



White Paper

Modernizing the Distributed Frontier with VMware Cloud Foundation Edge

Sponsored by: VMware by Broadcom

Dave McCarthy

May 2026

IDC OPINION

The traditional centralized datacenter model is rapidly evolving. Organizations are increasingly mandated to adopt a cloud operating model that extends well beyond the core datacenter and directly into operational environments. This transition to a distributed private cloud is driven by a non-negotiable requirement: Enterprise infrastructure must now deliver the elastic agility of the public cloud while maintaining the absolute control, data sovereignty, and low-latency performance characteristics of localized infrastructure.

For IT and operational technology (OT) architects, this represents a fundamental architectural shift. The edge is no longer a collection of isolated compute islands; it must become a seamlessly integrated extension of the enterprise cloud capable of supporting advanced workloads at the point of data creation.

The adoption of a distributed private cloud architecture fundamentally restructures enterprise IT economics and operational models, delivering three critical outcomes:

- **Aggressive reduction in TCO via "lean edge" topologies:** Historically, deploying infrastructure to remote sites required substantial hardware investments. Modern distributed private clouds utilize lean edge footprints, scaling down to as small as a single host. This dramatically minimizes both capital expenditure (hardware spend) and operational overhead (power, cooling, and physical space) across hundreds or thousands of remote locations.
- **Operational consistency at scale:** The industry is moving away from manual management. Managing a fleet of geographically distributed edge locations requires a shift to automated, GitOps-driven desired state deployments. By adopting pull-mode operations, the responsibility for maintaining synchronization shifts to the edge itself; this ensures that the local environment autonomously converges with the

desired state, even if temporary network disconnections occur. The decentralized approach enforces configuration consistency, mitigates configuration drift, and significantly accelerates deployment times across complex IT/OT environments.

- **Future proofing for next-generation workloads:** As compute requirements evolve, the edge must act as an enabler for innovation. Establishing a robust distributed private cloud (complete with native support for GPU-accelerated workloads) is a prerequisite for preparing infrastructure to ingest and process the next wave of AI-driven applications and complex, agentic workflows directly where the data is generated, supporting both VM-based and containerized applications on the same platform.

Navigating the complexities of distributed infrastructure requires a holistic strategy. VMware Cloud Foundation (VCF) Edge addresses this by providing a unified platform explicitly engineered to manage fleets scaling to thousands of edge nodes.

SITUATION OVERVIEW

Enterprise infrastructure is undergoing a massive decentralization. While the centralized public cloud remains a powerhouse for data aggregation, training, and heavy compute, the center of gravity for real-time execution is shifting rapidly to the edge.

Market data points to an aggressive and sustained transition. According to IDC's *Future Enterprise Resiliency and Spending Survey* (December 2025), 26.2% of organizations plan to increase spending to expand infrastructure in edge locations in 2026. This reflects a broader macroeconomic trend; IDC's *Worldwide Edge Spending Guide* (February 2026) estimates that the total market for edge hardware, software, and services will hit \$449 billion by 2029, representing a robust five-year CAGR of 14.6%.

For IT and OT architects, as well as the C-suite, edge computing has been elevated from a localized initiative to a mandatory architectural evolution required to unlock the next phase of operational agility.

Beyond centralization: The physics and economics of the edge

The shift toward edge computing is driven by the immutable physical limitations of networks and the financial realities of data transport. Next-generation, real-time applications cannot tolerate the round-trip latency of traversing a wide area network to reach a regional cloud datacenter. Workloads such as high-speed manufacturing robotics, automated quality control, and instantaneous fraud detection require sub-millisecond response times that only localized compute can guarantee.

Additionally, the volume of telemetry that modern OT environments generate is staggering. Backhauling massive, continuous data streams to a centralized cloud for processing incurs

prohibitive bandwidth and cloud egress costs. Processing data locally and only transmitting the refined insights or anomalies fundamentally improves the economic model of enterprise data management.

Bridging the management gap at "zero IT" sites

Deploying enterprise-grade infrastructure outside the datacenter introduces a critical operational hurdle: surviving the zero-IT footprint. Most edge locations, from retail storefronts to remote energy grids, lack on-site technical staff. To scale effectively, edge infrastructure cannot rely on traditional break/fix models. It must be engineered for complete remote autonomy, utilizing zero-touch provisioning (ZTP), nondisruptive patching of remote edge hosts, and seamless fleetwide life-cycle updates that require absolutely no local human intervention.

