

Businesses should complement investments in AI workloads with appropriate investments in digital infrastructure. This approach enables better scaling, agility, and operational efficiency. Organizations must prioritize datacenter modernization alongside application modernization, placement, and cyber-resilience.

Strategic Investments in Datacenters for the AI Era

May 2025

Written by: Ashish Nadkarni, Group VP and General Manager, Worldwide Systems Infrastructure Research

Introduction

AI is disrupting the status quo across the board. Businesses are increasingly investing in it to drive long-term differentiation, customer experience, product and service innovation, and operational efficiency. For enterprise companies, the focus is on IT optimization, digital assistants, supply chain management, and AI-driven marketing and digital initiatives. All of these benefit from investments in AI platforms and data architectures, which enable businesses to gain real-time customer insights, make informed data-driven decisions, and collaborate across the C-suite to influence long-term strategy.

Nowhere is this disruption felt more than in the IT organization, specifically on the datacenter infrastructure that enables these new AI initiatives to deliver consistent, tangible, and measurable outcomes. At the same time, businesses are racing to embed AI into existing workloads through modernization, building security and privacy at the core of their data architecture, and ensuring that their overall application portfolio is as "AI ready" as possible.

For CIOs and IT decision-makers (ITDMs), now is the time to examine their organization's IT strategy. They must reevaluate tried-and-true approaches, such as moving all applications to public cloud environments, with a focus on return on investment, data, performance, and security requirements for various AI workloads, both embedded and standalone.

AT A GLANCE

KEY STATS

- » By 2027, resilience, AI, and privacy regulations will drive 70% of the G5000 to reengineer most core business workloads to be portable across multiple private and public cloud infrastructures.
- » By 2027, 75% of large-scale datacenters will have implemented advanced energy management systems, incorporating AI and ML to optimize power usage and reduce carbon footprints.
- » By 2029, 50% of the G2000 will experience combined annual losses totaling over \$100 billion because of tech debt and lack of modernization as they struggle to protect older infrastructure and meet the demand for AI.

KEY TAKEAWAYS

- » Develop a hybrid IT strategy and operating model as a precursor to workload modernization and investments in AI workloads.
- » Invest in self-owned/self-operated datacenters as the anchor for a hybrid IT strategy.
- » Optimize a return on investment in datacenter facilities through careful planning and mitigation strategies.

Specifically, CIOs and ITDMs benefit from a hybrid IT strategy that places their own datacenters at the center of a connected, complex, and geo-distributed landscape that includes public clouds, private clouds, and edge environments. IDC finds that application modernization, enhanced digital resilience capabilities, and core IT modernization will see the most significant levels of IT spending increase in 2025. CIOs and ITDMs must invest strategically in datacenter modernization to create a fit-for-purpose environment anchored by highly efficient and high-performing silicon, networking, compute, and storage infrastructure that supports modern AI-enabled workloads. By investing in automation and AI-enabled observability, they can increase service quality and assurance. Furthermore, investments in alternative energy sources, cooling, and power technologies can help meet or exceed their organization's sustainability objectives.

Investing in Datacenters as Part of a Hybrid IT Strategy

The popularity of public cloud services has made IT organizations question the long-term viability of investing in and maintaining their own datacenters. Businesses often pivot to a public cloud-first strategy as part of their digital transformation initiatives. With careful planning, public cloud services offer better economics, scaling, and agility for some situations.

However, as businesses accelerate the adoption of AI workloads, an all-or-nothing strategy can prove expensive and even risky. AI workloads present a particular set of requirements, and as developers and ISVs infuse AI algorithms into business workflows, these requirements spread to a larger subset of business workloads. AI workloads pose specific data security and privacy concerns and using specialized hardware for extended periods on a pay-as-you-go basis can be expensive in the long run. CIOs and ITDMs should adopt a hybrid approach, anchoring in an on-premises datacenter while leveraging public cloud resources as needed when the economics, compliance, and performance requirements are favorable. IDC finds that hybrid infrastructure strategies, using a workload-optimized mix of dedicated on-premises datacenters, colocation facilities, edge, and public cloud infrastructure, continue to be the norm, with 77% of organizations reporting they operate their own datacenter. So what does a modern datacenter strategy look like? For starters, it includes:

