White Paper

Transforming the Enterprise Edge: The Critical Role of Unified Edge Infrastructure for the AI Era

Sponsored by: Cisco

Brandon Butler Jennifer Cooke

September 2025

EXECUTIVE SUMMARY

The rise of the AI era has thrust organizations into assessing how to optimize data-driven insights to accelerate their business goals. As they do so, IDC research shows that edge infrastructure has emerged as a critical component of their strategy.

Edge infrastructure and technology is supporting key business initiatives such as creating better customer experiences and protecting sensitive organizational and customer data. Edge infrastructure is driving significant improvements in operational efficiency and reduced data transmission costs across several industries. The benefits of edge infrastructure are measured by increased efficiency, lower costs, and greater control over data. The leading benefits of edge infrastructure are shown in Figure 1.

September 2025, IDC #US53740925

FIGURE 1

What Organizations Gain from Edge Investments

n = 800

Source: IDC's *Edge View Survey,* 2025

Edge infrastructure refers to the distributed infrastructure and applications outside of centralized cloud datacenters located closer to where data is generated and consumed. Edge infrastructure speeds the processing of data at the edge of the network. For latency-sensitive use cases, this is critical. Edge infrastructure supports near-real-time insights and actions without having to transmit data to a central datacenter or cloud platform. When data is sent to the cloud, only the most important information is transmitted, thereby reducing transmission costs.

There are a variety of factors driving edge transformation. Most notable are AI workloads, and particularly those that are latency sensitive, such as real-time data processing for fraud detection, machine sensors and video analytics in manufacturing for quality control and predictive maintenance, and in-store retail analytics to create new customer experiences. New AI workloads at the edge require high performance and better resilience, with critical operations and business processes relying on edge infrastructure. Greater efficiency and cost optimization are key benefits and will drive continued deployment of new edge use cases. Security and sovereignty considerations related to the gathering, analyzing, and storing of data in specific geographical areas for

compliance or regulatory reasons are another important driver. Edge IT is the foundation upon which digitally driven insights and operations are built.

The demands on edge infrastructure are increasing. New AI workloads require fast compute performance and higher availability. Many of the new workloads at the edge require higher levels of resilience and uptime because critical processes and operations depend on these resources. The deployment and ongoing management of distributed edge infrastructure introduces operational challenges that escalate as projects scale. Edge platforms today must support streamlined and zero-touch deployment, remote visibility, and troubleshooting and include ongoing support for life-cycle management. Overall, there are a variety of key opportunities relating to edge:

- AI is driving increased demand for real-time, latency-sensitive computing at the edge.
- Current edge deployments have a myriad of challenges: solutions are not optimized for the edge use cases and are difficult to deploy and manage.
- There's a need for unified, platform-based edge infrastructure systems with scalable operational models.
- Edge infrastructure should be a full-stack solution, inclusive of third-party ecosystems' integrations and targeted vertical use case designs.

This IDC white paper explores the critical importance of edge infrastructure in the AI era, review the key factors driving edge transformation, and the challenges organizations face in edge deployments. The white paper also includes detailed insights from multiple end users in manufacturing and healthcare industries where edge transformation is prominent and provides guidance to enterprises on how to succeed in edge transformation initiatives.

SITUATION OVERVIEW

The Critical Importance of Edge in the AI Era

As business and operations shift to more data-driven insights, the need for secure infrastructure platforms in close proximity to where data is generated or stored becomes increasingly important. Edge infrastructure minimizes latency, optimizes bandwidth utilization, and enhances data privacy, making it a crucial enabler for the widespread and effective deployment of AI inference use cases. This is especially important for use cases and applications that require real-time processing, high data volumes, and stringent security requirements.

AI-infused applications are all about extracting insights and understanding from data. Massive volumes of data are being created by machines, people, and a myriad of

devices — all of which exist outside of datacenters. The ability to gather, analyze, and protect data at the edge is a critical capability for modern IT organizations. When asked what they needed to do to prepare to run AI at the edge, 53% of organizations said they needed to upgrade their compute infrastructure (source: IDC's *Edge View Survey,* 2025; n = 800). On average, organizations expect to be storing and protecting 1.6PB of data at the edge in 2027, up from 544TB that they store today. The demands for performance and security are heightening awareness of the need for edge infrastructure that is up to the task.

