IDC

White Paper, sponsored by:

Broadcom

Self-Service Private Cloud as the Engine for In-House AI and Application Innovation

WRITTEN BY:

Matthew Flug
Research Manager

Adam Reeves
Research Director

PUBLISHED:
August 2026

Table of Contents

Introduction

Artificial intelligence has become the default context for new application development, not a specialized use case within it. IDC's research has found that organizations building or planning AI applications now represent the overwhelming majority, and that intelligent applications and AI agents now top the list of organizations' priorities for their application platforms, a position that is expected to hold and strengthen in the future. Capabilities that support AI development — including model libraries, inferencing, and AI frameworks — have become the most-cited functionality organizations weigh when selecting a platform, which indicates AI readiness has moved from differentiator to baseline requirement. AI applications are not one workload competing for infrastructure alongside the rest. They are becoming the primary driver of new infrastructure demand.

That demand is not landing exclusively, or even primarily, on public cloud. IDC's research shows that AI workloads running on-premises or in private cloud now approach parity with those running in public cloud. Security requirements, data privacy obligations, and the sensitivity of the proprietary data feeding these models are keeping a substantial share of AI infrastructure inside the datacenter, even as overall AI adoption accelerates. For platform engineers, this means private cloud is no longer a legacy environment to modernize away from. It is where a large share of the most strategically important workloads in the enterprise are already running, which raises the bar for what that platform needs to deliver.

A significant obstacle to shipping AI applications inside private datacenter environments is manual, ticket-driven provisioning. A request for a virtual machine (VM), a namespace, or a database often still moves through an approval queue, and each request lengthens the path from an idea to a production deployment, slowing the software development life cycle (SDLC).

Platform engineers carry the burden of resolving that friction. Developers need speed. They do not want to focus on security, compliance, or cost controls. Developers are better served by having those policies embedded in the infrastructure they consume. IT organizations need consistently enforced security, compliance, and cost control while maintaining resiliency. Self-service infrastructure can be the mechanism that satisfies both requirements at once, and private cloud is increasingly the choice for organizations to deliver it, given the security, compliance, and data sovereignty requirements associated with AI workloads.

Developer expectations have moved past what ticket queues can support. Developers expect instant access to runtime environments, automated deployment pipelines, managed databases, auto-scaling compute, and integrated monitoring and logging, alongside newer requirements such as Model Context Protocol (MCP) servers and gateways that connect AI-powered applications and agents to existing tools. Meeting that expectation within the private datacenter has become a baseline requirement for keeping application development in-house, enabling organizations to maintain control over their proprietary data.

This paper gives platform engineers the data needed to secure executive buy-in for a self-service private cloud built on VMware Cloud Foundation (VCF) Automation, Broadcom's automation and consumption layer for VCF. The argument is organized around the three outcomes that carry the most weight in that conversation: speed, security, and AI readiness.

87%	100%	81%
of application platforms are meeting or exceeding ROI expectations.	of organizations are building or planning generative AI and/or agentic AI applications.	of developers report productivity gains from agentic AI.

Sources: IDC's *Application Deployment Platforms Decision-Maker and Business Value Survey*, December 2025; IDC's *Agentic Application Development and DevOps Survey*, December 2025; *The Impact of AI on PaaS*, April 2026

On average, 49.9% of organizations' AI workloads run on-premises or in private cloud, nearly matching the share running in public cloud.

Source: IDC's *Application Deployment Platforms Decision-Maker and Business Value Survey*, December 2025