- » **High-performance networking:** A foundational element of the stack is low-latency, high-bandwidth silicon and networking to support AI compute and storage. Where scale-out clusters are used, this network serves as a fabric for node-level communication and coherency.
- » **AI-ready compute and storage:** This stack includes high-performance and accelerated compute, such as high-core count processors or accelerators such as GPUs, coupled with high-performance memory and high-performance storage for analyzing large data sets. The stack is secure by design to protect data and infrastructure.
- » **Cloud-scale architecture:** The stack offers scaling and parallelization capabilities that can scale out horizontally and scale up vertically, in addition to supporting efficient parallelization. This architecture usually relies on high-speed, low-latency interconnects that can enable low-latency communication for distributed tasks.
- » **Optimized software stack:** This stack includes deep learning frameworks and GPU-specific libraries, coupled with a robust data management layer that enables seamless in-place analytics. The software stack also provides compliance controls that enable security teams to maintain data security and privacy.

- » **AI-enabled automation:** The system provides comprehensive mechanisms for performance tracking, monitoring, and logging of security, compliance, and service quality issues. It also enables efficient allocation and management of resources, automation, and corrective action.
- » **Secure by design:** Security is more than just logging security issues. It is built into the fabric of the datacenter and protects everything from the infrastructure to the workloads to AI models.
- » **Streamlined operating costs:** Increasing operational efficiency and reducing IT operating costs are top priorities. IDC finds that more than half of surveyed CIOs and ITDMs say their organizations are prioritizing investments in infrastructure automation and AI-enabled observability in 2025.

Benefits to Organizations Investing in Modernizing Their Datacenters

By placing their own modern datacenters at the center of a connected, digitally resilient landscape, CIOs and ITDMs can usher in an era of agility, scalability, and digital resilience for modern AI workloads. The process of modernizing a datacenter involves updating and optimizing its infrastructure to improve efficiency, performance, and sustainability. With this kind of modernization, the following benefits are possible:

- » Accelerate time to deployment and value with validated solutions and modular, full-stack systems; increase rack density to enhance efficiency.
- » Power AI anywhere, supporting new modern workloads at scale by moving the computing stack closer to the data sources.
- » Infuse security from the ground up with distributed enforcement points and ensure compliance with evolving regulatory standards.
- » Gain full observability across the internet, the network, and the infrastructure stack with advanced tooling and automation systems.
- » Improve energy efficiency with advanced cooling technologies, rack standards, and alternative and renewable power sources.
- » Meet regulatory or sovereignty responsibilities without compromises.

Strategic datacenter investments, combined with operational transformation objectives, enable CIOs and ITDMs to optimize return on investment, enhance infrastructure efficiency, and meet corporate sustainability goals. This process is crucial for meeting the demands of digital transformation and maintaining a competitive advantage.

How CIOs and ITDMs Can Make Datacenter Investments Work in Their Favor

Investing in datacenter facilities can be resource intensive, and costs can quickly add up. However, with careful planning, CIOs and ITDMs can maximize their return on investment. Here are some familiar challenges and how CIOs and ITDMs can mitigate them preemptively:

Strategic Cost Distribution Through Hybrid Cloud Placement

- » **Concern:** Building and maintaining a datacenter requires significant capital expenditure, including hardware, software, and physical infrastructure.
- » **Mitigation:** Pursue a hybrid IT strategy that combines on-premises, hosted, and public cloud services into a single hybrid cloud operating model. Develop and implement a robust workload modernization and placement strategy in which the organization places the most data-sensitive and performance-critical workloads on premises. Leverage flexible consumption offerings from vendors to spread initial costs over time.

Operating Costs Management and Mitigation

- » **Concern:** Datacenters require ongoing maintenance, skilled personnel, and management, which can be resource intensive and lead to higher operating costs.
- » **Mitigation:** Where possible, leverage colocation facilities for expansion, ensuring that the network adjacency allows the facility to operate as an extension of the on-premises footprint. Invest in edge locations for more distributed workload placement. Simplify and reduce operating costs where possible with infrastructure choice. For example, the addition of AI workloads can be blended into existing operational models, using the same skills and technology currently implemented for enterprise applications. Invest in solutions designed to optimize infrastructure performance. Finally, consider outsourcing datacenter management to third-party providers or use managed services to reduce the burden on internal teams.