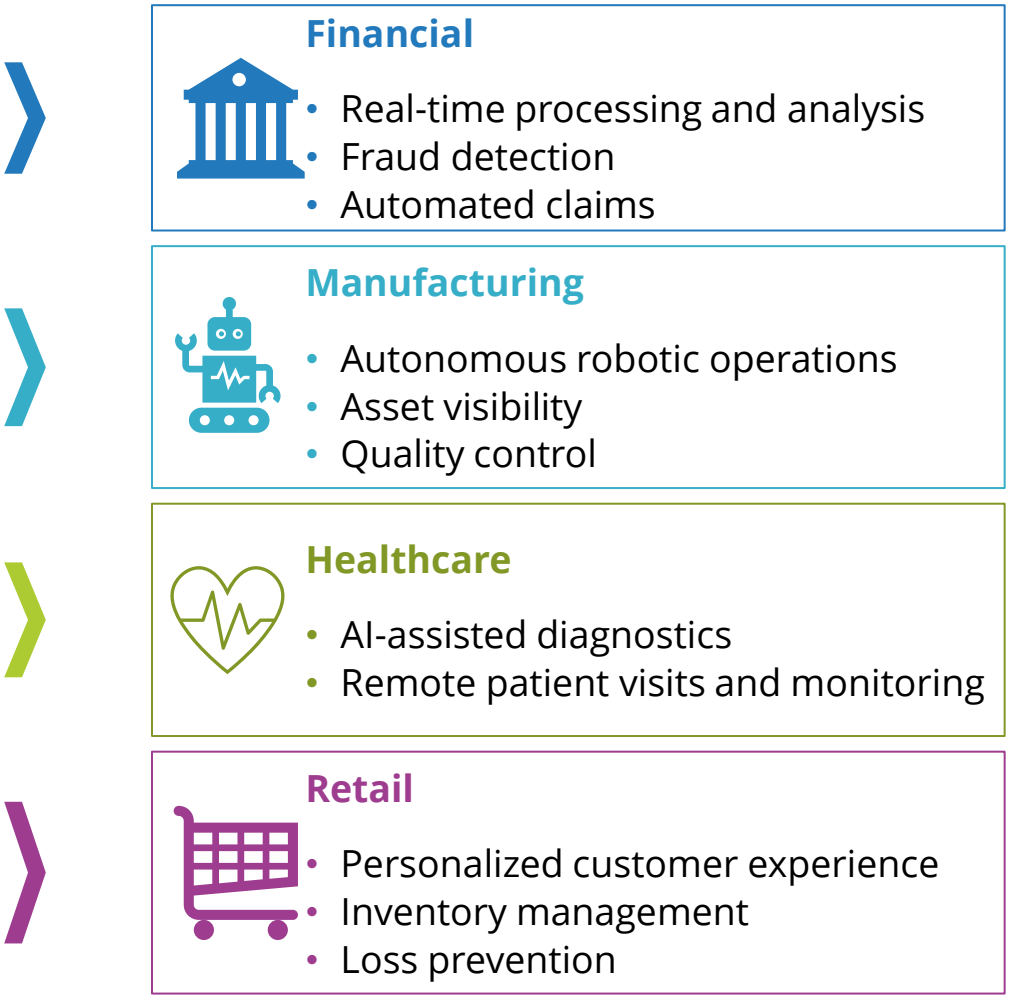
AI is driving a greater need for secure, resilient edge platforms driven by AI use cases that require real-time, latency-sensitive collection and inferencing of data closer to where it is collected/consumed. But edge deployments also introduce additional deployment and ongoing management complexity. The sheer scale of edge deployments is taxing IT organizations. The logistics of deploying infrastructure in numerous regions and in geographically distributed locations, often with little or no IT staff, will curtail progress. There's a big difference between deploying 100 servers at 1 location and deploying 1 server in 100 locations. The time and complexity of not only setting up but performing ongoing maintenance (patches and security updates) is overwhelming for most organizations. The complexity of integrating and assembling edge platforms and ensuring that they remain updated is a major challenge. With data security at stake, and operations often reliant on the uptime of compute resources, edge infrastructure needs to be highly resilient. The unique demands of edge computing are underscoring the need for modernized platforms. Full-stack solutions that are prebuilt with third-party integrations, comprehensive integrated security, and validated solutions are key. Each industry and use case will put unique demands on edge infrastructure. Over time, the use cases will grow and change. Edge systems need to be flexible to adapt to these evolving requirements.