Situation Overview

Private cloud is where production lives

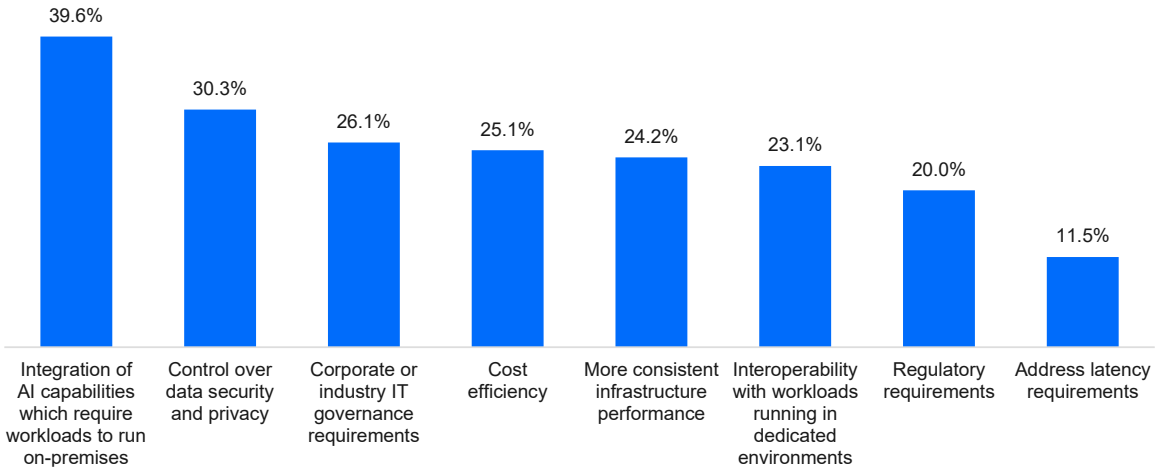
Non-public cloud deployments are growing at a CAGR of 7.3% (see IDC's *Worldwide Cloud Application Deployment Platforms Software Forecast, 2026-2030*, July 2026), and workload repatriation — the movement of workloads from public cloud back to on-premises infrastructure — is accelerating. Organizations moved 17.6% of workloads from public cloud platforms back to on-premises environments over the past 12 months, and nearly 30% of organizations repatriated more than a quarter of their total workloads (see IDC's *PaaS Decision-Maker and Business Value Survey — Cloud and PaaS Adoption*, March 2026).

As shown in Figure 1, the integration of AI capabilities is the leading driver of repatriation (see IDC's *2025 Future Enterprise Resiliency & Spending Survey*, Wave 9). Security, data privacy, cost efficiency, and governance also contribute significantly to the shift toward private cloud. These four factors recur throughout IDC's research as drivers of an organization's choice to run workloads on infrastructure it controls directly.

Figure 1

Top factors driving application repatriation

Q: You indicated your organization is prioritizing the migration of applications from public cloud to on-premises infrastructure in 2026. What are the top two reasons for this migration?



Source: IDC's *Future Enterprise Resiliency and Spending Survey* Wave 9, November 2025

AI workloads intensify this pattern rather than reversing it. Roughly half of AI workloads already run on private cloud or on-premises infrastructure (see IDC's *Application Deployment Platforms Decision-Maker and Business Value Survey*, December 2025). As AI initiatives move from pilot projects to production systems, data gravity and governance requirements pull those workloads toward infrastructure the organization already controls, rather than toward the public cloud capacity many of the same organizations expanded during the previous decade.

Platform engineering and the self-service imperative

Platform teams have shifted from maintaining infrastructure to operating an internal product for developers, and increasingly for data scientists and line-of-business users. Nearly 70% of organizations now purchase rather than build their platforms (see IDC's Application *Deployment Platforms Decision-Maker and Business Value Survey*, December 2025), indicating that the differentiated value of platform strategy has evolved from a focus solely on the underlying plumbing to include platform operations and governance.

Ticket-driven provisioning limits how far a platform team can scale, because every resource request becomes a queue, and each queue requires a person to clear it manually. Self-service infrastructure removes that constraint by giving developers compliant, governed, on-demand access to the resources they need, allowing platform teams to scale their impact without adding headcount in proportion to increased demand.

The goal of an application platform has always been to enable developers via compliant, governed self-service. AI adoption is the force that makes it a necessity. All organizations are building or planning AI applications, and intelligent applications and AI agents rank as the top two platform priorities today and over the next 24 months (see IDC's *Application Deployment Platforms Decision-Maker and Business Value Survey*, December 2025). A platform strategy built around conventional application workloads does not meet those

priorities, and closing that gap requires access to infrastructure and resources that developers can provision without waiting on a ticket.

IDC Opinion

What developers need from a platform

Developers influence 47% of organizational spend on cloud adoption and vendor selection (see IDC's *Which Technology Purchasing Decisions Do Developers Influence Most?*, IDC #US53186125, February 2025). That concentration of influence gives developers' platform preferences *significant* weight in an organization's technology direction, independent of who holds formal purchasing authority. To earn that preference, application platforms must serve and enable developers first.

As shown in Figure 2, security is the top consideration organizations weigh when selecting a platform to support developers, followed closely by the ability to build AI apps, application performance, alignment with cloud strategy, delivery of tools to make developers more successful, cost, and ease of use (see IDC's *Application Deployment Platforms Decision-Maker and Business Value Survey*, December 2025). The near parity between security and innovation-related capabilities indicates that organizations no longer treat platform security and platform capability as separate purchasing decisions weighed against each other. A platform now has to deliver both in order to be considered.

Figure 2

Organizational selection criteria to support developers

Q: How to support developers — what matters most when choosing an application platform?