Scalability Challenges

- » **Concern:** Scaling up a physical datacenter can be slow and costly compared with the flexibility that cloud services offer.
- » **Mitigation:** Implement scalable infrastructure solutions and consider cloud bursting for peak loads. Utilize scheduling and cloud-scaling software that can make hybrid cloud operations seamless.

Disaster Recovery

- » **Concern:** Ensuring robust disaster recovery and business continuity plans can be complex and expensive.
- » **Mitigation:** Develop comprehensive disaster recovery plans and use cloud-based backup solutions to ensure data redundancy. In addition, leverage cloud-based disaster recovery strategies that enable quick operational recovery in one or more public cloud services.

Rapid Technological Changes/Tech Debt

- » **Concern:** Keeping up with the latest technological advancements can be challenging and costly, leading to potential obsolescence.
- » **Mitigation:** Regularly update and upgrade infrastructure and consider flexible consumption or leasing options when procuring new infrastructure stacks. View technical debt as a strategic priority that impairs flexibility, stifles innovation, and reduces operational efficiency. Proactively seek to retire legacy systems, outdated architecture, and data debt. Prioritize application modernization, use modern development tools, and integrate technical debt remediation into strategic plans.

Energy Consumption

- » **Concern:** Building and maintaining a datacenter requires significant capital expenditure, including hardware, software, and physical infrastructure.
- » **Mitigation:** Invest in energy-efficient hardware and cooling solutions. Explore renewable energy options to reduce operational costs and environmental impact. If possible, place datacenter facilities in locations where network and power capacity are cheaper compared with densely populated metropolitan areas. Invest in lights-out management to minimize resource overhead.

By adopting these strategies, companies can mitigate the challenges associated with owning datacenters while leveraging their benefits. Cloud services offer greater flexibility, scalability, and cost efficiency, making them a preferable option for businesses.

Considering Cisco Solutions for Datacenters in the AI Era

As one of the leaders in networking and datacenter solutions, Cisco has a distinct perspective on the requirements for AI-ready datacenters. As a trusted network provider for many leading hyperscalers and enterprises, Cisco has seen the massive demands placed by both traditional and AI workloads. The company has also seen how datacenters operate within a hyper-distributed and interconnected landscape of owned and unowned infrastructure. Organizations of all sizes need to deploy, operate, and secure datacenters within this context. To help businesses meet their needs for AI datacenters, Cisco unifies silicon, network, compute, security, and observability. The company aims to help organizations build AI-ready datacenters and accelerate their AI initiatives.

AI workloads live at core and edge datacenters, colocation sites, and public clouds. Cisco has a comprehensive portfolio of networking that offers reliable, high-performance fabrics throughout and between these AI datacenters. Massively scalable fabrics are designed to meet demanding AI requirements for high throughput and low latency. Innovation with Cisco Silicon One offers robust switching and networking performance and a single silicon architecture that breaks down dividing lines across datacenter connectivity. The ability to deploy networking, compute, and storage in fully integrated, scalable, and modular systems and stacks such as Cisco AI PODs and Cisco Hyperfabric AI also helps power traditional and AI workloads. Cisco Validated Designs provide documented and tested approaches for designing, deploying, and configuring Cisco networking, compute, and complementary third-party partner technologies such as NVIDIA AI Enterprise software and Red Hat OpenShift, dramatically reducing the complexity that can slow AI initiatives. Cisco has a strategic partnership with NVIDIA and recently announced the Cisco Secure AI Factory with NVIDIA solution. This solution

enables a secure, scalable, and high-performance AI infrastructure needed to develop, deploy, and manage AI workloads securely and at scale.

Seamless operations and observability are intended to enhance an organization's agility by offering the ability to work with familiar operational models across multiple environments. This provides the visibility to help ensure timely resolution of datacenter issues. Cisco helps remove silos with unified management, observability, and assurance for traditional and AI workloads across all environments. Organizations can leverage existing tools, skill sets, and operational models to manage both traditional and AI workloads. All compute resources can be managed centrally with Intersight, and all datacenter networks are run from one place with Nexus Dashboard. In addition, having access to visibility and insights everywhere in the stack drives digital resilience. To resolve issues faster, Cisco's observability solutions provide one place to look across every aspect of a company's digital operations (owned and unowned), including on-premises datacenters, public clouds, internal networks, and the internet.

