanix Cloud Platform

Cisco Unified Edge is a co-engineered, AI-ready edge platform developed in collaboration with Nutanix, integrating Cisco's hardware and management capabilities with Nutanix's Hyperconverged Infrastructure (HCI) software stack into a unified

system. Purpose-built to address the operational, security, and infrastructure challenges that make edge deployments complex and costly, the platform consolidates compute, storage, networking, security, and SaaS-based management in a single modular chassis, with Nutanix Acropolis Operating System (AOS) and Acropolis Hypervisor (AHV) running natively on the infrastructure.

The hardware platform directly addresses a gap in which IDC research confirms that 44% of organizations deploying AI-intensive workloads at the edge said they need to invest in more powerful servers to support them. Key elements of the platform include:

- » **Compute density.** The three-rack unit (3RU) chassis supports up to five hot-swappable compute nodes with Intel Xeon 6 system-on-a-chip (SoC) processors, up to 768GB of memory per node, nonvolatile memory express (NVMe)-based storage, and optional GPU support, providing the compute density required for AI inferencing at the edge.
- » **Built-in redundancy.** Built-in 25GbE redundant internal networking ensures high-speed, resilient node-to-node connectivity, whereas redundancy throughout power supplies, cooling fans, and network uplinks ensures availability where onsite intervention is not readily available.
- » **Edge-ready design.** A lockable bezel, front-accessible components, and acoustically optimized fans make the platform well suited for branch offices, retail locations, and light industrial facilities.

Addressing the skills gap and operational complexity that plague large-scale edge deployments, hardware is shipped from the factory with Nutanix AOS and AHV pre-installed. Clusters of one to five nodes can be provisioned and initialized remotely via Cisco Intersight and Nutanix Foundation Central, enabling large-scale deployments with no onsite IT expertise required and eliminating the costly truck rolls that drive up deployment complexity at scale.

For ongoing operations, Intersight serves as the centralized SaaS-based management plane, directly addressing the visibility and integration gaps that IDC identified as a top challenge for 43% of edge deployers. A single dashboard spans hardware telemetry, Nutanix cluster visibility, security advisories, life-cycle management, and role-based access control across edge and core environments. For organizations with data sovereignty or air-gapped requirements, Intersight also supports on-premises deployment via its Private Virtual Appliance (PVA). The integration of Nutanix Life Cycle Manager with Intersight enables automated, rolling firmware and software upgrades without cluster downtime, mitigating the configuration drift and missed patch risks that create persistent security exposure across distributed sites.

The platform supports both VM-based workloads via Nutanix Cloud Infrastructure and containerized workloads via the Nutanix Kubernetes Platform on the same hardware, allowing organizations to run traditional applications today and transition to modern environments as needs evolve without hardware replacement. A coordinated support model between Cisco and Nutanix provides a single point of contact, synchronized case management, automatic tech support bundle uploads, and proactive return merchandise authorization (RMA), which reduces resolution times and operational burden at remote sites.

The solution is designed for remote office and branch office environments, including in retail, healthcare, manufacturing, and mining, in which ease of deployment, centralized management, and integrated security are priorities.

Challenges Cisco and Nutanix will face in bringing this solution to market

Translating a technically strong solution into broad market adoption requires navigating challenges that are as much commercial as technical. The unified edge HCI space is increasingly crowded, with established players such as Dell Technologies, Hewlett Packard Enterprise (HPE), and existing Nutanix OEM partners offering overlapping capabilities, making differentiation an ongoing requirement.

Customer inertia presents an equally significant obstacle. Many enterprises have existing edge infrastructure investments and established vendor relationships, and migration to a new platform requires a compelling total cost of ownership case and confidence in transition support.

Finally, scaling the solution will depend on channel and ecosystem readiness. The co-engineered platform represents a go-to-market motion that differs from either vendor's traditional approach, and broad partner enablement will be essential to reach the industries and geographies in which edge demand is growing.