The AI frontier: Inference and agentic workflows

The edge is rapidly establishing itself as the primary execution venue for the next generation of AI, moving the compute closer to the point of data creation. The architectural focus is actively migrating from resource-intensive AI training in the core cloud to optimized AI inference at the edge. Processing distilled large language models (LLMs) and multimodal models locally ensures immediate insight generation, reduces latency, and removes the dependency on continuous cloud connectivity.

Simultaneously, the industry is entering the era of agentic workflows. The edge provides the necessary low-latency execution environment for autonomous AI agents that require immediate, unimpeded access to local sensors and physical actuators to make real-time, closed-loop decisions.

Navigating data sovereignty, privacy, and isolation

Beyond performance and economics, localized infrastructure solves complex compliance and security challenges that the public cloud cannot address alone. As global data privacy regulations become increasingly fragmented and stringent, the edge enables strict compliance by design. It ensures that sensitive data, whether intellectual property or personally identifiable information, never leaves a specific physical site, geographic border, or legal jurisdiction.

Organizations also increasingly require highly secure and physically and logically isolated infrastructure footprints. This concept of running independent "mini clouds" within a broader facility allows for compartmentalization, often referred to as "pristine environments." A prime example is operating a heavily regulated, high-security pharmacy environment safely within the shared physical and network space of a standard grocery retail store, ensuring absolute isolation of sensitive healthcare operations.

INDUSTRY USE CASES AND VERTICAL INSIGHTS

The deployment of edge infrastructure manifests differently across industries, addressing highly specific operational constraints and business outcomes. Analyzing these vertical use cases highlights the transformative nature of localized compute.

Retail

In the retail sector, the primary challenge is managing unprecedented scale, often encompassing continuous deployments across thousands of distributed storefronts. Edge infrastructure ensures consistent delivery of advanced, high-bandwidth applications, such as computer vision for loss and fraud prevention, real-time inventory tracking, and highly personalized in-store shopping experiences. By leveraging automated, no-touch deployment models, retailers significantly reduce opex while bolstering site-level resiliency against network outages.

Manufacturing

The manufacturing floor is undergoing a profound hardware consolidation. Unified, software-defined infrastructure is aggressively replacing traditional, siloed industrial PCs and programmable logic controllers (PLCs). This transition demands absolute deterministic performance and availability, requiring infrastructure capable of consistently achieving 2–3 millisecond PLC cycle times across fully redundant data paths. This software-driven approach dramatically improves operational efficiency, reducing critical system upgrade windows and minimizing costly production downtime.

Logistics and transportation

For logistics and transportation networks, the edge is critical for maintaining real-time operational oversight in highly distributed, frequently mobile environments. Localized compute facilitates real-time fleet monitoring to ensure stringent safety compliance and driver security. Simultaneously, it processes massive data sets locally to optimize fleet routing operations and enables continuous, real-time inventory tracking across complex global supply chains.

Energy

The energy sector relies on edge computing to modernize and secure decentralized critical infrastructure. Virtualized protection and automation functions are executed locally at substations to ensure immediate grid stability and rapid fault response. Additionally, edge infrastructure provides the low-latency processing necessary to power advanced safety initiatives, such as deploying augmented and virtual reality for the remote operation and maintenance of equipment in hazardous or geographically isolated locations.

Government and defense

In government and defense applications, edge infrastructure must operate flawlessly under the most extreme, uncompromising conditions. These mission-critical environments frequently require architecture capable of supporting fully air-gapped, disconnected operational scenarios. In these theaters, continuous cloud connectivity is impossible, yet localized compute, AI-driven intelligence, and secure data processing remain critical to mission success.

VMWARE CLOUD FOUNDATION EDGE

As data gravity continues to shift away from centralized datacenters, organizations face the complex challenge of deploying, managing, and securing infrastructure in highly distributed, often constrained environments. VCF Edge addresses this imperative by extending enterprise-grade infrastructure directly to the point of data generation. Designed for environments ranging from retail footprints and branch offices to highly ruggedized or disconnected locations, such as hospitals, oil rigs, and defense sites, VCF Edge provides a unified, highly scalable architecture.

VCF Edge represents a strategic shift from siloed edge deployments to a centralized, fleet-managed paradigm that balances robust workload performance with aggressive TCO optimization. By centralizing the management control plane, organizations can achieve significant cost savings by eliminating the requirement for dedicated hardware to host the management stack at each individual edge site. This streamlined architecture ensures that local resources remain dedicated to high-value workloads while maintaining the operational simplicity of a unified control plane.

Feature highlights

VCF Edge is engineered to balance high availability with a minimized hardware footprint. It allows architects to standardize operations while supporting diverse hardware profiles.