Factors Driving Edge Deployments

Edge infrastructure is deployed in a wide variety of use cases. Typical edge locations include retail, factory floors, branch offices, hospitals, warehouses, along railways, on cruise ships, and a myriad of unique locations where data is created and stored. Some have industry use case–specific requirements. Edge infrastructure can be in many forms. What are the common traits in modernized edge solutions? The ability to be remotely monitored and managed, and the ability to operate with a degree of autonomy, as most edge locations do not have IT staff nearby. Further, the wide variety of edge use cases necessitates having infrastructure platforms that can meet the needs of different use cases. Figure 2 shows examples of key edge use cases by industry.

FIGURE 2

Edge Use Cases by Industry



Source: IDC, 2025

Another common thread among edge use cases is the performance needs of edge workloads, notably real-time processing of data for latency-sensitive applications. Modern edge infrastructure should be capable of supporting near-real-time operations and should support unified management and optimized connections to core and cloud IT resources.

Securing data and protecting company IP is a high priority for all IT organizations. Unified resources that streamline updates and ensure all security patches are performed are table stakes for edge use cases in all industries. Physical security varies by location, with most edge deployments factoring in vulnerabilities that are present

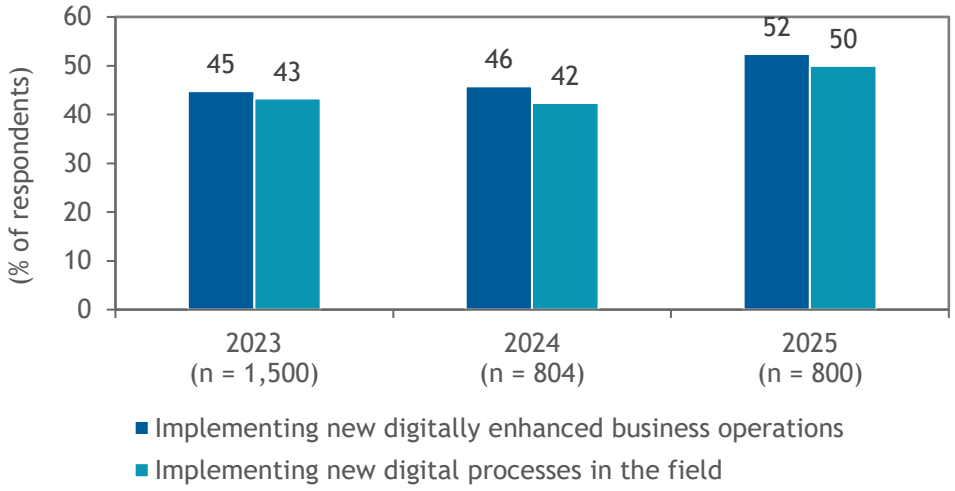
with IT gear deployed outside of datacenters. Modern data management practices and integrated system designs reduce the risk of data loss or breaches.

Resilient edge resources should be built for highly distributed and diverse locations and enable IT organizations to maximize the value of data through support across cloud, edge, and core locations.

IDC observed a sharp uptick in 2025 of organizations using edge for digitally enhanced business operations and new processes in field locations (see Figure 3). The additional investment signals that edge projects are maturing and providing operational efficiency gains. The research found that 54% of organizations say that they've observed improved operational efficiency from their edge projects, 51% of organizations report improved data security, and 42% of organizations reported better business resilience (source: IDC's *Edge View Survey,* 2025; n = 800).

