Powering What's New — And What's Next

Your AI-Ready Data Center Strategy

The day the infrastructure couldn't keep up

You've been here before: Another day, another performance spike. That's because you're still running mission-critical apps on legacy infrastructure that's stretched thin and wasn't designed for today's scale. Taking it offline to rework everything? Not an option. Your workloads keep growing, with compute, networking, and storage pushed to its limits. And your team likely spends more time firefighting than forward-planning.

Meanwhile, the business is pushing into AI and you're suddenly facing computational and infrastructure demands you've never seen before. You're being asked to do more: modernize, move faster, stay secure – all without downtime, budget spikes, or extra headcount.

What you need now isn't just an upgrade – it's a data center that powers all workloads, scales for exponential growth, secures the entire stack, unifies management, and keeps resilient. And the leaders who act now can position their organizations for long-term success no matter what comes their way.

The challenge:

From infrastructure tipping point to transformation

The AI era has introduced a new reality for CIOs and senior IT leaders: incredible opportunities, but also intense cost pressures, business demands for rapid delivery of AI capabilities and use cases, and an intensifying threat landscape.

- Despite increased investment in AI, the overall cost is not always well understood, and Gartner analysis[1] has indicated the potential for ongoing costs to get out of control by 5-10x.
- Organizations are in a race to develop AI-native applications for competitive advantage. This means they must reduce time-to-inference and training cycles to stay relevant. The rise of GenAI, LLMs, and agentic AI heightens the pace of change and innovation in IT.
- Bad actors utilize AI in their attacks, and AI itself introduces a whole new range of safety and security concerns.

What does this mean for infrastructure? Traditional data centers were never designed for this reality. Consider this: inferencing and training workloads are now up to 20x more compute-intensive, consuming 10x the power of traditional applications.[2] Networking, compute, and storage systems are straining under the sheer data footprint of AI. Add in the need for real-time processing at the edge, and the cracks in legacy infrastructure become impossible to ignore.

The bottom line: We've reached an inflection point.

What makes a data center ready for AI?

1

Powers all workloads – everywhere

Traditional and AI workloads must coexist. Data center infrastructure must handle increasing numbers of traditional workloads, plus AI workloads, no matter where they reside. Whether it's on-prem, in the cloud, or at the edge, AI should move closer to the data – not the other way around.

2

Scales for exponential growth

High-density compute, lossless low-latency networking, and sustainable growth must be prioritized. Energy efficiency is no longer optional – it's a strategic imperative.

3

Secures the entire stack

Security must be fused into the data center fabric, with enforcement at every layer – from infrastructure to networks to applications, and even the AI models themselves.

4

Simplifies operations

Unified management of traditional and AI workloads allows teams to operate seamlessly and manage all workloads uniformly.

5

Resilience baked in

When disruptions happen, organizations must detect, respond, and resolve – fast.

How Cisco fuels the modern data center

The critical components to power data centers

Robust and flexible infrastructure

+

Reimagined security

+

Unparalleled observability and operations

Robust and flexible infrastructure

AI demands a new class of infrastructure — one that delivers massive scale, ultra-low latency, and seamless integration across compute, networking, and storage. Cisco delivers exactly that, with flexible options to deploy as full-stack systems or individually as data center building blocks.

Building blocks: Networking, silicon, and compute

- With Nexus fabric, you can build automated, high-performance Ethernet fabrics that support both traditional workloads and AI networking for the most demanding AI workloads in frontend and backend networks.
- At the core is Cisco Silicon One — the industry leading unified network silicon architecture — driving breakthrough performance, operational efficiency, and simplified management from edge to core.
- Nexus Hyperfabric, our cloud-managed data center fabric, simplifies operations and accelerates scale.
- Our Unified Compute System (UCS) revolutionizes data center architecture by tightly integrating compute, networking, and storage into a single, agile platform.

In addition to these data center building blocks, Cisco can help you build a Secure AI Factory with fully integrated and pre-validated, full-stack systems that are purpose-built for AI workloads and accelerate the development and deployment of AI applications.