- **Lean infrastructure:** VCF Edge supports one, two, and three+ host clusters. It offers flexible storage options, including vSAN for software-defined resilience and VMFS for traditional storage integration. This allows for a "right sized" approach where storage performance can be tuned to specific site requirements without changing the management model.
- **Mixed vendor support:** Recognizing that edge hardware is rarely uniform due to regional supply chain variances, VCF Edge allows for clusters comprising multiple server models within a single image. This significantly simplifies supply chain constraints and hardware refresh cycles because it decouples the software life cycle from specific OEM hardware revisions as well as ODM self-certification.

- **Observability and monitoring:** A centralized operations interface provides the ability to manage and monitor a globally distributed fleet of edge devices that provides simplified provisioning, operations, and security management with life-cycle management of hundreds or thousands of remote distributed edge devices.
- **Integrated services:** The platform integrates critical north-south services (DHCP, NAT) and high-performance switching specifically tuned for OT traffic. This is critical for industrial environments where deterministic performance for Modbus, PROFINET, or EtherNet/IP is required alongside standard IT traffic.
- **Edge AI readiness:** VCF Edge helps to position organizations for the next wave of innovation through a dedicated AI stack across different AI use cases such as computer vision and generative AI:
 - **vGPU sharing:** Democratizing and lowering the cost of GPU access by allowing multiple VMs or containers to share a single physical GPU is essential for localized computer vision and real-time inference.
 - **Model gallery:** This is a centralized repository providing consistent model packaging for reliable deployment across the fleet, ensuring that only the approved or validated model is used within the organization for a consistent user experience and maintaining corporate governance.
 - **Automated virtual large language model (vLLM) deployment:** This accelerates the rollout of LLMs at the edge utilizing CPU- and GPU-based AI inference engines. This enables support for local generative AI and natural language processing needs without backhauling data to the cloud.
 - **Private AI services at the edge:** VCF Edge includes Private AI Services to ensure privacy and control of their data, providing choice of open source and commercial AI solutions while maintaining privacy and compliance. VCF Private AI Services are designed for enterprises to develop, deploy, and manage AI applications — entirely at the edge.

Operational and scaling enhancements for VCF Edge

Scaling from 10 sites to 5,000 requires advanced capabilities in life-cycle management (LCM) and resource utilization.

- **Massive fleet scale:** VCF Edge supports up to 10,000 hosts and can execute 256 parallel LCM upgrades. This concurrency is a force multiplier, effectively removing the bottleneck of sequential maintenance windows that previously prevented massive edge rollouts.
- **Non-disruptive upgrades:** The introduction of Live Patch for ESX allows for host security updates and bug fixes without reboots. In single-host edge sites, where a reboot equals a total site outage, this is a transformative capability for maintaining 24 x 7 uptime and adhering to strict service level agreements.

- **NVMe memory tiering:** By utilizing lower-cost NVMe disks as a memory tier, VCF Edge reduces the need for expensive physical RAM. This tiering allows for higher consolidation ratios at a lower price point. For edge use cases where cost sensitivity is high, reducing the hardware bill of materials per site dramatically drives down the total cost of ownership.
- **Multi-NIC network isolation:** Multi-NIC support enables hardware-level traffic isolation between IT and OT paths. This ensures that a breach in the IT environment cannot move laterally to the OT controller network, providing a zero trust posture for modern containerized workloads.
- **Self-service edge deployments:** Desired state management and automated zero-touch provisioning simplify provisioning and provide intelligent resource placement for automated workload balancing across the distributed edge fleet.

Topology and deployment options

A primary challenge for infrastructure architects is sizing edge environments without over-provisioning hardware. VCF Edge introduces highly adaptable topologies tailored to specific site constraints. Organizations can select from distinct deployment topologies based on workload criticality and site requirements:

- **Single host/VCF instance:** This represents the lowest footprint option for simple VM and container workloads. These clusters integrate into the central management domain, drastically minimizing local management overhead.
- **Thin edge (two-host):** Ideal for critical applications requiring VMware High Availability (HA) without an expansive hardware footprint, this topology leverages remote or local witness capabilities to ensure resilient, WAN-independent shared storage.
- **Low-footprint containers:** For maximum scalability per vCenter, this non-Kubernetes option runs native container applications directly via vSphere Pods, optimizing resource consumption.
- **Full VCF instance:** This topology delivers the comprehensive VMware stack, including NSX virtual switching, designed for sites requiring complex networking, micro-segmentation, and advanced workload isolation.
- **Air-gapped/disconnected sites:** Purpose-built for defense or remote industrial sites, these provide independent workload domains with localized vCenters to guarantee self-contained operations regardless of WAN connectivity.

