

Modernize your Edge Operations with Red Hat and Cisco



The move to the network edge

Global spending on edge computing has already been projected to reach

US\$261 billion in 2025 and **US\$380 billion by 2028**,¹ and organizations that invested in edge computing have:²

- Improved operational efficiency (54%).
- Improved data security (51%).

¹ IDC press release. “[IDC Estimates Global Spending on Edge Computing to Grow at 13.8% Reaching Nearly \\$380 Billion by 2028.](#)” 17 Mar. 2025.

² IDC Market Forecast. “[Worldwide Edge Enterprise Infrastructure Forecast, 2025–2028.](#)” Document #US52263925 May 2025. (purchase required)

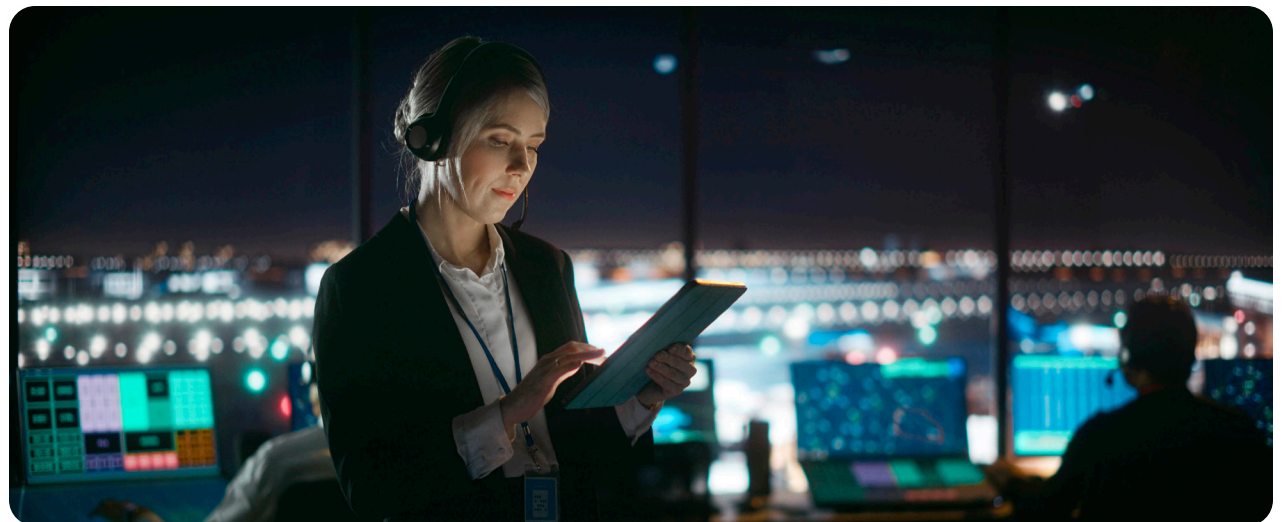
Edge computing will be the key to success for many organizations

Deploying complex computing workloads at the edge of networks with localized data storage and processing has quickly become a key strategy for data-powered organizations in the industrial, retail, and healthcare industries, among others.

More and more devices, sensors, and cameras are being used to collect vital operational data, power key processes, and create significant operational efficiencies. This ranges from supporting predictive equipment maintenance in industrial environments to tracking real-time inventory in retail locations to offering modern, data-powered healthcare.

As technology has evolved to meet the demands of modern edge operations, it has become much more viable to process that data where it is collected, rather than transmitting it back to datacenters. By minimizing data transmissions and reducing the attack surface, this helps improve performance, reduce costs, and mitigate the threat of security risks.

Yet, as the volume of workloads and data at the edge continues to increase, the skill sets required to support this approach are not necessarily growing at the rate they need to, and the rise of AI has introduced more operational complexity and potential security risks.



The critical role of unified edge infrastructure

Unified edge infrastructure has emerged as a critical component of edge transformation, and has been proven to provide key competitive advantages:³

- Better performance.
- Improvements in operational efficiency.
- Increased data security.
- Improved business or operational resilience.
- Reduced cost of data transmission.
- Ability to attract new customers.
- Improved physical safety.

³ IDC White Paper, sponsored by Cisco. “[Transforming the Enterprise Edge: The Critical Role of Unified Edge Infrastructure for the AI Era.](#)” Document #US53740925, Sep. 2025.

