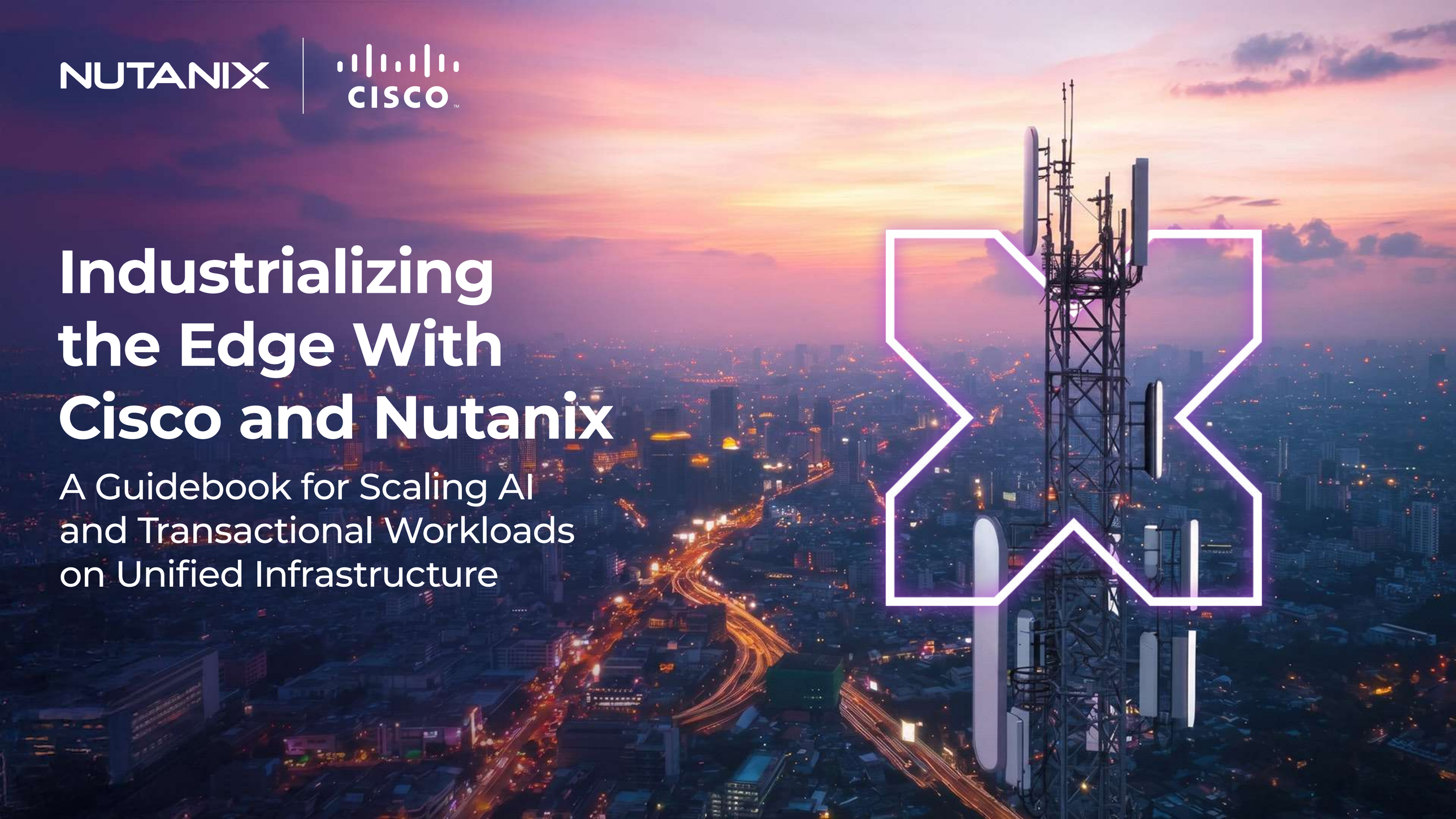
NUTANIX | cisco

Industrializing the Edge With Cisco and Nutanix

A Guidebook for Scaling AI and Transactional Workloads on Unified Infrastructure



Scaling Diverse Workloads at the Edge With Consistency

AI is moving to the edge, driven by sovereignty, latency, and cost constraints. Yet most organizations lack the infrastructure to scale beyond pilots into production. **Cisco Unified Edge with Nutanix** delivers a unified platform that integrates compute, networking, virtualization, and security into a single system. **Nutanix Agentic AI** extends this foundation with the software and services needed to build, run, and govern AI agents anywhere, at scale. Together, they enable consistent deployment, automated operations, and embedded governance across distributed environments, from retail locations to factory floors.

