White Paper

Essential Elements for Private Cloud Strategies

Sponsored by: Broadcom

Jevin Jensen Natalya Yezhkova

August 2024

IDC OPINION

Looking back 20 years ago, most workloads and IT resources organizations had resided in their datacenters, with some limited hosting environments for off-premises workloads. The last 12 years have seen a rush to public cloud services, including infrastructure-as-a-service, platform-as-a-service, and software-as-a-service solutions. While the adoption of public cloud services was on the rise, it got an extra boost during the pandemic, as IT teams had to make split-second decisions and skip critical architecture design and planning steps in an attempt to quickly address the unexpected needs of changing work environments. The public cloud is here to stay, but the pendulum is moving toward equilibrium between on-premises and the public cloud. Today, many enterprises have adopted private cloud architecture. This natural evolution occurred as companies modernized their legacy applications with containers and web services. The rise of edge applications also drove companies to leverage cloud resources for local datacenters and in company-operated remote locations. Companies wanted the cloud's flexibility, scale, and on-demand agility for their on-premises applications and infrastructure. IDC's research shows that 49% of production workloads are still on-premises, and in two years, enterprises estimate that 47% will still run on-premises. This fact demonstrates that enterprises should carefully consider workload placement; some move workloads back on-premises.

Companies realized that the public cloud isn't an optimal environment for every workload for a variety of reasons, including costs, security, privacy, manageability, and other requirements and expectations. Companies must be agile as the pendulum returns to on-premises while maintaining the flexibility to handle an uncertain future. In fact, according to IDC's *March 2024 Cloud Pulse* survey, 29% of applications retired will be from the public cloud versus 28% from

the private cloud. Furthermore, new applications will be equally added to public cloud growth (35%) versus private cloud growth (34%). This equilibrium has now been reached as enterprises reconsider the strategy around how to leverage their private cloud investment effectively. Private cloud platforms now contain the popular features and capabilities initially found in public cloud offerings, showing a maturity level on par with public providers. These powerful on-premises capabilities give enterprises the flexibility of the public cloud with the predictable costs of the private cloud. More importantly, the cloud is the operating model for modern applications, and this model is now available on-premises, empowering enterprises with more granular control over how these applications operate.

As companies modernize their legacy applications into containers, the portability of modern applications increases and on-premises applications become more attractive. Fully optimizing the public cloud requires rearchitecting and disrupting all interdependencies for each application or business process. As many companies found out the hard way, a lift and shift to the public cloud can dramatically increase costs compared with on-premises IT environments. Also, some legacy workloads do not need to be migrated to the cloud or even to a new architecture for various reasons. Companies need to maintain and operate them efficiently. New workloads, such as AI-enabled applications and large language models, require a radical new approach to designing and running infrastructure to provide resilience and well-performing applications. Important data privacy and sovereignty considerations need to be considered as part of the governance of these new workloads. With containers and AI workloads effectively running in either the public cloud or on-premises private clouds, enterprises should evaluate what offers the best business outcome.

While there is no doubt that public cloud services deliver benefits, organizations should carefully consider private cloud when outlining their IT strategies. There are ample benefits to the private cloud for enterprises. These include increased control of security, improved performance for mission-critical applications, and more predictable ongoing costs. According to IDC's *Enterprise Infrastructure Survey, Workloads*, from November 2023, the top benefits of using private cloud are the availability of computing and storage resources, performance, ease of use and management, data security, scaling, and data protection. However, the challenges customers experience at the technical or business levels sometimes outweigh these benefits. It happens more often when decisions about workload migration into the public cloud aren't well analyzed and are rushed at the request of the C-suite, when the workload ecosystem has changed, or when external factors force companies to adjust the use of public cloud services. As a

result, companies choose to repatriate workloads back into dedicated environments.

