

**Vanguard
Report**

October 2023

On-Premises, in Cloud or Hybrid — the Risk of Status Quo

Introduction

The performance demands of new application types, such as generative AI, advanced data analytics, augmented reality/virtual reality (AR/VR) and industrial IoT, are accelerating the rate at which IT infrastructure must evolve. Existing enterprise datacenter assets often fail to meet the demands of newer applications that require expanded storage capacity, faster networking and enhanced compute processing power. These new requirements, along with the ongoing struggle to drive datacenter efficiency improvements (e.g., power, cooling and physical capacity) and find available space to lease, are leading enterprises to take a fresh look at their IT growth and modernization options.

At this juncture, enterprises have three obvious choices (which are not necessarily mutually exclusive): upgrade and modernize existing on-premises datacenter assets to accommodate more resource-intensive workloads; extend or migrate the current datacenter environment to the cloud or to cloud-adjacent colocation facilities (thus gaining access to modern IT infrastructure without the capex); or rearchitect the applications using cloud-native frameworks and containers, thus allowing them to operate across modernized on-premises IT environments *and* public cloud platforms.

Every enterprise IT workload has an operating environment or “best execution venue” that optimally balances performance, cost and other considerations such as capacity requirements, data isolation/sovereignty, latency, sustainability and application dependencies. With these factors in mind, some enterprise applications may not be good candidates for public environments. Nevertheless, the need remains for IT infrastructure that accommodates the resource and performance requirements of new enterprise applications.

The Take

New business requirements and technology advancements are forcing enterprises to rethink and revamp their IT environments to meet the increasing resource demands of new applications that incorporate generative AI, machine learning and advanced data analytics. Cloud-native “run anywhere” workloads are becoming the rule rather than the exception (especially for net-new applications) and will execute across both public cloud and enterprise datacenter/colocation environments. This confluence of events has created the ideal circumstances for organizations to consider their IT modernization options — which workloads to keep on-premises in modernized server fleets and which to migrate or extend into public cloud environments. The public cloud offers the advantage of immediate infrastructure modernization (in addition to other benefits), but business, regulatory or technology considerations may dictate that some workloads remain on-premises, requiring modernized IT infrastructure in the enterprise datacenter or colocation facility as well.

While the up-front cost of on-premises server modernization may be daunting, inaction may prove to be more costly in the long run, resulting in business risk through hasty wholesale migration to the public cloud, missed opportunities to incorporate new digital technologies into business operations or inability to sustain competitive advantage in the market. The hybrid approach offers the best of both worlds: Speed, scale and agility for workloads that can move to public clouds and more performant, efficient and sustainable infrastructure for those that cannot.

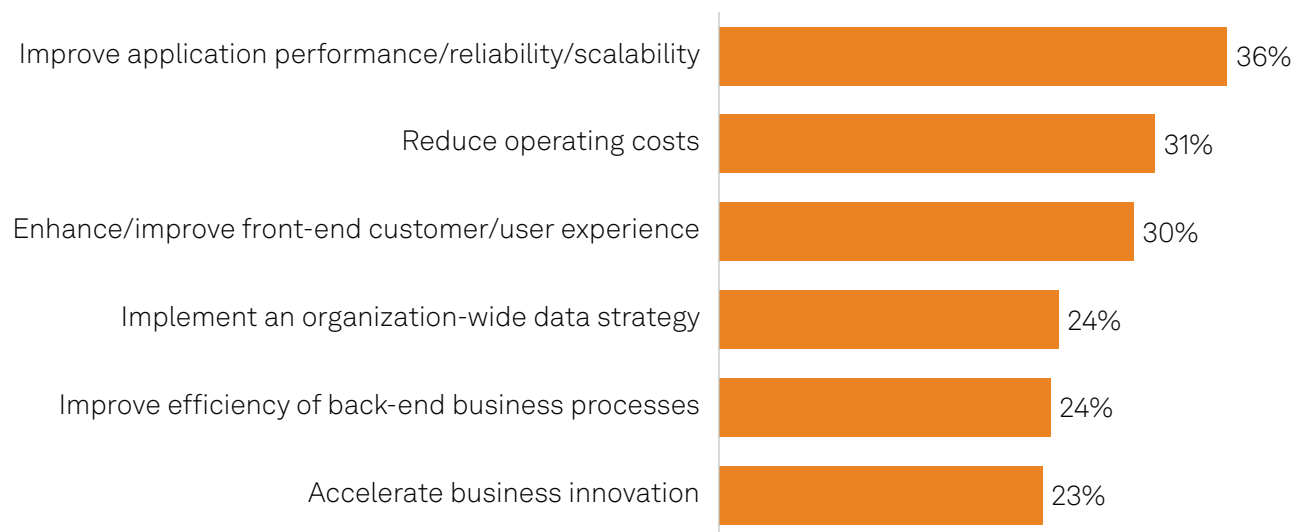
Business requirements drive the need for IT modernization