What This eBook Explores

- **AI scaling depends more on operational infrastructure than model capability**
- **Unified, automated platforms are required to manage distributed edge environments at scale**
- **Security and governance must be embedded from hardware through software**

Table of Contents

The Gravity of Modern Data

Ready or not, AI is moving to the edge

AI is no longer a centralized discipline. It is being pulled to the edge by forces that organizations cannot ignore.

First, **data sovereignty** is redefining where AI can operate. In fact, 77% of organizations say the location of AI development is now a key factor in technology decisions,[1] signaling a shift toward localized control.

Second, **latency** is now a business constraint. Real-time use cases – from fraud detection to clinical diagnostics – cannot tolerate delays introduced by distant data centers.

Third, **bandwidth and cost** make centralized processing impractical. Data volumes continue to surge, while moving that data remains expensive and inefficient.

Together, these forces are reshaping AI architecture.

[1] Deloitte, State of AI in the Enterprise, January 2026

The Implication Is Clear

Organizations must bring compute to the data, transforming the edge into a primary execution environment, not a secondary extension of the core.

The Readiness Gap and Challenges

Why scaling AI across the edge remains difficult

While AI is rapidly moving to the edge, most organizations are not equipped to run AI and transactional workloads there at scale.

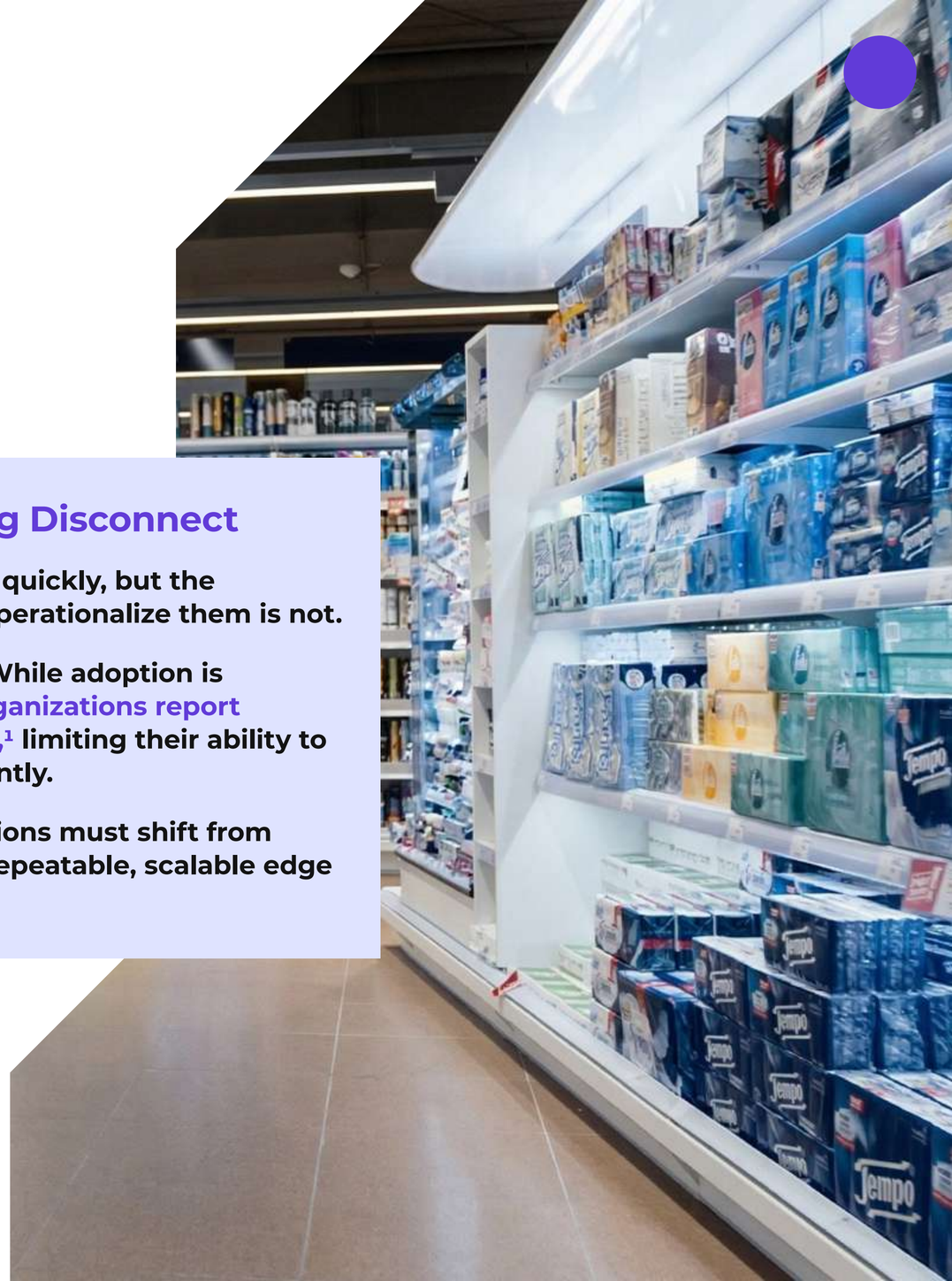
Edge environments introduce a fundamentally different operating model: thousands of distributed sites, limited on-site expertise, constrained physical environments, and expanded cyber attack surfaces.