WORKLOAD REQUIREMENTS DRIVE REPATRIATION

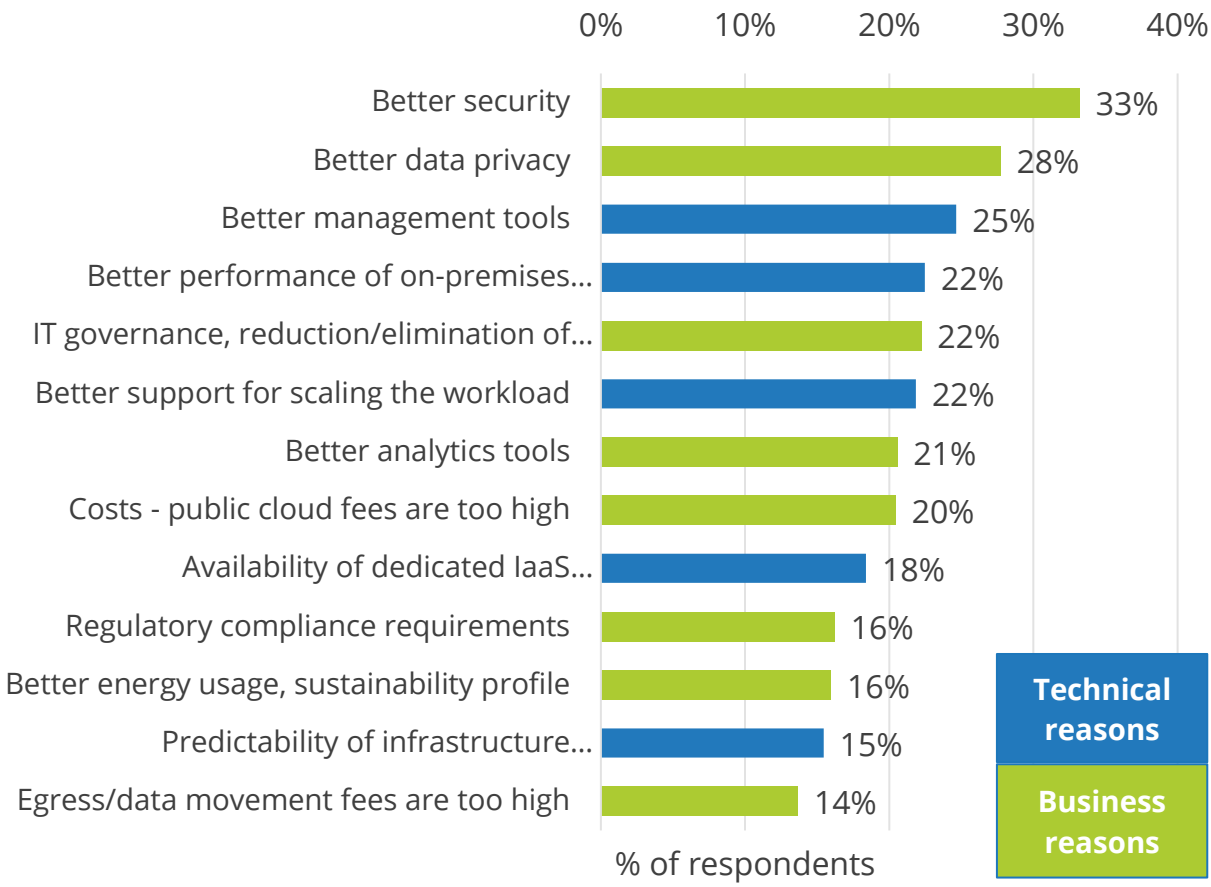
The concept of workload repatriation isn't new. Since the public cloud became a mainstream element of IT strategies, IDC has been following the phenomenon of repatriation, which hasn't slowed down significantly for years. IDC's *Server and Storage Workloads* survey of 2,250 IT professionals across the globe, completed in March 2024, reveals that 81% of respondents expected some level of repatriation of computing and 83% of storage resources in the next 12 months. There is a slight slowdown compared with the previous surveys, but it still indicates significant activity related to the continuous search for the optimal IT environments for workload placement. The same survey confirmed that the repatriation of an entire workload is a rarity, with an average of only 7% of respondents outlining plans for complete workload repatriation. Technical and AI-related workloads are more often than others considered for full repatriation, with 13% and 10% of respondents, respectively, expecting full repatriation.