Full stack systems

AI PODs

Cisco AI PODs simplify AI deployment and scale with business needs. AI PODs combine high-performance UCS-X compute, Red Hat OpenShift for Kubernetes orchestration, and NVIDIA AI Enterprise for optimized AI workflows — while embedding Cisco security throughout the stack to protect data, models, and infrastructure. This integrated approach simplifies deployment, accelerates AI adoption, and ensures consistent results.

Nexus Hyperfabric AI

Additionally, Nexus Hyperfabric AI is a cloud-managed, full-stack solution that simplifies deployment and streamlines management of AI infrastructure at scale. It combines Cisco networking and compute with NVIDIA GPUs, DPUs, and VAST storage — delivering AI-scale performance with 800G.

Reimagined security

As workloads become increasingly distributed across hybrid environments, consistent security enforcement can be elusive. Our security portfolio is not only comprehensive, but tightly integrated, which means you get better threat detection, consistent policy enforcement, and simplified management. With Cisco's Hybrid Mesh Firewall, you can secure from on-prem environments to the cloud with enforcement points built into every layer of the stack – from infrastructure, workloads, AI models, and more.

And you can manage policies across the stack with ease using Security Cloud Control to define a policy once and enforce everywhere.

And we're delivering on unique innovations in data center security with Hypershield and AI Defense. Hypershield delivers AI-native security for data centers, building a mesh of distributed enforcement points. It offers deep visibility and fine-grained enforcement at workload, network, and infrastructure layers.

AI Defense offers a groundbreaking way to protect the development of AI-enabled applications as they are built. Using a novel algorithmic red teaming approach, AI Defense validates the AI model on a continuous basis to check if the model is behaving the way it should. Whereas most organizations need 7-10 weeks to validate a model, AI Defense can do it in minutes.

Unparalleled observability and operations

Beyond security, Cisco gives you visibility and insights across your entire environment to drive digital resilience. To ensure seamless connectivity, ThousandEyes helps you see across owned and unowned networks, which is critical since the majority of publicly reported outages stem from networks you don't own. Splunk Observability Cloud extends this visibility to applications and infrastructure, speeding detection of performance problems and issue resolution. Nexus Dashboard and Intersight empower IT teams to manage traditional and AI workloads easily with unified management. Teams can streamline operations across compute and network environments, driving faster issue resolution and improved reliability.

Driving outcomes that matter across the business

IT leaders today must move faster, scale AI, and manage complex hybrid environments. Cisco uniquely helps organizations meet these demands by unifying networking, compute, security, and observability. This platform approach compounds value over time — delivering more value and less complexity with every new investment.

Cisco enables these four essential outcomes:

Faster deployment

We accelerate time to value and deployment with the critical infrastructure for the most demanding environments, pre-validated designs, and deep expertise.

Faced with rapid growth and scaling challenges, Workday's data centers were hitting capacity limits and its network couldn't keep pace. By deploying Cisco solutions, Workday cut network deployment time by 83% — unlocking the speed and scalability needed to support expansion.[3]
"We had a great experience with Cisco Nexus switches in our legacy data centers, and it became clear Cisco Nexus 9000 Series Switches were the best choice for our cloud fabric network. They give us exceptional density, capacity, and scalability as well as highly desirable line rate encryption capabilities," said Ahsan Ghayas, senior principal engineer.