At the same time, infrastructure remains fragmented across compute, networking, virtualization, and security, each managed independently.

This complexity slows deployment, creates configuration drift, and makes it difficult to maintain consistency across locations.

The Result Is a Growing Disconnect

- **AI strategies are advancing quickly, but the infrastructure required to operationalize them is not.**
- **This gap is already visible. While adoption is accelerating, only 21% of organizations report mature governance models,[1] limiting their ability to scale AI safely and consistently.**
- **To move forward, organizations must shift from isolated deployments to a repeatable, scalable edge operating model.**



[1] Deloitte, State of AI in the Enterprise, January 2026

The ‘No Compromise’ Edge

A unified platform for AI and transactional workloads

Closing the readiness gap requires a new approach to edge infrastructure – one that eliminates tradeoffs between performance, simplicity, and security.

Cisco Unified Edge with Nutanix delivers a **purpose-built, unified platform** designed specifically for distributed environments. By converging compute, networking, storage, virtualization, and security into a modular system, it enables organizations to run both **AI inference and traditional workloads** on the same infrastructure, without added complexity.

This is not a collection of integrated components. It is a validated, engineered platform designed to operate consistently across locations.

With zero-touch deployment, centralized management, and built-in lifecycle automation, organizations can scale from a single site to thousands, while maintaining control, visibility, and performance.

The Result Is a 'No Compromise' Edge

A platform that delivers the consistency of the core with the responsiveness of the edge, ready for both today’s workloads and tomorrow’s intelligent systems.

The Complete Edge Solution

A unified platform for AI and transactional workloads

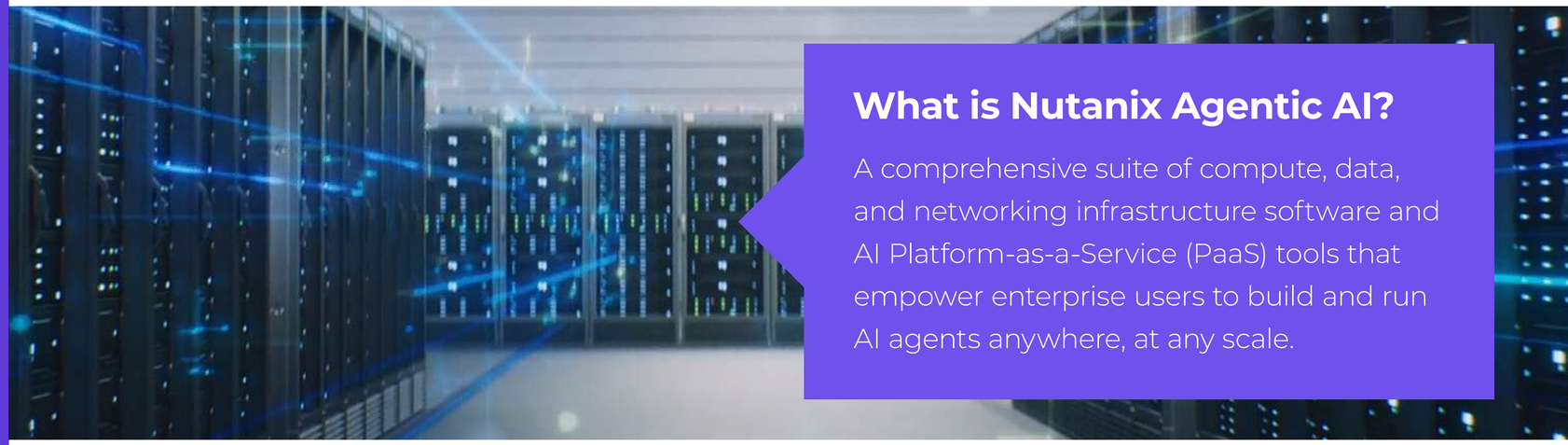
Delivering a “no compromise” edge requires both convergence and full-stack integration.