Another survey, completed in November 2023 as part of IDC's U.S. Cloud Migration study, provides insights into the primary reasons for workload repatriation (Figure 1).

FIGURE 1

Top Reasons for Workloads Repatriation from Public Cloud into Dedicated IT Environments

Q. Select the top three reasons for repatriating workload(s) from public cloud to dedicated environment.



IDC, 2024. n = 783. Source: *Cloud Migration Data Collection Survey*, November 2023.

As this chart shows, while technical reasons such as management tools and performance play important roles, business reasons such as security and data privacy constitute the top reasons for workload repatriation, appealing to 33% and 28% of respondents to this survey. These reasons are even more critical for certain workloads related to sensitive corporate data, such as artificial intelligence and, specifically, generative AI (GenAI). The recent IDC *Future Enterprise Resilience and Spending Survey*, Wave 4, completed in April 2024,

showed that data security and privacy exposures pose the greatest risk of launching GenAI solutions for organizations' C-suites, with 22.6% of respondents worldwide selecting this as a top concern. In the same survey, half of the respondents indicated that in the next 18 months, they would consider one of the forms of dedicated environments (on-premises, hosted/colocated, or at the edge) as the primary approach to support the initial GenAI training, tuning, and inferencing, and 54%–56% will consider them as a secondary approach.

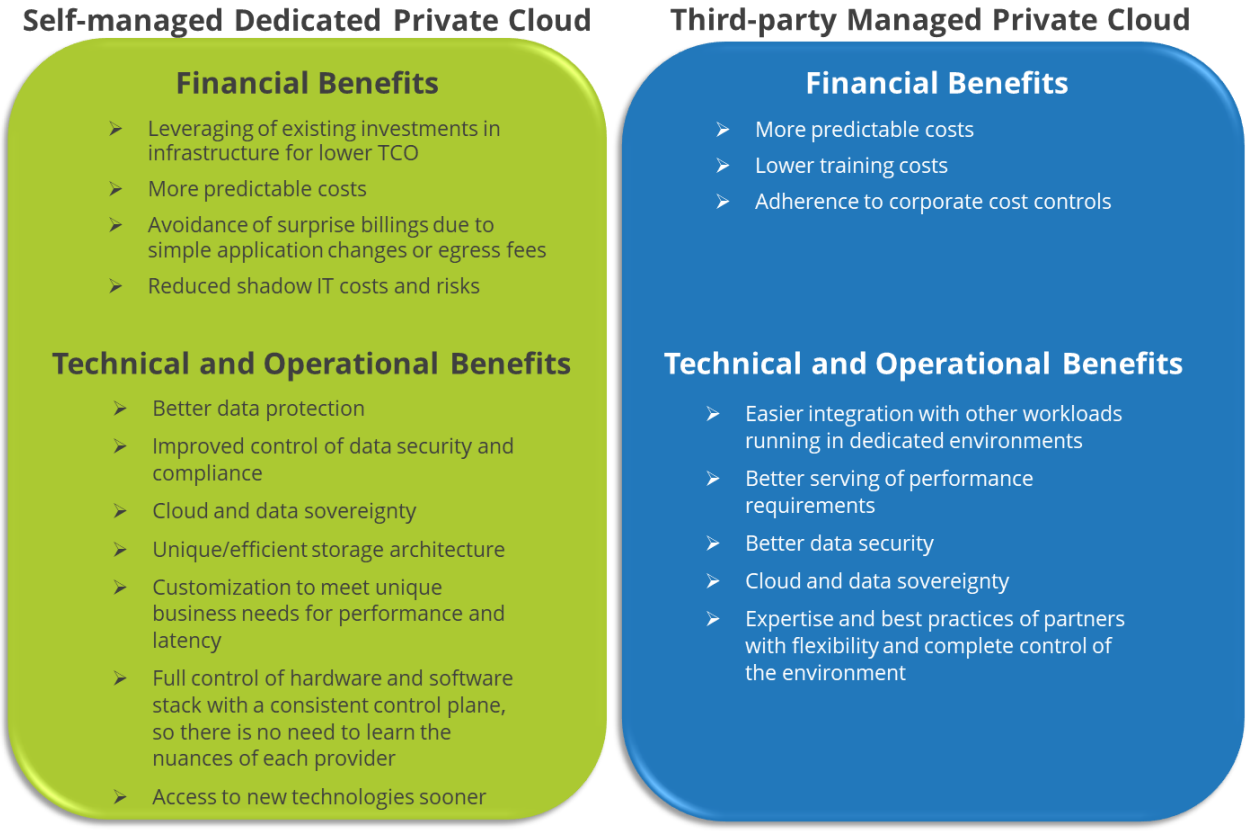
IDC's *Enterprise Infrastructure Pulse, Computing Infrastructure*, survey from April 2024 showed even greater interest in using dedicated environments for running AI workloads. Three-fifths (61%) of respondents indicated using on-premises private clouds for deploying their AI (not only GenAI) workloads. However, 36% use on-premises non-cloud, 30% use off-premises dedicated infrastructure, and 23% also use public cloud services.