Fleet-level operations and life-cycle management

Operationalizing the edge at scale requires a departure from traditional, site-by-site management. VCF Edge introduces sophisticated fleet management capabilities designed to support massive distributed deployments:

- **GitOps workflow:** Utilizing Argo CD and VCF Edge enables declarative, desired state management across VMs, Kubernetes clusters, and native containers. This self-service deployment model ensures intelligent resource placement and automated workload balancing across the entire edge fleet, backed by centralized observability.
- **Automation:** Day-0 automation scripts streamline the deployment of the VMware vSphere Kubernetes Service (VKS), Harbor registries, and Content Library synchronization.
- **Fleet scale:** VCF Edge provides centralized observability and troubleshooting for thousands of sites across multiple regions.
- **Zero-touch provisioning (ZTP):** Organizations can execute remote imaging of ESX hosts using standard protocols (UEFI, HTTP/S).

Strategic business value

For executive leadership, the adoption of VCF Edge transitions edge computing from a fragmented operational burden into a strategic enabler, delivering five core business outcomes:

- **Optimized TCO and a consistent operating model:** Through NVMe memory tiering, mixed-vendor hardware support, and right-sized topologies, capital expenditure is strictly controlled, while centralized life-cycle management reduces the task of maintaining a global fleet of distributed edge devices.
- **Operational efficiency at scale:** GitOps workflows, ZTP, and the ability to manage 5,000+ hosts from a single control plane drastically reduce opex and truck rolls.
- **Enhanced security and compliance:** Air-gapped capabilities, secure boot, confidential computing, multi-NIC IT/OT isolation, and centralized desired state management are designed to ensure that remote sites adhere to strict corporate security and compliance mandates.
- **Business resiliency:** Features such as Live Patch and WAN-independent, thin-edge VMware HA enable critical applications to remain online.
- **Agility for AI and innovation:** Built-in AI readiness prepares the enterprise infrastructure to support future data science initiatives, bringing intelligent processing directly to the source of data creation.

VCF Edge effectively bridges the gap between the datacenter and the real world, providing the robust, automated, and scalable infrastructure necessary for the modern distributed enterprise.

CHALLENGES AND OPPORTUNITIES

As organizations look to implement these advanced edge architectures, they must navigate a landscape where technology intersects with deeply ingrained operational cultures.

A significant hurdle in edge deployments is often cultural rather than technical. Field operations teams work with a strict OT mindset, prioritizing absolute stability, determinism, and safety. Consequently, there is often a natural reluctance to change existing systems, particularly when introducing platforms from a vendor historically associated with centralized IT datacenters. Overcoming the perception that VCF is strictly an "enterprise IT" solution requires demonstrating proven reliability and an understanding of site-specific operational constraints.

While these perception challenges exist, they are eclipsed by the strategic necessity of edge modernization. The robust feature set of VCF Edge, specifically its workload isolation, lean footprint capabilities, and fleet management, creates a clear pathway to modernize legacy OT environments. Bridging this IT/OT divide with a unified, enterprise-grade infrastructure platform has become a fundamental prerequisite for organizations looking to adopt and scale the next wave of high-value, AI-driven capabilities at the edge.

CONCLUSION

The transition from centralized datacenters to a distributed cloud model has emerged as a critical operational necessity. As organizations push compute power to the physical frontier to overcome the latency challenges and bandwidth economics, the definition of the private cloud has fundamentally expanded. For IT and OT infrastructure teams, the mandate is clear: To survive the scale and complexity of managing thousands of locations, the traditional approach to managing edge environments must change.

VCF Edge provides an architectural answer to this distributed challenge. By delivering technical parity across the entire fleet, VCF Edge allows enterprises to deploy lean infrastructure tailored to site-specific needs, accommodate mixed vendor hardware seamlessly, and enforce integrated networking and security services. This software-defined approach transforms thousands of isolated sites into a cohesive, secure, and highly available private cloud fleet.

Looking ahead, IDC believes that AI inference and autonomous agents operating directly at the edge will define the next frontier of competitive advantage. However, these agentic

workflows cannot function without a robust, performant, and standardized infrastructure layer. With native edge AI readiness, VCF Edge is designed for organizations that are actively deploying the distributed private cloud infrastructure necessary for successful, large-scale AI workloads.

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

One Beacon Street
Suite 33100
Boston, MA 02108
USA
508.872.8200
X: @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 2026 IDC. Reproduction without written permission is completely forbidden.