Cisco Unified Edge provides the modular foundation for compute, storage, and high-speed networking, while Nutanix delivers a unified software platform across virtualization, containers, cluster management, and Agentic AI capabilities, including **Nutanix Kubernetes Platform** (NKP), **Nutanix AHV** hypervisor, and **Nutanix AI Enterprise** (NAI).

Together, they create a single platform capable of running both traditional applications and next-generation AI-driven workloads at the edge.

Deployment and operations are unified through **Cisco Intersight** and **Nutanix Prism Central**, enabling zero-touch provisioning, centralized visibility, and coordinated lifecycle management across distributed environments.

This integration ensures infrastructure and software operate as one system, eliminating silos, reducing operational overhead, and maintaining consistency across sites.



What is Nutanix Agentic AI?

A comprehensive suite of compute, data, and networking infrastructure software and AI Platform-as-a-Service (PaaS) tools that empower enterprise users to build and run AI agents anywhere, at any scale.

The result is a complete edge platform: engineered for performance, simplified for operations, and designed to scale AI and transactional workloads seamlessly across the enterprise.

The Complete Edge Solution

Purpose-built infrastructure for distributed AI environments

Cisco edge infrastructure delivers performance, resiliency, and connectivity for AI and distributed enterprise workloads.

Designed for space-constrained and operationally demanding environments, Cisco's modular edge hardware enables consistent deployment, secure connectivity, and scalable performance across retail locations, factory floors, branch sites, and remote operations.

The Complete Edge Solution

One Platform. One AI Team.

Advancing AI requires coordination across teams with different responsibilities and technical requirements – from infrastructure operations and governance to AI services and application development. Nutanix Agentic AI brings these functions together through a consistent platform for building, running, and scaling AI across edge, core, and cloud environments.

Platform Eng

Infra Admin

Autonomous Agents

Agentic Apps

Physical AI

Models
NVIDIA, Llama, OpenAI, Anthropic, Google, Mistral, Qwen, etc.

AI Services and Kubernetes Platform
Nutanix Kubernetes Platform, AI Catalog, Nutanix Enterprise AI
Nutanix Database Service, Nutanix Data Hub

AWS

Azure

Google

Infrastructure Optimization and Security
AHV Topology Aware Hypervisor,
Flow Virtual Networking and Security

CPUs, GPUs, DPUs, Networks
NVIDIA, AMD, Intel

Data Infra
Nutanix Unified Storage

Tenant and AI Management
Nutanix Central, Service Provider Central
Nutanix Cloud Manager

AI Engineer

Infra Admin

Business Users

Deploying, Orchestrating, and Governing

Day 0: Deploy at Scale

From first deployment to continuous control

Scaling edge AI requires more than infrastructure. It requires a repeatable operational model that works from day one and holds over time.

Day 0: Deploy With Speed and Consistency

Edge deployments cannot rely on manual processes. **Cisco Intersight**, integrated with **Nutanix Foundation Central**, enables zero-touch provisioning of infrastructure and clusters across distributed locations

Systems can be claimed, configured, and deployed remotely. This eliminates staging complexity and reduces time-to-value.

At the same time, policy-based controls establish a foundation for governance from the start, ensuring consistency across every site.