In the digital economy, long-standing ways of doing business are giving way to new models fueled by data-driven, processing-intensive workflows. Technology is a key enabler of the digital economy, requiring accelerated IT modernization efforts. As a result, digital business imperatives are driving organizations to reassess the state of their enterprise IT environments to make them more responsive to business needs. Enterprises are exploring use cases for generative AI and embedding the capabilities into business processes to drive speed, efficiency, productivity and creativity across numerous domains. Alongside generative AI, organizations look to other advanced technologies such as industrial IoT, robotic process automation and AR/VR to improve product/service delivery, support, customer engagement and other business functions.

However, new applications and workloads incorporating these technologies come with new resource capacity and processing requirements, and many organizations' current datacenters and IT environments are not up to the task. IT modernization is never a one-and-done endeavor — yesterday's top-of-the-line servers are today's legacy servers. To address the datacenter capacity deficit, enterprises must think about where they want to take their businesses, which technologies can help them get there and how to modernize the IT estate in ways that enable the applications (i.e., the business ideas that execute digitally as software) to run as scalable, operationally efficient and cost-effective processes.

Application modernization efforts have accelerated over the past year, according to data from 451 Research's Voice of the Enterprise: Cloud, Hosting & Managed Services, Application Modernization 2022 survey. A majority of organizations with IT modernization plans in progress (59%) report that activity has intensified over the past 12 months, with almost 20% saying it has "significantly accelerated." Improved performance of IT environments tops the list of modernization drivers, followed closely by cost reduction (see Figure 1). This indicates close alignment between improved operational efficiency (processing power, energy consumption, datacenter footprint utilization, etc.) and total cost of ownership.

Figure 1: IT modernization drivers



Q. Which of the following — if any — are the primary drivers for your organization's efforts to modernize legacy applications? Please select up to three.

Base: Organizations that are currently planning or developing a plan for application modernization (n=523).

Source: 451 Research's Voice of the Enterprise: Cloud, Hosting & Managed Services, Application Modernization 2022.

As organizations evolve toward data-driven decision-making, inadequate infrastructure poses a barrier to AI and machine learning (ML) implementation and becomes a bottleneck as workload demands expand. As a result, enterprises are ramping investment in high-performance infrastructure. A recent Voice of the Enterprise: AI & Machine Learning, Infrastructure survey reveals that nearly 90% of organizations expect to increase their AI/ML spending in the next 12 months, and more than half anticipate moderate to significant expenditure increases.