TOTAL COST OF OWNERSHIP AND BENEFITS OF PRIVATE CLOUD

So, why do dedicated environments, especially dedicated or private clouds, play such an essential role in the market? IDC's *Enterprise Infrastructure Pulse, Workloads,* survey found that data privacy/protection was the number 1 factor in running a self-managed dedicated infrastructure, with better total costs of ownership (TCOs) closely following (*Enterprise Infrastructure Pulse, 4Q23: Enterprise Workloads — Executive Summary*, IDC #US51791224, March 2024). Unsurprisingly, lower TCO was reported as a financial benefit. IDC found 34% lower infrastructure costs when running a private cloud platform solution (*Business Value White Paper, sponsored by VMware by Broadcom*, #US52312224, August 2024). The survey and IDC's discussion with IT leaders suggest the following additional benefits of dedicated environments over the public cloud as seen in Figure 2.

FIGURE 2

Benefits of Dedicated Private Cloud Environments



Source IDC, 2024

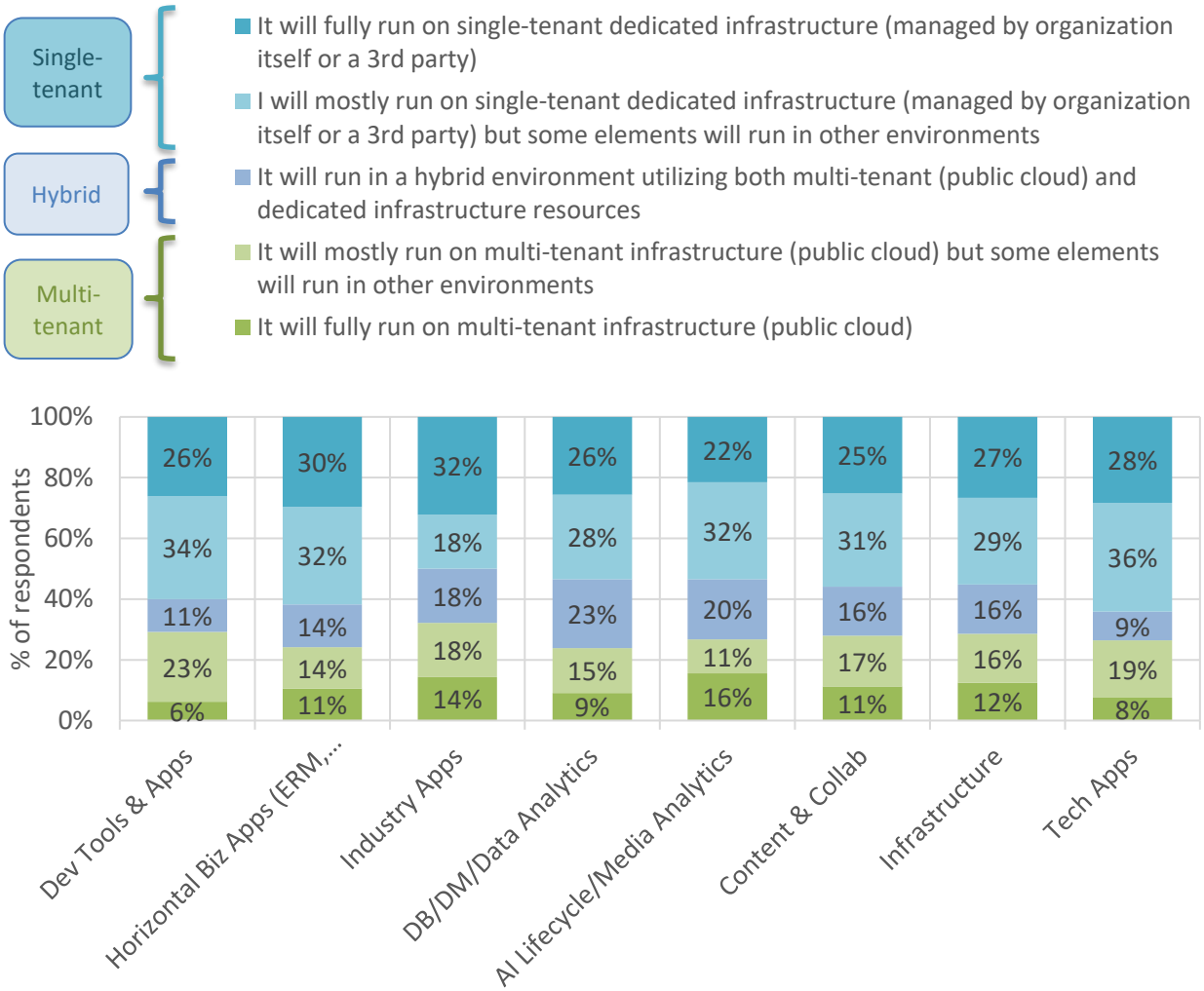
WHY ENTERPRISES ARE ADOPTING PRIVATE CLOUD STRATEGIES

These developments in the private cloud-enabled a larger transformation of the IT world into a hybrid. IDC's research shows that for most workloads, only 35%–40% of companies expect to run fully in public cloud or dedicated environments. In most cases, companies expect workloads to run across environments, leaning toward one or another — either public or dedicated (see Figure 3).

FIGURE 3

IT Environments in Which Companies Plan to Run Workloads in Two Years

Q. Which of the following statements best describes deployment models for the following workloads in two years?



IDC, 2024. n = 600. Source: IDC *Enterprise Infrastructure Pulse Survey*, December 2023.

This tendency is enabled by the increasing consistency of the infrastructure layer and management tools, which allow workloads to run anywhere — on-premises, at the edge, in colocation environments, or in the public cloud — without modification or changes in the processes or management practices. The enablement that the infrastructure and management layers provide is a fundamental part of the hybrid multicloud world. The other critical element that makes it possible is the application side. The application layer must be transformed and modernized for flexibility and movement between environments.

APPLICATION CONTAINERIZATION IMPROVES THE RESILIENCE OF PRIVATE CLOUD

Application containerization has seen increased adoption and is common in legacy application modernization strategies. By enabling DevOps teams to package applications and all dependencies into isolated environments, enterprises can achieve many business benefits, including improved consistency, portability, and efficiency. Containers ensure that applications run similarly across different environments and increase portability between servers or cloud environments.