Resilience is enhanced by securing digital infrastructure from ground to cloud. To protect hyper-distributed workloads more fully, organizations can infuse security everywhere with a comprehensive Cisco security portfolio that can build integrated, layered defenses, from datacenters to multicloud environments. For example, Cisco Hypershield is a distributed, AI-native security architecture that puts security wherever it needs to be (i.e., in every software component of every application running on the network, on every server, and in public or private cloud deployments). Cisco AI Defense is an end-to-end AI security solution that protects the development, deployment, and usage of AI against safety and security risks. Cisco also offers security analytics built on a great breadth of data.

Challenges

CIOs and ITDMs consider IT vendors as partners in their digital transformation journey. The level of trust a company places in an infrastructure provider is related to its ability to support efficient datacenter infrastructure. CIOs and ITDMs are looking to invest in secure and resilient datacenter infrastructure. Such an infrastructure enables the organization to achieve or exceed its sustainability goals while reducing its datacenter footprint, including through workload consolidation. This includes the ability to host performance-intensive AI workloads alongside other business- and mission-critical enterprise workloads. The datacenter infrastructure must be capable of managing latency- and bandwidth-sensitive workloads alongside network, memory, and compute-intensive workloads.

Cisco recognizes the challenges and opportunities presented by the pace of AI innovation. In fact, 85% of respondents of Cisco's AI Readiness Index say they have less than 18 months to deploy an AI strategy, or they will see negative business effects. To enable businesses to realize their strategy, Cisco must continue to take a more prescriptive approach that leans into its own products and solutions. Cisco must continue to articulate its value proposition in a way that resonates with CIOs and ITDMs. The differentiation for Cisco is to deliver efficient, agile, and secure infrastructure solutions that transform its relationship with CIOs and ITDMs into that of strategic and reliable partner.

Conclusion

As organizations look to leverage AI more across their business, CIOs and ITDMs need to develop a strategy to ensure their datacenters are modernized to handle the requirements today and can scale to meet the needs of the business. A hybrid IT strategy and operating model ensures that technology investments today do not become liabilities tomorrow, enabling quicker product launches, improved customer experiences, and better resource management. This includes datacenter facilities that are essential for housing data sets crucial to business differentiation. With regular tracking and reporting of technology currency, CIOs and ITDMs can ensure an optimal return on investment.

A hybrid IT strategy and operating model, with strategic investments in self-owned or self-operated datacenters, ensures technology investments can enable quicker product launches, improved customer experiences, and better resource management.

About the Analyst



Ashish Nadkarni, Group VP and General Manager, Worldwide Systems Infrastructure Research

Ashish Nadkarni leads IDC's worldwide research on compute and storage infrastructure systems, platforms and technologies, enterprise, emerging and performance-intensive workloads, cloud and edge infrastructure and infrastructure services, and infrastructure software platforms.

MESSAGE FROM THE SPONSOR

Cisco helps transform data centers to power traditional and AI workloads anywhere. Only Cisco unifies networking, compute, security, and observability to help organizations modernize and advance their data centers. Cisco uniquely brings together comprehensive infrastructure including modular building blocks of computing, networking, and silicon, plus full-stack systems, seamless operations and observability, and security from ground to cloud. With Cisco, organizations can accelerate AI initiatives without compromise. We help maximize agility with faster time to value at massive scale, while building digital resilience to keep data centers securely up and running.

Learn more: <https://www.cisco.com/site/us/en/solutions/artificial-intelligence/infrastructure/index.html>.



The content in this paper was adapted from existing IDC research published on www.idc.com.

IDC Research, Inc.
140 Kendrick Street
Building B
Needham, MA 02494, USA
T 508.872.8200
F 508.935.4015
blogs.idc.com
www.idc.com

IDC Custom Solutions produced this publication. The opinion, analysis, and research results presented herein are drawn from more detailed research and analysis that IDC independently conducted and published, unless specific vendor sponsorship is noted. IDC Custom Solutions makes IDC content available in a wide range of formats for distribution by various companies. This IDC material is licensed for external use, and in no way does the use or publication of IDC research indicate IDC's endorsement of the sponsor's or licensee's products or strategies.

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 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.

©2025 IDC. Reproduction is forbidden unless authorized. All rights reserved. [CCPA](#)