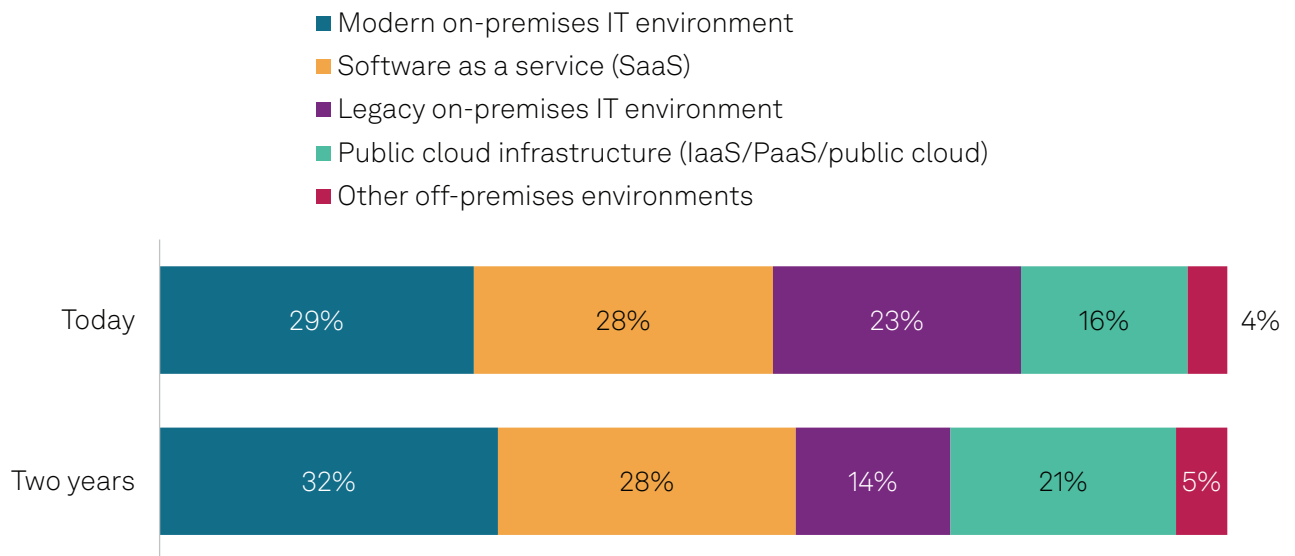
Emerging generative AI workloads are also driving changes in the underlying IT infrastructure requirements due to the tremendous processing capacity required to train large language models (LLMs) and the storage and network capacity essential to run applications and business processes that leverage these LLMs. AI has unique infrastructure requirements, and in the absence of integration with existing systems, AI infrastructure planning requires a different approach to the one used for general-purpose workloads. IT leaders must also factor in the datacenter dimensions, given that higher-performing application requirements of generative AI and other emerging technologies deployed on legacy infrastructure may strain the power and cooling capacity of current facilities. This will require an assessment of existing IT assets' operational costs and capacities against modern platform capabilities.

The path to modernization

Today, enterprise workloads are trending toward modernization, both in public cloud and on-premises. Primarily, this means moving workloads out of legacy environments and into modern, advanced infrastructure environments whether in on-premises datacenters, in a public cloud or a mixture. Modernized IT environments must support varying workload execution venues (public cloud, private cloud, on-premises and colocation), application packaging (VMs, containers, bare metal) and diverse systems with processing speed, storage and network throughput capacity optimized for different types of workloads.

Organizations have choices about how they procure, deploy, integrate, manage and transform their digital infrastructure stacks. The “either/or” IT binaries of public versus private cloud and on-premises versus off-premises have given way to converged approaches and hybrid/multicloud IT estates spanning private and public cloud environments. For example, 26% of organizations anticipate that public cloud infrastructure or other off-premises environments will be the primary workload execution venue in the next two years, up from 20% today, according to 451 Research’s Voice of the Enterprise: Digital Pulse, Workload Placement 2023 survey. Modern on-premises IT environments remain in the mix, with nearly one-third of organizations expecting this to be the primary venue for their workloads, thanks in large part to migration from legacy on-premises IT environments (see Figure 2).

Figure 2. Workload migration trends toward modernization



Q. Which of the following best describes the primary IT environment used to operate your organization’s workloads today?
Q. And which of the following best describes the primary IT environment your organization will use to operate workloads two years from now?

Base: All respondents (n=918).

Source: 451 Research’s Voice of the Enterprise: Digital Pulse, Workload Placement 2023.