Rank your organization's top 3 criteria when selecting an application platform	Rank up to three capabilities/ functionalities that are key to choosing an application platform
→ Security — 30%	→ Capabilities for building AI apps — 30%
→ Application performance — 29%	→ Security capabilities — 27%
→ Align with cloud strategy — 25%	→ Intelligent capabilities with which to build apps — 26%
→ Cost — 22%	→ Deployment across hybrid cloud — 24%
→ Ease of use — 21%	→ Available components for building apps — 23%

Source: IDC's *Application Deployment Platforms Decision-Maker and Business Value Survey*, December 2025

Intelligent applications and AI agents are the top two priorities for application platforms today, cited by 49% and 38% of organizations, respectively (see Table 1). However, while AI agents grow as a priority over the next two years, intelligent applications (while still a top priority) fail to grow over that span. That signals a shift from prioritizing software that "answers" to prioritizing software that "acts." As a result, application platforms must expand

to govern AI agents, which raises the bar on two fronts: security posture around what an agent is permitted to do, and integration with the tools and data it depends on to act. Both have to be rock-solid, because an agent taking the wrong action poses a materially different risk than an application returning the wrong answer.

Table 1

Top purposes for application platforms

Q: Which of the following are the top three priorities for your organization's application platform today and in 24 months? [Rank top 3; table shows overall results]

Platform priority	Today	In 24 months
Intelligent applications	49%	48%
AI agents	38%	47%
Mobile applications	36%	27%
Modernized legacy applications	33%	30%
Net new cloud native applications	32%	30%

Source: IDC's *Application Deployment Platforms Decision-Maker and Business Value Survey*, December 2025

Capabilities that support building AI applications, including model libraries, inferencing, and AI frameworks, are now the most-cited functionality organizations weigh when selecting a platform (see IDC's *Application Deployment Platforms Decision-Maker and Business Value Survey*, December 2025). Platform engineers evaluating vendors should treat AI-building capability as a baseline requirement rather than a differentiator.

Measuring what matters

IDC recommends that platform investment be measured against business and developer outcomes rather than infrastructure uptime alone. While important, uptime shows that a platform is running, not how well it is running. It does not demonstrate that the platform is changing how fast an organization ships and patches software, the levels of innovation, or how much that software is worth.

Table 2

Measuring what matters

Metric type	Importance	Example metrics	Impact
Business	Links technical performance to tangible outcomes such as revenue and retention, ensuring the platform delivers real value to the organization	• Time to market • Revenue attributable to applications • SDLC cost • Technical debt reduction	In total, 41% of organizations cite platform-enabled revenue as a top three driver of application platform adoption.
Technology	Links system behavior to measurable indicators such as uptime, latency, and error rates, ensuring the platform performs reliably at scale	• DORA metrics • Mean time to remediation • Vulnerabilities identified pre-production • Uptime • Latency • Error rates	Fifty-eight percent of organizations cite security as a top-three application platform selection criterion.
People	Links team engagement to measurable indicators such as adoption rates and time-to-first-commit, ensuring the platform is used and delivering value	• Adoption rates • Time from onboarding to first commit • Time to access resources	High-performing organizations are 10%–11% more likely than underperformers to track these metrics.

Sources: IDC's *Application Deployment Platforms Decision-Maker and Business Value Survey*, December 2025; *IDC Perspective: Quantifiable Metrics to Measure the Success of Application Platform Investments*, March 2025

Organizations that track people metrics (e.g., platform adoption) alongside business and technology metrics are the same groups that IDC research identifies as high performers, those getting the most from their platforms. Tracking of all three indicates that the discipline of actively driving adoption and refining the platform to keep developers satisfied, not just the initial platform investment, is what separates high performers from the rest of the field. Establish baselines before launch, measure continuously, and assign a named owner accountable for the platform's road map and outcomes.

VCF Automation: Self-Service for the Private Cloud

VMware Cloud Foundation: Unified cloud platform for AI

VMware Cloud Foundation (VCF) is a unified private cloud that provides compute, storage, and network services for consumption, with layered capabilities for AI development, automation, Kubernetes cluster management, and applications. Organizations adopt it in place of assembling a private cloud from separately managed virtualization, storage, and networking products, each with its own operational model and its own team responsible for keeping it running.

That unification is what Broadcom positions VCF as delivering. It gives an organization one platform capable of running traditional VM-based applications, containers, and AI and machine learning workloads side by side under a single operating model that holds whether the workload sits in a datacenter, at the edge, or on a VMware-supported public cloud endpoint. The company claims this combination is the foundation an organization needs to keep pace with public cloud providers in terms of agility and developer experience while retaining the control, security posture, and data locality that come with infrastructure that the organization owns and operates itself. VCF Automation, covered in the rest of this section, turns that foundation into something developers and platform engineers can actually consume on demand.