According to IDC's *Enterprise Infrastructure Pulse, Workloads*, survey, the top 5 benefits of using containers for enterprise workloads are better security (31% of aggregate responses across all workloads running in containers), better utilization of compute and storage resources (29%), easier management (27%), faster deployment (27%), and easy scaling (26%).

Kubernetes, the most popular container orchestration platform, further enhances the advantages of containerization. Kubernetes can dynamically adjust the number of running containers based on demand, ensuring optimal resource utilization and performance. This scalability is crucial for applications with fluctuating workloads. If a container fails, Kubernetes automatically restarts or replaces it, providing high availability and reliability of applications, which is critical in today's digital business.

Public cloud providers offer Kubernetes as a service for easy consumption and use of containers. This capability is now available for on-premises clouds. Enterprises can now deploy virtual machines and containers side by side in their private datacenters. With a single pane of glass, IT teams can monitor their entire application landscape on-premises. They can instantly provision, scale up, or scale down this environment. Enterprises can use automation to deploy and

update the whole environment with rolling updates from firmware to container, saving companies time and reducing costs.

Another benefit of running Kubernetes in a private cloud is improved resilience. Kubernetes pods will restart if they fail to provide self-healing capability. The distributed nature of Kubernetes eliminates the single point of failure in the control plane, even during node outages. Load balancing is built in to provide a well-performing application experience and consistent user experience. Finally, circuit breaker–style features prevent a failure from cascading across the Kubernetes landscape. This resilience is essential for today's digital business and to ensure customer satisfaction with private cloud-based applications.

TOOL MATURITY

Supporting Kubernetes and virtual machines together is just one example of how modern private cloud platforms have matured rapidly. Other examples included the single pane of glass to provision new private cloud resources and manage day 2 operations with an automated resolution to common issues. In addition, analytics and reporting for the private cloud allow site reliability engineering (SRE) teams and executives to have customized dashboards to plan resources and track resilience proactively. This insight makes new workload placement equally attractive for on-premises private clouds. Companies can look at the overall application performance, uptime requirements, and total cost of ownership/investment to achieve the business objectives when determining where to place future workloads.

Today's private cloud platforms are also essentially equal in ease of use and cloud features compared with public cloud providers. Therefore, it's no longer a critical component in workload placement decisions. For example, private clouds enable DevOps teams to leverage self-service tools to build new environments. SRE and CloudOps teams can leverage infrastructure as code to provision or update private cloud resources. The latest on-premises Kubernetes management systems mean that software engineers can scale up container pods automatically for new applications or increased demand. When combined with the flexibility to implement the latest software and hardware technologies (AI, GPUs, DPUs, storage, etc.) on the company's schedule, private clouds offer a mature alternative to public clouds.

PRIVATE CLOUD AND VMWARE CLOUD FOUNDATION

Broadcom's solution for the private cloud is VMware Cloud Foundation (VCF). This platform is an integrated software-defined stack that combines VMware vSphere for compute virtualization, vSAN for storage, NSX for networking, vSphere Kubernetes Services (formerly Tanzu Kubernetes Grid) for containers, and VCF Operations and Automation for management. VCF is a comprehensive hybrid cloud platform that integrates these products into a comprehensive software-defined private and public cloud management platform.

VCF aims to streamline IT operations by providing a standard, single platform to manage all enterprise cloud resources. Providing an integrated platform can reduce IT's TCO by better leveraging cloud resources, increasing the productivity of operations teams, empowering developer efficiencies, reducing training costs, and improving agility and resilience. IDC recently found that VCF benefited IT by providing a 42% lower three-year cost of operations (*Business Value White Paper, sponsored by VMware by Broadcom*, IDC #US52312224, August 2024). Other vital benefits of VCF include:

- Security, governance, and compliance: Built-in security and compliance tools ensure robust data protection and regulatory adherence. Micro-segmentation and distributed firewalls are also supported.
- Hybrid cloud: VMware Cloud Foundation supports hybrid cloud strategies, enabling integration across all major public cloud providers and applying VCF entitlements with public cloud providers through license portability. So far, Microsoft and Google have announced their support for VCF license portability.
- Scalability: It allows organizations to scale their infrastructure quickly, accommodating growing workloads in virtual machines and modern applications such as containers. Additionally, it enables easier provisioning of new workloads and automatic allocation of cloud resources on demand to meet real-time application needs.
- Hyperconvergence: VCF provides a hyperconverged infrastructure for deploying and managing private and hybrid public clouds. It includes capabilities to manage networking, storage, and computing in software, all via a centralized console.
- Infrastructure-as-code option: It enables developers and CloudOps teams to provision, test, and version control (git-style repository) cloud resources. This allows for code reuse and the ability to shift left provision new cloud resources, as well as improve reliability and agility.

- Flexibility: The VCF platform can run on premises or via an extensive partner ecosystem for hosting options. Hardware support is agnostic, with nearly unlimited support for the underlying infrastructure hardware that enterprises utilize today.
- Performance: VCF optimizes resource utilization and workload mobility to control costs and maximize performance through smart automation, load balancing, and integrated resource management.
- Automation: VCF provides complete life-cycle management control, including enabling the software-defined datacenter. IT automation can be enabled throughout the life cycle to create, maintain, or decommission private cloud resources. Guardrails can be enforced to ensure that VCF automation runs efficiently and safely. Data services management to deliver DBaaS for developers is also included.

CONCLUSION

Enterprises tell IDC that hybrid cloud strategies will dominate IT management. The private cloud is a crucial component of this evolving IT strategy. With feature and capability parity with the public cloud, a private cloud enables a foundational approach to facilitating and supporting today's digital business. IT leaders must equally weigh the placement of new and existing workloads between public and private clouds. By understanding the unique requirements of each application and data set, executives can leverage the strengths of each cloud type to serve the needs of their workloads and business operations optimally. Enterprises should consider increasing future investments and growth in their private cloud. IT leaders may find, after careful analysis, that private clouds offer a lower TCO than alternatives.

IDC provides the following guidance for investing in private clouds and determining workload placements:

- It is widely recognized that a homogeneous IT environment can rarely address all technical requirements for all applications and data. Consider the public cloud and the private cloud in equilibrium when deciding on workload placement and future investments.
- Workload placement may vary based on application and business environments. Enterprise workload placement decisions should balance business and regulatory constraints and include analysis of various factors, including but not limited to costs, data security, privacy, resilience, performance, interoperability, and compliance requirements.

- Transformation is common in both private and public cloud environments. It's a necessary evolution to future proof applications and ensure smoother integration and management.
- Application modernization is essential for easier workload portability. This capability will make moving workloads between public and private clouds commonplace and allow companies to maximize the business value of each cloud.

MESSAGE FROM THE SPONSOR

About VMware Cloud Foundation (VCF)

VCF is a comprehensive private cloud platform that delivers a unified experience by integrating compute, storage, networking, and cloud management. It is designed to support and accelerate traditional, modern, and AI applications. VCF combines best-in-class foundational technologies with a layer of native, built-in automation and operations for a consistent experience.

This unified approach simplifies operations and enhances efficiency by automating the build, deployment, and operations of the entire private cloud infrastructure. With VCF, organizations can focus on developing a robust consumption layer that accelerates developer productivity and benefits immediately from a platform rooted in performance, trust, security, and compliance.

By reducing complexity, operational overhead, and costs, VMware Cloud Foundation empowers IT teams to concentrate on strategic initiatives and application teams to easily consume the latest features without friction. This allows IT leaders and businesses to remain competitive and continue innovating.

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. IDC helps IT professionals, business executives, and the investment community make fact-based decisions on technology purchases and business strategy. More than 1,100 IDC analysts provide global, regional, and local expertise on technology and industry opportunities and trends in over 110 countries worldwide. For 50 years, IDC has provided strategic insights to help our clients achieve their key business objectives. IDC is a subsidiary of IDG, the world's leading technology media, research, and events company.

Global Headquarters

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