Edge modernization brings opportunity and also added complexity

As organizations continue to modernize their edge operations—adopting more complex Operational Technology (OT) solutions in tandem with modern, innovative IT approaches to support them—their challenges have only become more intricate. This includes:

- **Integration and system gaps.** If not managed properly, bringing together modern edge IT systems with established OT infrastructure can result in a lack of cohesion and integration across the technology stack. This in turn can create isolated data, technical debt, and interoperability and management inefficiencies.
- **Scaling with consistency, reliability, and security.** Fleets of edge devices can often number in the hundreds or thousands, especially in the retail industry, where keeping track of large inventory caches or chains of stores requires massive fleets. Managing and safeguarding this many distributed, resource-limited edge devices and the volumes of data they collect can cause strain on IT operations, creating significant performance, configuration, and security challenges.
- **Disparate skills and expertise.** Designing, deploying, and managing modern edge environments requires a diverse range of specific skill sets, especially as IT and OT systems begin to converge. This technology convergence often brings together professions with drastically different fields of expertise, ways of working, and tools, which can make it difficult to collaborate.
- **Deployment and management complexity.** IT personnel do not often operate where these edge devices live, and on-site staff are usually equipped with specialized OT skill sets specific to the equipment they use. This can complicate the deployment and management of complex edge systems—leaving organizations to either bring outside experts into these remote locations, or find remote management options to help set up, diagnose, and service equipment.

These challenges have caused significant complications for industries that rely on edge operations to collect and process their vital operational data. As a result, many of these organizations are looking for a complete edge platform built for modern needs.

The value of centralized cloud and edge management

Cisco Intersight, the IT operations platform for Cisco Unified Edge, has been proven to provide significant operational improvements, including:⁴

- 50% reduction in mean time to recovery.
- 80% reduction in build and deployment times.
- Accelerated time to value resulting in an incremental rise of US\$289,000 in profit.
- Reduced security and compliance risk valued at US\$267,000.

Support edge workloads now and in the future

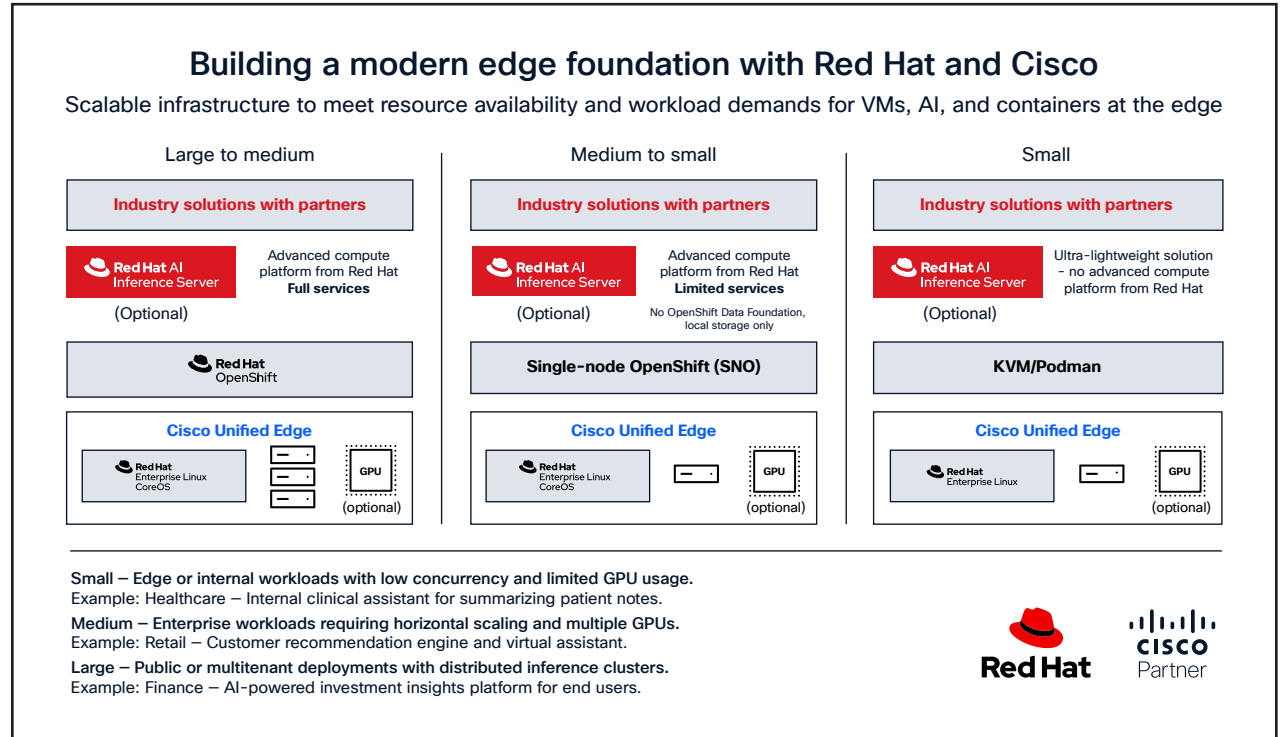


Figure 1. Modernizing the edge with Cisco Unified Edge and Red Hat.