FIGURE 3

Drivers for Edge Investment, 2023–2025



Source: IDC, 2025

CUSTOMER INSIGHTS: USE CASES DRIVING EDGE INVESTMENTS

IDC interviewed two senior IT executives regarding their current and future edge infrastructure strategies. There are a variety of factors driving executives toward increasing their use of edge infrastructure. Most revolve around the need for real-time processing of latency-sensitive operational workloads, reducing the amount of data traversing the network for data-heavy use cases, leveraging AI tools to improve operations and creating new business opportunities.

Global Industrial Manufacturing CIO

The CIO of a global manufacturing organization noted that over the past two decades, his organizations moved a variety of applications and workloads to the shared cloud resources. But, in recent years, there's been a reckoning that high-performance edge infrastructure is needed for mission-critical use cases. "We are continuously monitoring the performance of these machines … through sensors measuring the machine's status — up, down, scrap rate, benchmarking data," he noted. "The tests need to be done in real time, (because) the factory machinery makes real-time decisions, which potentially change production parameters. So, we cannot have latency, we need high performance, we cannot have connectivity issues to a remote datacenter or cloud."

The company operates about 65 sites, including 20 in the United States and Canada, with production facilities ranging in size of 200,000-300,000 sq ft. In addition to processing machine production data on the edge infrastructure, the company also runs AI-powered maintenance applications, as well as health and safety applications tied to video surveillance feeds. The company is exploring a variety of additional use cases too: "As we deploy more edge and AI applications, that will necessitate more edge infrastructure because most of the applications we use or plan to use require real-time processing of data for actionable inputs and autonomous decision-making, which cannot have latency," the CIO noted.

When it comes to edge infrastructure to support these use cases, the CIO was clear: "What we want is a self-contained, turnkey, easy-to-deploy infrastructure, with compute (servers), storage, and network; start up the server and it's ready to go," he noted. "The benefit is cost savings via easy manageability … plus, we don't need too many resources to deploy and operate it." He added that other key capabilities include having centralized management of distributed edge infrastructure at multiple sites, including visibility and analytics into the infrastructure performance, along with an ability to remotely diagnose and troubleshoot issues that arise.

Vice President of Architecture and Infrastructure at a United States–Based Healthcare Organization

The healthcare infrastructure VP said there are a variety of use cases that would drive his organization to consider edge infrastructure. For one, researchers at his organization have atomic-level cameras that generate more than a petabyte of data per day. "That absolutely needs some sort of edge device" for processing; it's untenable, he said, to build infrastructure to support that on premises or to traverse the data to a shared cloud resource. Hospitals within his organization are increasingly using edge infrastructure for patient-facing use cases, such as real-time monitoring of patients who need 24-hour surveillance. "This edge infrastructure is monitoring the telemetry of

the patient themselves — if the skin changes color or breathing is off — it uses computer vision and some AI to automatically kick off an alert that goes right to the appropriate clinician. But for me to pipe that all from a patient's room to a datacenter, that's not going to work. I'm going to need some sort of edge device to manage that," he said, noting the life-saving criticality of the workload necessitates a need for real-time processing.

In terms of the infrastructure to power edge use cases, the executive noted that given the wide variety of use cases the hospital system needs to support — from clinical to research and patient facing to internal — having flexible systems that enable modifications of the infrastructure as needs change would be an important quality. "It's all application driven," he noted. Ease of management, high availability/redundancy, and integrated security are other top qualities he noted of ideal edge infrastructure. "What the clinicians want is for it just to work," he said. "You want clinicians focused on the patients, period. The mission of my department is to make sure the technology works so we can let the end users focus on what they do best."

FACTORS TO CONSIDER IN EDGE DEPLOYMENTS

Challenges with Legacy Approaches to Edge Infrastructure

Despite the continued momentum behind edge infrastructure deployments, the market continues to be challenged by a variety of factors. Key challenges with legacy approaches to edge deployments include:

- **Infrastructure constraints:** Existing edge infrastructure is not typically optimized nor flexible enough for modern edge use cases and workloads. In addition, many edge use cases have environmental constraints related to low power and limited footprints. In an IDC study of 800 organizations with edge deployments, one in five reported that they had difficulty diagnosing and servicing equipment remotely. In field locations, one in four have difficulty servicing equipment and 24% reported damage due to the surrounding environment (see Figure 4). Edge workloads and use cases inherently increase attack surface and expand security vulnerabilities.
- **Deployment and operational complexity:** The complexity of edge deployments frequently causes projects to take longer and cost more than originally planned. Staging sites and truck rolls take time and resources. Configuration "drift," where system changes happen gradually and often unnoticed, introduces security risk. If not managed well, large-scale updates can cause service disruptions. IDC's research shows that one in three edge projects cost more than planned. Why? Half reported that they needed to upgrade their compute infrastructure first and 43% said they had difficulty integrating systems.