VCF Automation: Delivering self-service

VCF Automation is Broadcom's self-service automation and consumption layer within VCF, built for AI, Kubernetes, and VM-based application development. The platform gives infrastructure administrators three levers that shift their role from virtual infrastructure administrators who provision resources to cloud administrators who define the rules that enable provisioning without them. Instead of manually provisioning VMs for each request, administrators establish projects and namespaces mapped to specific teams or business units, enabling those teams to self-provision infrastructure resources within defined governance guardrails. They can set infrastructure policies (to help govern where workloads land) and approval and lease policies, as well as using policy-as-code to create policies that support Kubernetes management. They also delegate network configuration to application teams through virtual private clouds (VPCs) scoped to each tenant, removing networking from the admin's per-request workload. The result is an administrator who operates the foundation of a multi-tenant, self-service private cloud rather than clearing a queue of individual infrastructure tickets.

Broadcom describes a two-layer model built on top of that foundation. Cloud administrators establish the multi-tenant environment, and platform engineers then consume those capabilities within each tenant to build and operate an internal developer platform (IDP), curating service catalogs, publishing golden-path templates, and managing governance and compliance at the application-team level. This division of labor mirrors the shift identified in

IDC's research, in which platform teams increasingly function as internal product owners rather than infrastructure custodians.

Mechanically, VCF Automation runs as a clustered set of VMs directly on vSphere and consumes the vSphere Supervisor, which establishes a Kubernetes control plane directly on the hypervisor layer rather than as a separate system that administrators must integrate themselves. VCF Automation integrates with and manages the Supervisor to run VMs, containers, and Kubernetes-based workloads side by side. That control plane enables a single IaaS foundation to provision VMs (via VM Service), containers (via Container Service), and Kubernetes clusters (via VMware vSphere Kubernetes Service [VKS]), network configurations, and storage, instead of routing VM and container requests through separate tools. Application teams reach that foundation through the Modern Cloud Interface — a single UI, CLI, and API endpoint that exposes both a curated, pre-approved self-service catalog and a consumer-facing self-service IaaS model, with upfront cost estimates and workload performance data surfaced at the point of provisioning rather than after the resource is already running.

Broadcom reports that one oil and gas company reduced infrastructure provisioning time from 3–4 days to approximately 30 minutes after moving its catalog onto VCF Automation.

Source: Broadcom, 2026. The figure is vendor-reported and illustrates achievable improvement rather than an industry benchmark.

Key self-service capabilities

VCF Automation's self-service capabilities include the following components:

- **Modern cloud interface:** A unified UI, command-line interface (CLI), and API for on-demand provisioning, with built-in cost visibility, service-specific pricing, and chargeback
- **Private cloud services:** Out-of-the-box IaaS building blocks consumable via the IaaS Services Console, spanning core, extensible, and pre-bundled services such as data protection, encryption, and multi-tenant disaster recovery
- **Curated service catalog:** Reusable blueprints published directly to the catalog, including live application topologies captured via App Stack Formation
- **Embedded guardrails and policy controls:** A layered governance model spanning approval, lease, and day 2 action policies, infrastructure policies, policy-as-code, RBAC with tenant identity management, as well as per-tenant or per-business-unit isolation with chargeback visibility
- **Multi-tenancy:** Per-tenant or per-business-unit isolation with chargeback visibility
- **Private AI services:** VCF Private AI Services, delivering self-service GPU-enabled VMs and VKS clusters with model runtime, vector databases, data indexing, and an AI agent builder
- **MCP servers and gateways:** Connectivity between AI-powered applications and the tools already present in a developer's environment

- **Fast Deploy:** VM Fast Deploy reduces the time required to provision VMs, containers, and VKS clusters
- **Extensibility:** Native Terraform and Ansible integration with event-driven orchestration

The governance layer is the component that determines whether the rest of the catalog is safe to open to developers and agents. Approval and lease policies control resource life cycles, placement policies control locations, and policy as code and RBAC control access. Platform engineers evaluating any self-service private cloud, not VCF Automation specifically, should treat this layered governance model as the baseline with which other platforms are compared, since it addresses the security concern that IDC's survey data has identified as the top platform selection criterion.

How VCF Automation helps organizations enable developers

IDC views developer enablement as the practices and tooling that remove technology and business friction from software development, improving the experience of building software itself. With its focus on delivering developer value, developer enablement functions as a primary mechanism through which vendors differentiate, a consequence of the 47% of spending influence developers hold. Key elements of IDC's view of developer enablement include safe, smooth, innovative, and easy.