IT is a significant contributor to many organizations' environmental impact. Therefore, the energy efficiency of enterprise IT resources (whether on-premises or in the cloud) is also becoming an important factor in IT decision-making and modernization planning. In fact, 63% of enterprises report that datacenter efficiency is "very important" to their organizations, according to 451 Research's Voice of the Enterprise: Datacenters, Sustainability 2023 survey data. Similarly, IT vendors' sustainability goals and the credibility of their commitment to mitigate environmental impacts are key factors in technology and purchasing for 75% of respondents, as noted in 451 Research's most recent Voice of the Enterprise: Digital Pulse, Sustainability survey.

For organizations with aging datacenters, legacy applications and technical debt, migrating to modern architectures and deploying upgraded servers will reduce power and cooling requirements, enhance performance, improve IT efficiency via infrastructure consolidation and reduce the physical server footprint. This reduction in required space through platform modernization coupled with the hybrid approach of offloading some workloads to cloud will enable enterprise datacenters to free up capacity to introduce AI/ML workloads in existing facilities. The improved scale, agility and cost profile that comes with more powerful infrastructure and consolidation means that organizations can spend more time leveraging technology and data to develop applications that drive business efficiency and less time patching and managing obsolete infrastructure.

The organizational dynamics of modernization will also require new IT skills and possibly new IT roles and personas such as platform and cloud-native engineers. Tighter alignment between business and IT requirements is also a must-have on the road to modernization. Continuous change is now business as usual in the digital economy. Organizations need to be confident that their IT environments can support a range of business outcomes with flexibility, agility and security. Scalable, performant and efficient IT infrastructure is an important strategic foundation for building this confidence.

Modernization and the benefits of hybrid

Despite widespread adoption of public cloud, on-premises IT estates persist because few organizations have the luxury of starting over on a clean sheet of (IT) paper. As a result, hybrid environments are becoming the default IT architecture. However, enterprise workloads have different resource, performance, security and compliance requirements, and one size may not fit all when it comes to IT infrastructure environments.

Hybrid IT environments offer significant advantages for on-premises and off-premises workloads. Cloud-scale infrastructure radically accelerates the dynamics of IT provisioning and infrastructure change management, which saves time and money that would otherwise be spent on infrastructure deployment and scaling. Additionally, hyperscale cloud providers offer an array of computing options, allowing enterprises to choose the CPU and/or graphics processing unit (GPU) instances that best suit their application requirements and pay for resources using various on-demand, resource-commit and savings-plan models. The operational elasticity of public cloud allows organizations to scale compute up or down based on processing, memory and storage demands at runtime. Ongoing supply constraints and costs in the GPU space make the hyperscalers and their growing service portfolios an especially attractive option for enterprises with traditional AI/ML and generative AI use cases.

At the level of individual workloads, public cloud's self-service provisioning and on-demand access provides an off-ramp for enterprises looking to lift and shift applications from space- and capacity-constrained datacenters to make room for new AI/ML workloads running on modernized on-premises infrastructure. Public cloud offload and upgraded servers in modernized datacenters also facilitate IT rationalization and consolidation for workloads that enterprises want to keep on-premises due to regulatory or compliance mandates, data or other application dependencies, data I/O volume and costs, performance, and security/risk considerations. As applications and use cases for generative AI emerge, enterprises are exploring hybrid approaches that enable model training to occur mostly in the cloud with inferencing (i.e., making data-driven decisions based on the training) on-premises in centralized or edge locations.