Red Hat and Cisco have partnered to create a joint solution that is built for the specific needs of the edge, bringing together Cisco’s purpose-built hardware with Red Hat’s flexible, scalable software to support the edge workloads of today and tomorrow.

The Cisco Unified Edge solution powered by the advanced compute platform from Red Hat and Red Hat AI Inference Server delivers a

consistent, reliable, and standardized foundation for critical and AI-powered workloads across datacenter and edge environments. It delivers unified edge infrastructure—including the required computing, networking, storage, and security capabilities—that supports performance, scalability, and a focus on security, and offers simplified deployment, management, and lifecycle operations.

⁴ Forrester Consulting, commissioned by Cisco. “The Total Economic Impact™ Of Cisco Intersight.” Nov. 2025.

This hardware and software combination provides a consumable experience that allows users to onboard new technologies—such as Virtual Machines (VMs), containerized applications, and innovative AI-powered applications—without needing to completely replace or rebuild existing, business-critical workloads.

The platform also addresses a fundamental need created by the shift to data-driven operations: the rise of AI and gen AI models in diverse environments, including the edge. It helps accelerate the adoption and management of traditional AI workloads while also supporting gen AI workloads with the use of AI Inference Server. This combination allows organizations to deploy, scale, and operate AI-powered applications with confidence and control anywhere across the hybrid edge.

This end-to-end platform, while specifically architected to meet the unique needs of each organization, is primarily composed of a few key components. This includes Cisco Unified Edge and the advanced compute platform from Red Hat, as well as an IT operations platform that is fully integrated with the enterprise-wide automation capabilities of Red Hat® Ansible® Automation Platform, and for organizations ready to accelerate their AI journey, AI Inference Server.

Cisco Unified Edge is a full-stack, converged edge computing architecture that combines compute, networking, and storage into a single AI-ready platform. It supports modern edge computing with optimized datacenter performance, modularity, centralized management, and built-in security focus at the edge. Cisco Unified Edge's modular design, high-performance networks, and prevalidated solutions deliver consistent, agile, and uninterrupted edge operations.

The advanced compute platform is built on the foundation of Red Hat's open source solutions, and is designed to simplify deployment, maintenance, and operations at the edge. It offers the interoperability, automation, and security focus needed to modernize edge operations as organizations prepare for the future of their respective industries.

Cisco Intersight is the IT operations platform for Cisco's computing portfolio, including Cisco Unified Edge. It streamlines the delivery of resources for both traditional and AI workloads from datacenters to the edge and unifies and simplifies IT management with a comprehensive view, precise controls, and fully integrated automation capabilities through Ansible Automation Platform.

AI Inference Server optimizes model inference across environments, including remote edge locations, to support more efficient and cost-effective AI model deployments. It helps improve performance per allocated investment by optimizing throughput and latency and minimizing compute utilization while preserving accuracy. It also offers a preoptimized model repository to help simplify model deployment.

This offering from Cisco and Red Hat provides a unique, all-encompassing, edge-optimized solution that meets modern edge hardware requirements for power, cooling, acoustics, space, and physical resilience, while offering the necessary computing capabilities and security focus to support all workloads at the edge, including AI.

These combined capabilities help organizations modernize edge operations on an AI-ready platform that supports the most innovative industry use cases.



Optimize for efficiency, scalability, and security focus

This edge platform offers a fundamentally different infrastructure and operational approach to help organizations redefine their edge operations with optimized efficiency, consistent scalability, and increased security focus.

It provides edge-optimized computing, networking, storage, and security capabilities, while offering added efficiency and scalability through software-defined systems, zero-touch deployment, curated blueprints, automated orchestration, and centralized management with end-to-end observability and real-time analytics.

These fleet management capabilities also support the security focus needed for modern edge operations. This is reinforced by an embedded, multilayered, zero trust security approach that safeguards infrastructure, applications, models, and data from device to datacenter, and a wide range of key security capabilities, including:

- Physical security features (locking bezel, intrusion detection, and more).
- Digital protection features (Trust Anchor module, signed firmware, and more).
- Policy-based templates that offer consistent security profiles and automated policy enforcement.
- Advanced networking security features, including threat defense and network segmentation provided by Virtual Network Function (VNF) capabilities.
- A Software-as-a-Service (SaaS) platform that allows only verified and authenticated users to gain access, safeguarding data and minimizing unauthorized entry.