- **Safe:** Developer environments, platforms, and infrastructure that are intrinsically secured, with security, privacy, and governance built into access, execution control, and auditability. Broadcom states that VCF Automation delivers source-controlled blueprints and enforces security and compliance through embedded guardrails at provisioning time, continuously afterward through lease and day 2 action policies, and consistently across cluster fleets through centralized policy management.
- **Smooth:** Environments where commercial friction is minimized, with straightforward contracting and clear service-level agreements (SLAs). Broadcom positions VCF Automation's on-demand catalog as a way to eliminate multi-team approval chains that would otherwise separate a developer from the resources they need.
- **Innovative:** Platforms that expose developers to capabilities that support building AI-driven, intelligent applications. Broadcom points to VCF Private AI services and GPU optimization as support for intelligent application development inside the organization's own security perimeter, a claim consistent with the survey finding discussed above that AI-building capabilities are now a baseline platform expectation.
- **Easy:** Developers want to write and generate code in the languages and environments of their choice, without contending with infrastructure complexity. Broadcom describes VCF Automation's unified interface as an abstraction layer that keeps developers focused on code rather than provisioning.

The "safe" and "smooth" dimensions map most directly onto the governance and catalog capabilities described earlier in this section, while "innovative" and "easy" depend more heavily on how a given platform engineering team configures and curates the platform for its own developers. VCF Automation supplies the underlying mechanism; platform engineering teams determine how much of that mechanism developers actually experience as reduced friction.

Conclusion and Essential Guidance

Self-service private cloud requires an operating model change from platform teams that extends beyond a one-time infrastructure deployment. It is the mechanism that lets organizations build AI applications, modernize legacy systems, and ship software at the pace developers now expect, without asking IT to give up the governance for which it is accountable.

Executive conversations about platform investment should lead with outcomes that boards and finance organizations already track: time to market, developer productivity, security consistency, and AI readiness. Infrastructure specifications do not make that case on their own.

Implementation should proceed incrementally. Target high-value problems first, establish baselines before launch, and measure from day 1 using the business, technology, and people categories outlined above (see *IDC Perspective: Quantifiable Metrics to Measure the Success of Application Platform Investments*, March 2025).

Treat the platform as a product with a road map, a changelog, and a continuous feedback loop with the users it serves. VCF Automation supplies the underlying self-service and governance mechanism, and the discipline platform engineers apply on top of it determines whether that mechanism translates into the outcomes executives are asking to see.

About the IDC Analyst(s)

Matthew Flug
Research Manager, IDC

Matt Flug is a research manager within IDC's Cloud Application Deployment Platforms research practice. Matt's research focuses on software vendors, cloud providers, and end customers' software development/deployment plans and experiences. He is responsible for insights and analysis on emerging PaaS solutions that support developers deploying highly performant, modern, and cloud-native applications.

Adam Reeves
Research Director, IDC

Adam Reeves leads IDC's worldwide research on PaaS for Developers of Modern and Edge Applications. His analysis leverages demand-side research with market modeling expertise to identify the evolving needs of PaaS buyers as they implement the products and services needed to accelerate application development and deployment in the cloud, across multiple clouds, and at the edge. His focus on PaaS and edge applications is oriented around the foundational cloud services, supporting components/services used by developers, evolving public multicloud and cross-cloud PaaS use cases, and the growing use of cloud services at edge locations.

Message from the Sponsor

A modern cloud operating model

As organizations scale AI and intelligent applications, private cloud has emerged as the primary engine for innovation. Delivering developer autonomy without sacrificing enterprise governance requires shifting from ticket-driven IT to an automated, policy-based infrastructure. VMware Cloud Foundation (VCF) Automation delivers this modern cloud operating model, empowering cloud admins and platform engineering teams to operate a secure, multi-tenant, self-service private cloud. By offering a curated catalog and private cloud services, including VMs, containers, Kubernetes clusters, and AI services alongside embedded guardrails, VCF Automation drastically accelerates the software development life cycle. This unified approach reduces infrastructure provisioning times from days to minutes, directly optimizing developer velocity and maximizing your infrastructure ROI.

- Learn more about how your private cloud is the new public cloud and how to deliver private cloud services on your terms in our new eBook.
- Learn more about VCF Automation.

IDC Custom Solutions

IDC Research, Inc.
One Beacon Street, Suite 33100, Boston, MA 02108, USA
T +1 508 872 8200

blogs.idc.com | LinkedIn @IDC | www.idc.com

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