Conclusion

Edge infrastructure has moved from a peripheral IT consideration to a strategic priority, driven by AI inferencing workloads, data sovereignty requirements, and the operational demands of highly distributed enterprise environments. Legacy approaches characterized by siloed management, limited automation, and general-purpose hardware are increasingly inadequate for the scale and complexity modern edge deployments demand.

Purpose-built, hyperconverged edge platforms that unify compute, storage, networking, and centralized management represent the next evolution, enabling organizations to deploy and operate at scale without proportional growth in IT staffing. The Cisco Unified Edge with Nutanix HCI exemplifies this approach, combining co-engineered hardware and software with zero-touch deployment, full-stack life-cycle management, and unified visibility via Cisco Intersight.

The shift to AI at the edge is not a future consideration; it is underway now. With 27% of organizations already deploying edge AI and an additional 54% planning to do so within two years, the infrastructure choices made today will determine whether organizations can respond at the speed that competitive and operational pressures demand.

About the Analyst



Brandon Butler, Senior Research Manager

Brandon Butler's research focuses on market and technology trends, forecasts, and competitive analysis in enterprise campus and branch networks. His coverage includes technologies used in local and WAN, such as Ethernet switching, routing/software-defined-WAN, wireless LAN, and enterprise network management platforms.



Max Pepper, Senior Research Manager

Max Pepper is a senior research analyst in IDC's enterprise infrastructure research domain, focusing on the enterprise workload infrastructure market. His work includes contributing to IDC's Semiannual Enterprise Infrastructure Tracker: Workloads, covering computing and storage trends for virtualized and containerized workloads, cloud-native applications, and edge deployments.

MESSAGE FROM THE SPONSOR

The rapid expansion of edge computing is driven by the need for real-time data processing and localized AI capabilities. Customers today are increasingly concerned with the complexity of managing distributed infrastructure, ensuring high availability at remote sites, and maintaining consistent security policies across their entire footprint.

The Cisco and Nutanix edge solution addresses these challenges by integrating Cisco's AI-ready Unified Edge platform and cloud-based Intersight management with the industry-leading HCI software Nutanix Cloud Platform. This hyperconverged infrastructure (HCI) provides a unified, software-defined environment that simplifies operations from the datacenter to the edge.

By combining Cisco's robust networking and compute hardware with Nutanix's flexible software stack, organizations can achieve greater agility and improved cost efficiency. This collaboration empowers businesses to deploy applications rapidly, ensuring consistent, secure, and reliable performance at the edge, ultimately driving better business outcomes through a simplified and resilient IT architecture.

To learn more, visit [Cisco Compute Hyperconverged with Nutanix](#) and nutanix.com/cisco.

 IDC Custom Solutions

The content in this paper was adapted from existing IDC research published on www.idc.com.

IDC Research, Inc.
140 Kendrick Street
Building B
Needham, MA 02494, USA
T 508.872.8200
F 508.935.4015
blogs.idc.com
www.idc.com

IDC Custom Solutions produced this publication. The opinion, analysis, and research results presented herein are drawn from more detailed research and analysis that IDC independently conducted and published, unless specific vendor sponsorship is noted. IDC Custom Solutions makes IDC content available in a wide range of formats for distribution by various companies. This IDC material is licensed for external use, and in no way does the use or publication of IDC research indicate IDC's endorsement of the sponsor's or licensee's products or strategies.

International Data Corporation (IDC) is the premier global provider of market intelligence, advisory services, and events for the information technology, telecommunications, and consumer technology markets. With more than 1,300 analysts worldwide, IDC offers global, regional, and local expertise on technology and industry opportunities and trends in over 110 countries. IDC's analysis and insight helps IT professionals, business executives, and the investment community to make fact-based technology decisions and to achieve their key business objectives.

©2026 IDC. Reproduction is forbidden unless authorized. All rights reserved. [CCPA](#)