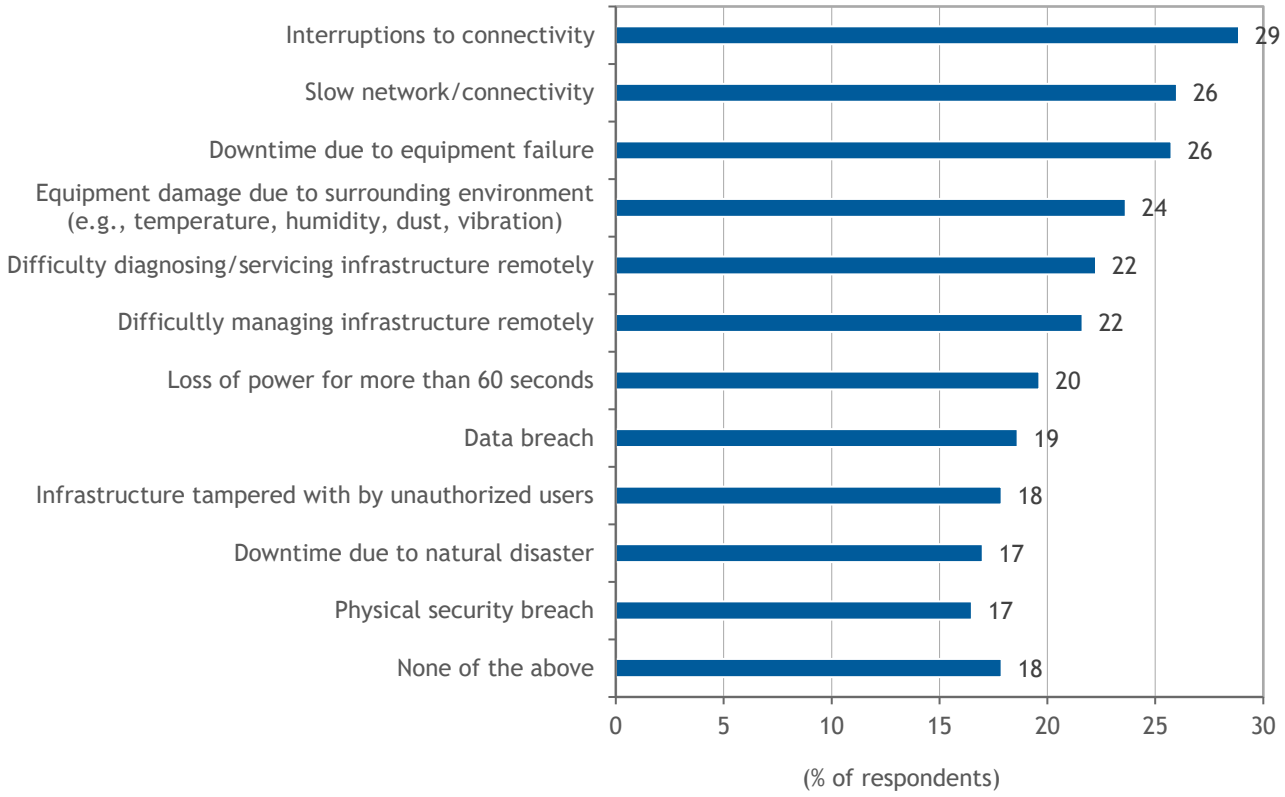
The scale and reach of edge projects amplify the difficulty of integrations and deployments.