The journey to a hybridized modern IT environment is worth the trip, but watch out for speed bumps

The hybrid approach gives enterprises options for the timing and manner of IT modernization, but it also introduces challenges that include a mix of technology, organizational and economic factors, including the following:

- **Operational complexity.** In hybrid IT scenarios, workloads and data operating in different environments may be siloed, making it difficult to operate infrastructure and application resources consistently across the IT estate and move data and workloads freely and securely between different environments. The oft-referenced “single pane of glass” that provides a unified operational perspective takes time to implement, and many organizations lack skills and/or IT headcount to make the move to hybrid without incurring opportunity costs elsewhere.
- **Data management and security.** From a security perspective, hybrid cloud entails the tracking and monitoring of different workloads in different states. This requires protecting data in transit and at rest, as well as maintaining governance and controlling who is granted access to what and at which points — whether this is on-premises or off-premises, hosted or non-hosted.
- **Skills shortages and lack of personnel.** Hybrid IT requires a diverse set of skills: programming, tools development, project management and operational management. According to recent 451 Research Voice of the Enterprise survey data, public cloud integration with on-premises IT systems is a key capability currently in short supply. This means that recruiting and retaining personnel with both cloud and integration/migration expertise may require a premium.

The modernization imperative: Doing nothing is not an option

In the digital economy, organizations’ ability to drive innovation, business value and competitive differentiation is data-centric and IT-dependent. As a result, organizations must ensure their business strategies can be executed in modernized IT environments that are optimized for change and designed to handle both current and future workload demands. The technical debt that builds by continuing to support aging on-premises infrastructure undermines the IT organization’s ability to respond to evolving business requirements.

Modernizing on-premises infrastructure enables enterprises to support new and existing system and application requirements at scale, with improvements in price/performance, operational efficiency and sustainability. Embracing modernized IT and integrating upgraded components from the chipset up positions enterprises to reap the rewards today on a foundation primed to support tomorrow’s data- and processing-intensive workloads. Taking a hybrid approach and adding public cloud environments to the mix provides a fast track to modernized infrastructure for certain workloads while enabling the emerging enterprise preference for AI training in the cloud and inferencing at on-premises and edge locations.

Conclusion

The advent of generative AI/LLM applications is transforming content creation, data analytics, customer service and, especially, software development and maintenance. This technology is bringing innovative business offerings to market at a pace that will be tremendously disruptive and require substantial IT infrastructure capabilities that are not currently resident in most enterprise datacenters. For enterprise IT estates of all architectures, modernizing servers will be necessary to answer the capacity demands of these new applications. Modernization will also reduce power and cooling requirements and offer enhanced performance, stronger security and efficiency-enabling infrastructure consolidation.

These benefits, in conjunction with reducing the physical server footprint and freeing up valuable datacenter space, justify the investment. IT infrastructure modernization is not a challenge that can be put off until tomorrow, but one enterprise IT leadership must address today. The emergence of a new competitive AI-driven landscape demands attention now, so postponing the enablement of these new applications could have significant business consequences.



AMD addresses the needs of enterprises with a complete portfolio of innovative and proven data center products. A unique combination of security features, flexibility, and outstanding TCO supports your applications and platforms in the cloud, on-premises, and at the edge. An industry leader, AMD brings numerous strengths to help companies achieve their business goals: close engagement with customers to develop real-world solutions; a vibrant partner ecosystem offering; a well-defined roadmap; and stability and consistent growth to power advanced R&D. Discover more about [AMD EPYC™ processors](#).

About the author



Melanie Posey

Research Director, Cloud & Managed Services Transformation

Melanie Posey is a research director within the Technology, Media and Telecommunications group of S&P Global Market Intelligence. She leads the Cloud & Managed Services Transformation team and manages the Voice of the Enterprise cloud survey. Melanie has a long history as a technology industry analyst. In 2016, she joined 451 Research (acquired by S&P global in 2019), bringing experience in strategy analysis, forecasting and enterprise/consumer insight for the cloud, hosting, IT services and telecommunications markets.



Bobby Clay

Senior Consulting Analyst, TMT Consulting

Bobby is a senior consulting analyst/manager within S&P Global Market Intelligence's Technology, Media and Telecommunications Research group. He is focused on cloud, datacenter, generative AI and metaverse technologies and has a unique mix of technical and business skills with more than 30 years of industry experience. Bobby's recent areas of concentration include hyperscalers, generative AI cloud tools, metaverse virtual world and avatar construction technologies, and datacenter modernization and sustainability.