AI at scale

From data center to edge, Cisco provides the scalable, energy-efficient infrastructure needed for AI — built on Nexus networking, UCS compute, and Silicon One. Cisco IT's own experience proves the model. With urgency from the product team, Cisco IT needed to build an AI cluster to develop and test new AI capabilities for its extensive product portfolio, including model training and inferencing, within three months. The team deployed an AI-ready data center 80% faster than through traditional methods.[4]

Security everywhere

Security is embedded across the stack with enforcement points distributed everywhere from Cisco's Hybrid Mesh Firewall. With innovations like Hypershield, Cisco is able to fuse security into the data center fabric itself. For example, you can run Hypershield on Cisco Smart Switches to manage security policies on the switches and enable segmentation inline. This results in enhanced security, operational efficiency, and less hardware to manage. United Airlines has gained deep visibility, automated threat response, and microsegmentation across its hybrid multi-cloud environment with Cisco Secure Workload, part of Cisco's Hybrid Mesh Firewall.[5]

Complete visibility

With capabilities from ThousandEyes and Splunk integrated throughout the technology stack, organizations gain complete visibility across public and private clouds, owned and third-party networks, and digital services. This means you can detect and resolve issues faster, optimize performance, and maintain a clear line of sight across the entire environment.

Wintrust Financial, for example, leveraged Cisco ThousandEyes and Splunk to gain deep visibility, accelerate threat detection, and streamline incident response.

"ThousandEyes gave us a window into networks we don't own but still rely on every day. We were able to proactively escalate the issue before customers felt the pain," said Jermaine Mason, vice president of network infrastructure at Wintrust Financial.[6]

These benefits lead to better economics, cost and energy efficiencies, and better sustainability. Cisco's portfolio is built for efficiency and sustainability, with networking and compute solutions that optimize energy use, reduce footprint, and extend product life. The UCS X-Series, named Sustainable Product of the Year by the SEAL Awards, and the Nexus 9000 Series — recognized in the 2025 Gartner Magic Quadrant — help organizations cut costs and support sustainability goals with efficient power, cooling, and real-time energy insights.

The 3 steps to becoming AI-ready

1

Modernize

what you have — before scaling for what's next

Even if AI isn't yet at the center of your strategy, now is the time to modernize and prepare. A strategic refresh allows you to update legacy systems, secure your hybrid and multicloud environments, enhance network performance, and lay the foundation for what's coming next. Focus on building an integrated stack that works together. And don't overlook power and thermal efficiency — early design choices here can significantly reduce long-term costs and complexity as your infrastructure scales to support AI.

2

Optimize

your environment for AI workloads

As you begin deploying AI workloads, visibility, resilience, and integrated security across your environment become critical. Optimization isn't just about adding compute — it's about building a modular, GPU-optimized foundation that can scale easily and securely. With AI comes new complexity, and without full visibility into workloads and infrastructure, performance bottlenecks and operational risks can quickly emerge. This is the stage to embed telemetry, automate performance monitoring, and ensure security is woven into every layer, from infrastructure through to your AI models. Laying this groundwork now ensures your AI deployments remain resilient, efficient, and protected as they grow.

3

Accelerate

with confidence when it's time to scale

Once AI moves from pilot to production, speed becomes a competitive differentiator — and infrastructure must be ready to respond. CIOs should prepare to scale with high-performance GPU-accelerated compute, low-latency network fabrics, and bandwidth-rich connectivity that can support advanced use cases like GenAI and large language models. At this stage, operational agility depends on complete visibility across the AI lifecycle: from workload scheduling and GPU utilization to model performance and business impact. With the right infrastructure in place, you can move fast — without losing control — while enabling teams to innovate with confidence and deliver measurable value from AI.

Resources

1. Gartner: “AI Pricing: Strategies to Prevent Escalating Costs”

2. CIO.com: “CIOs face mounting pressure as AI costs and complexities threaten enterprise value”

3. Cisco: “Revamping a global network architecture.”

4. Cisco: “Cisco deploys AI-ready data center quickly, scales for the future.”

5. Cisco: “United Airlines takes flight with Cisco: building a foundation for digital resilience”

6. BizTech: “Cisco Live! 2025: How Wintrust Financial Gained Better Visibility Into Their Data.”

The future won't wait. Get AI-ready now.

The AI era demands more than incremental upgrades – it requires a bold rethinking of your data center strategy. With Cisco's platform approach, you can modernize with confidence, scale securely, and unlock the full value of AI – today and into the future.

Learn how you can reimagine your data center for the AI era.