- **Limited skill sets at the edge:** Manual or disparate management of edge resources leads to configuration and policy drift. A lack of coordination among technology teams leads to operational and performance inefficiencies. The downside to this management gap ranges from cost and wasted time at best and interruptions to business and personal safety risk at worst. 54% of organizations said that edge disruption led to disruption to operations. 41% reported reduced profitability and 31% reported being fined for noncompliance. Most concerning was the 33% that reported people's safety was put at risk. The more organizations rely on digital technologies, the more important it becomes to ensure resilience.
- **Lack of integrations:** A major contributing factor to problems at the edge is a lack of integration support across compute, network, storage, and security. Connecting IT with operating technologies is a major challenge as well, leading to technical inefficiencies related to interoperability, visibility, management, automation, and upgrades.

FIGURE 4

Challenges at the Edge

Q. Has your organization experienced any of the following with its edge infrastructure deployments?



n = 800

Source: IDC's *Edge View Survey,* 2025

Key Technology and Operational Considerations for Edge Infrastructure

As organizations consider ways to implement an edge solution, there are a variety of considerations related to what makes a modern edge infrastructure different from legacy approaches. Some important elements to consider are:

- **Modular and unified platforms:** Given the diversity of edge use cases and environments, it's critical that edge infrastructure be modular to allow for customization. For example, some customers may want more compute, other may want to run GPUs in addition to CPUs, while others may need more storage capacities depending on their use case. Within this context, however, it's also important that the modular components of an edge infrastructure stack can be

managed as a cohesive stack, enabling centralized visibility, management, and automation.

- **Operational simplicity:** The life cycle of managing an edge deployment is a critical factor to consider. Starting with the initial deployment, edge platforms should leverage full-stack blueprints and include zero-touch provisioning capabilities. This ensures proper and efficient onboarding without a "truck roll," which is key given that edge infrastructure is typically deployed at sites with little or no IT staff. From an ongoing management perspective, automatic updates, comprehensive and real-time visibility into infrastructure health, and assurance capabilities should be native. Combined, these capabilities enable any issues or vulnerabilities to be identified and fixed before they impact operations. Edge platforms should have an ability to be centrally managed, with policy-based security and automation.
- **AI-powered management**: A key benefit of a unified platform-based approach to edge infrastructure is to leverage AI-powered capabilities for managing the deployment and ongoing operations of the infrastructure. Key AI-powered management capabilities include anomaly detection, guided or automated problem resolution, proactive maintenance and patches, performance optimizations, and security enhancements.
- **Physical and digital security:** Comprehensive physical and digital security safeguards applications and data at the edge. It's achieved by having security capabilities natively built into the edge infrastructure, such as a firewall and intrusion prevention system and intrusion detection system (IPS/IDS), and ensuring strong, consistent, and dynamic policies can be centrally managed. Another consideration of physical security is enclosures that prevent accidental and intentional tampering and technology to remotely enable and disable access to systems and track location.
- **Multilayered application security:** Beyond physical and digital security, edge systems must also protect workloads and data. Organizations need to ensure that only the right users/agents/devices get access to applications and that the workload is protected at the network level. In addition, there is a need for assessment and protection for AI models against adversarial attacks and model vulnerabilities. Plus, security for Kubernetes and cloud-native environments is needed to protect, monitor, and scale modern applications running in distributed, containerized environments.
- **Seamless integration with third-party technologies:** A comprehensive edge infrastructure platform would include the compute, network, storage, and security capabilities, but those still must work within the ecosystem of the customer's environment. Having native integrations with third-party tools, from

silicon to IT automation, there are a range of integrations that are valuable for the edge platform to leverage.

- **Centralized fleet management of distributed sites**: A critical factor contributing to operational complexity is the lack of centralized management across disparate sites. Leveraging SaaS-based and cloud-delivered infrastructure systems enables centralized policy creation and management, global insights, and analytics.
- **Future-proof strategies**: Given the rapid pace of development of AI technology, organizations are looking for platform-based approaches that allow them to not only meet their business needs today but also are flexible enough to support future use cases. Embracing a modular platform enables the ability to upgrade individual components (e.g., CPUs and GPUs) without having to replace the entire system.
- **Tested solutions for specific industry use cases:** Given the potential complexities of edge infrastructure, having pre-validated designs for edge infrastructure deployments, as well as vertical-specific capabilities can ease operational and technical implementation and ongoing management.