About S&P Global Market Intelligence

At S&P Global Market Intelligence, we understand the importance of accurate, deep and insightful information. Our team of experts delivers unrivaled insights and leading data and technology solutions, partnering with customers to expand their perspective, operate with confidence, and make decisions with conviction.

S&P Global Market Intelligence is a division of S&P Global (NYSE: SPGI). S&P Global is the world's foremost provider of credit ratings, benchmarks, analytics and workflow solutions in the global capital, commodity and automotive markets. With every one of our offerings, we help many of the world's leading organizations navigate the economic landscape so they can plan for tomorrow, today. For more information, visit www.spglobal.com/marketintelligence.

CONTACTS

Americas: +1 800 447 2273

Japan: +81 3 6262 1887

Asia Pacific: +60 4 291 3600

Europe, Middle East, Africa: +44 (0) 134 432 8300

www.spglobal.com/marketintelligence

www.spglobal.com/en/enterprise/about/contact-us.html

Copyright © 2023 by S&P Global Market Intelligence, a division of S&P Global Inc. All rights reserved.

These materials have been prepared solely for information purposes based upon information generally available to the public and from sources believed to be reliable. No content (including index data, ratings, credit-related analyses and data, research, model, software or other application or output therefrom) or any part thereof (Content) may be modified, reverse engineered, reproduced or distributed in any form by any means, or stored in a database or retrieval system, without the prior written permission of S&P Global Market Intelligence or its affiliates (collectively S&P Global). The Content shall not be used for any unlawful or unauthorized purposes. S&P Global and any third-party providers (collectively S&P Global Parties) do not guarantee the accuracy, completeness, timeliness or availability of the Content. S&P Global Parties are not responsible for any errors or omissions, regardless of the cause, for the results obtained from the use of the Content. THE CONTENT IS PROVIDED ON "AS IS" BASIS. S&P GLOBAL PARTIES DISCLAIM ANY AND ALL EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, ANY WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE OR USE, FREEDOM FROM BUGS, SOFTWARE ERRORS OR DEFECTS, THAT THE CONTENT'S FUNCTIONING WILL BE UNINTERRUPTED OR THAT THE CONTENT WILL OPERATE WITH ANY SOFTWARE OR HARDWARE CONFIGURATION. In no event shall S&P Global Parties be liable to any party for any direct, indirect, incidental, exemplary, compensatory, punitive, special or consequential damages, costs, expenses, legal fees, or losses (including, without limitation, lost income or lost profits and opportunity costs or losses caused by negligence) in connection with any use of the Content even if advised of the possibility of such damages.

S&P Global Market Intelligence's opinions, quotes and credit-related and other analyses are statements of opinion as of the date they are expressed and not statements of fact or recommendations to purchase, hold, or sell any securities or to make any investment decisions, and do not address the suitability of any security. S&P Global Market Intelligence may provide index data. Direct investment in an index is not possible. Exposure to an asset class represented by an index is available through investable instruments based on that index. S&P Global Market Intelligence assumes no obligation to update the Content following publication in any form or format. The Content should not be relied on and is not a substitute for the skill, judgment and experience of the user, its management, employees, advisors and/or clients when making investment and other business decisions. S&P Global keeps certain activities of its divisions separate from each other to preserve the independence and objectivity of their respective activities. As a result, certain divisions of S&P Global may have information that is not available to other S&P Global divisions. S&P Global has established policies and procedures to maintain the confidentiality of certain nonpublic information received in connection with each analytical process.

S&P Global may receive compensation for its ratings and certain analyses, normally from issuers or underwriters of securities or from obligors. S&P Global reserves the right to disseminate its opinions and analyses. S&P Global's public ratings and analyses are made available on its websites, www.standardandpoors.com (free of charge) and www.ratingsdirect.com (subscription), and may be distributed through other means, including via S&P Global publications and third-party redistributors. Additional information about our ratings fees is available at www.standardandpoors.com/usratingsfees.