Modernize on a consistent, reliable, and standardized foundation

The Cisco Unified Edge solution powered by the advanced compute platform from Red Hat and AI Inference Server provides organizations with a consistent and reliable foundation for their complete modernization journey—starting with VMs or containers before growing into modern AI workloads.

Its flexible, modular, energy-efficient, and simplified design can be readily adapted to meet the needs of future workloads—allowing organizations to onboard new functionality without needing to rip and replace their existing workloads.

It also offers interchangeable support for compute and network nodes, a choice of Operating Systems (OSes) deployed in clustered or nonclustered environments, and an extensive joint ecosystem from Cisco and Red Hat that delivers full-stack integration and interoperability for maximum flexibility and choice.

Simplify deployment, management, and lifecycle operations

The convergence of IT and OT technologies and skill sets being seen across industries has created a need for edge solutions that offer simplified deployment, management, and lifecycle operations.

Cisco Unified Edge solution powered by the advanced compute platform from Red Hat offers:

- Simplified hardware setup to support on-site staff with specialized OT skills in industries where remote IT management is the only option.
- Zero-touch infrastructure and application deployment that eliminates the need for highly skilled IT staff to be present at edge locations.
- Prevalidated, full-stack blueprints, including policy-based templates, hardware, OS, and cluster blueprints, to support automated deployment of consistent, repeatable configurations across distributed edge infrastructure.
- A common set of tools, capabilities, terms, and ways of working that break down the barriers between cross-functional units and allow different professions to “speak the same language”.

This joint solution also helps simplify management and lifecycle operations with:

- Centralized management at edge scale that supports automated global deployment of policies and settings to streamline management across locations.
- Global server fleet visualization and cross-domain context visibility for both server and network teams to allow them to deploy, monitor, and manage the complete edge infrastructure effectively and efficiently across locations.
- End-to-end observability with real-time analytics that support rapid error detection and correction across globally distributed edge infrastructure.
- A modular design that simplifies service and upgrades, with chassis power and cooling infrastructure that can be reused across multiple generations of compute, networking, and GPU technologies.

Accelerate adoption and integration of AI-powered workloads

AI-powered applications have quickly become key to the future of industries who rely on edge deployments to collect, analyze, and build value from their operational data.

The Cisco Unified Edge solution powered by the advanced compute platform from Red Hat and AI Inference Server helps organizations develop, deploy, scale, and operate AI applications across the hybrid edge efficiently, effectively, and with a focus on security.

It provides a full-stack, AI-ready system with integrated compute, networking, storage, and security, that supports complex AI-powered workloads at the edge with:

- Cisco Validated Designs built specifically for rapid, low-risk deployment and agile, uninterrupted AI operations.
- AI Inference Server streamlining model inference across environments, including remote edge locations, to maximize throughput and minimize latency.
- The flexibility to deploy across any hardware accelerators and in any IT environments across the hybrid cloud or hybrid edge.
- Optimized gen AI models that reduce compute requirements to support more efficient and cost-effective AI deployments.
- An extensive, joint partner ecosystem that offers a comprehensive range of tested and validated AI hardware and software solutions.



Try AI Inference Server

Explore these interactive demos:

- [Introduction to Red Hat AI Inference Server](#)
- [Benchmarking Gen AI Inference: The Business Impact of Performance Optimization](#)

Or [start your no-cost trial](#) to further discover the value of AI Inference server.

Support innovative, industry-specific use cases

The Cisco Unified Edge solution powered by the advanced compute platform from Red Hat also offers validated edge solutions tested and certified for vertical-specific use cases. This allows IT teams to reliably and confidently deploy solutions that are tailored to meet the unique challenges of their specific industries, including but not limited to:

- **Manufacturing and industrial IoT.** Real-time monitoring, machine control, predictive maintenance, and quality assurance helps industrial organizations use their operational data to boost efficiency and reduce downtime.
- **Retail.** In-store analytics, personalized customer experiences, and improved inventory management allows retailers to process data quickly and respond to customer needs in real time.
- **Healthcare.** Remote patient monitoring, telemedicine, and real-time analytics of medical data helps healthcare organizations improve patient care and promote better healthcare outcomes.
- **Financial.** Fraud detection, branch office experience delivery, and security powered by modern edge operations help financial services organizations strengthen their brand and mitigate the many risks of their industry.

Start optimizing your edge operations with Red Hat and Cisco

Learn more about [Cisco Unified Edge, the advanced compute platform from Red Hat, and AI Inference Server](#).

Explore [this CVD](#) for a full-stack solution reference architecture based on Cisco Unified Edge with the advanced compute platform from Red Hat and AI Inference Server.

Contact your Cisco, Red Hat, or partner sales consultant to discuss how the full-stack solutions offered by Red Hat and Cisco can help your organization.