What Is Gained from Edge Platforms

Unified edge infrastructure can enable a variety of benefits. Although many projects are still in the pipeline or yet to be discovered, the edge projects in full production to date are delivering measurable returns. Figure 5 shows the benefits and value of edge projects. Better performance, increased efficiency, and improved data security are the key criteria on which edge projects are measured. Other benefits of unified edge infrastructure include:

- **Simplified, scalable, and flexible deployment options**: A pre-validated, highly automated operational model supports consistent, repeatable infrastructure deployments and management across multiple diverse sites.
- **Automated management**: Automation that reduces the need for manual intervention, eliminates configuration drift, and reduces the need for on-site IT expertise can be very valuable in edge deployments, particularly when managing a large number of distributed edge environments simultaneously.
- **Comprehensive security**: Having security built into the system that enables strong, consistent, and dynamic policies at scale is critical for edge infrastructure.
- **Move at the speed of business:** As organizations explore a range of use cases for their business, some will have mission-critical performance needs and be latency sensitive or require real-time processing of data. Having right-sized infrastructure for individual use cases, with an ability to change those as business needs change, is a key quality of modular unified edge infrastructure.

FIGURE 5

Measuring the Value of Edge Projects

n = 800

Source: IDC's *Edge View Survey,* 2025

CONCLUSION

Edge computing is a key, foundational element of modern digital business and operations. The volume of data at the edge is growing rapidly, and the ability to gather, analyze, and react to data in near real time is critical.

Today, edge compute supports innovation in the business through more accurate, AI-driven insights and digitally enhanced operations. Edge use cases today are tightly linked with critical and mission-critical operations, underscoring the need for resilient, modernized platforms. Edge infrastructure and technology is supporting key business initiatives such as creating better customer experiences and protecting sensitive organizational and customer data.

A highly distributed IT ecosystem introduces new challenges in building, deploying, and maintaining remote infrastructure. The scale and reach of edge deployments creates a significant burden on IT organizations. Without rigorous management, data security is at risk. And edge environments often come with unavoidable physical security risks. For these reasons, infrastructure that is designed to streamline and simplify deployment is preferred. Modernized edge platforms automate management, ensuring that configurations remain intact and that systems can remain updated even when no IT staff is physically present.

Building an effective strategy for edge infrastructure choices begins with understanding the unique needs of highly distributed IT landscapes and industry-specific needs. Infrastructure and architecture choices need to consider the level of use case criticality and prioritize platforms that ensure data and physical security requirements. Unified, automated platforms that reduce management burden while enabling greater use of data and analytics are well suited to support the shift to data-driven decisions and digitally transformed operations.

MESSAGE FROM THE SPONSOR

The emergence of the edge as the intersection of physical and digital environments requires real-time data processing for effective decision making and productivity. Legacy infrastructure is not equipped to meet the demands of AI-driven operations, especially as inferencing increasingly shifts from data centers to the edge. Challenges such as inconsistency, inflexibility, and security complicate edge AI implementation. The scale and complexity of distributed edge environments necessitate a new approach to infrastructure and operations. Integrated solutions that combine compute, networking, security, storage, observability, and cloud management are needed to address these evolving requirements.

To explore solutions for edge computing and AI, visit Cisco Computing

ABOUT IDC

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, IT benchmarking and sourcing, 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. Founded in 1964, IDC is a wholly owned subsidiary of International Data Group (IDG, Inc.).

Global Headquarters

140 Kendrick Street
Building B
Needham, MA 02494
USA
508.872.8200
Twitter: @IDC
blogs.idc.com
www.idc.com

Copyright Notice

External Publication of IDC Information and Data — Any IDC information that is to be used in advertising, press releases, or promotional materials requires prior written approval from the appropriate IDC Vice President or Country Manager. A draft of the proposed document should accompany any such request. IDC reserves the right to deny approval of external usage for any reason.

Copyright 2025 IDC. Reproduction without written permission is completely forbidden.