Deploying, Orchestrating, and Governing

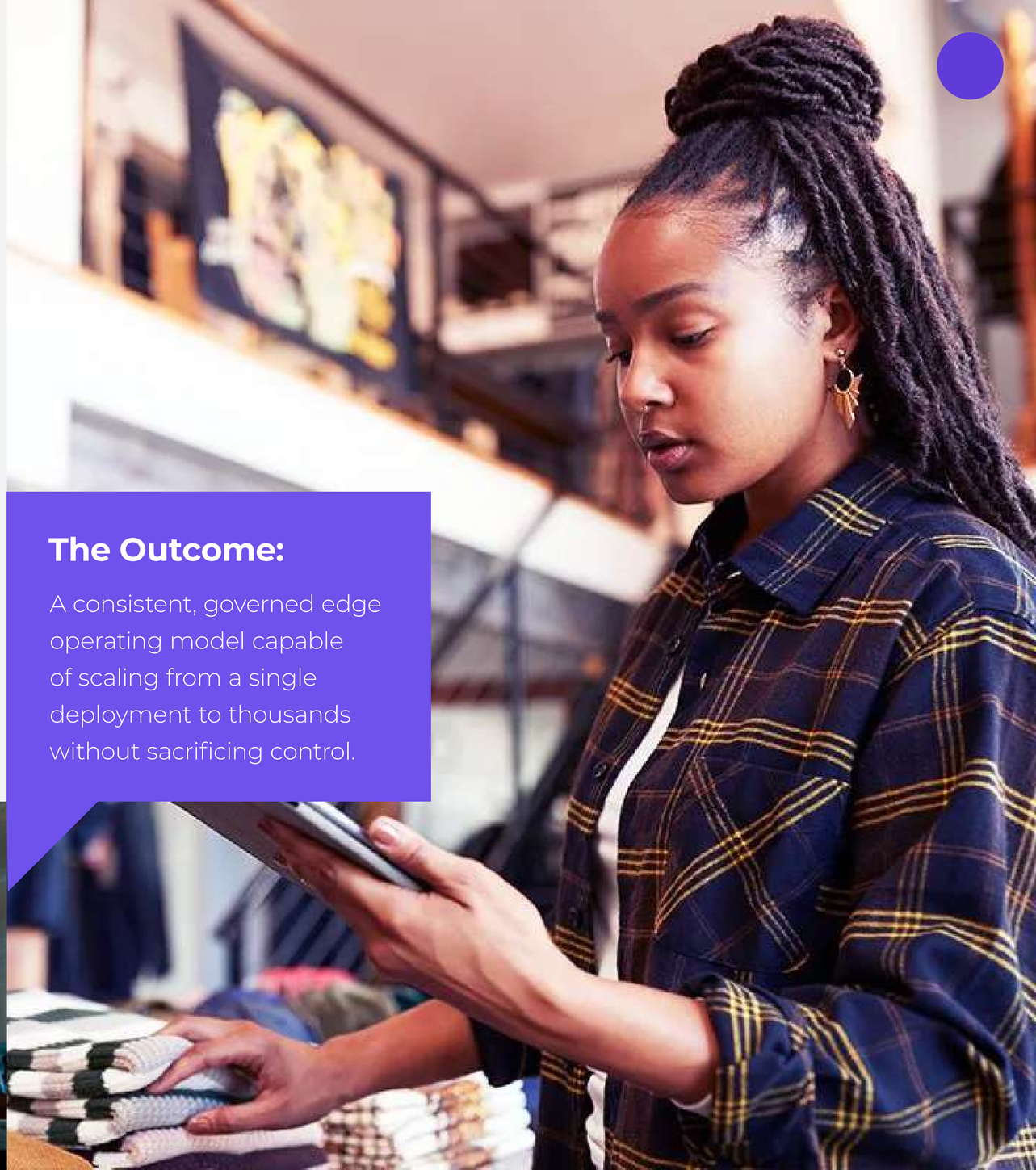
Day 1–2: Operate, Govern, and Scale

Cisco Intersight and **Nutanix Prism Central** provide unified visibility across infrastructure and workloads, enabling real-time monitoring, alerting, and performance management.

Coordinated lifecycle management ensures firmware and software updates remain aligned, eliminating drift and maintaining system integrity across sites.

This is critical as organizations scale. With only **21% of companies reporting mature governance models,**[1] most lack the operational discipline required to manage AI at scale.

[1] Deloitte, State of AI in the Enterprise, January 2026



The Outcome:

A consistent, governed edge operating model capable of scaling from a single deployment to thousands without sacrificing control.

Security Rooted in Silicon

Built-in protection across every layer

As AI moves to the edge, the attack surface expands. Distributed environments, physical access risks, and inconsistent configurations make security harder to enforce – and more critical to get right.

Cisco Unified Edge with Nutanix addresses this by embedding security directly into the platform, from hardware to hypervisor.

At the **hardware level**, root of trust and secure boot ensure system integrity from the moment of power-on, while tamper detection provides visibility into physical intrusion attempts.

At the **platform level**, **Nutanix AHV** delivers secure virtualization, workload isolation, and cluster resilience.

Together, these capabilities create a consistent security posture across sites, reinforced through lifecycle management that keeps systems patched, aligned, and compliant.

Security is never bolted on.
It’s built into every layer of the edge.

Edge Use Cases in Action

Where unified edge infrastructure delivers impact

The value of a unified edge platform is realized through the workloads it enables, combining real-time data processing with AI-driven insight across industries.

In **retail**, organizations can analyze in-store activity, optimize inventory, and deliver personalized experiences at the point of engagement.

In **manufacturing**, sensor data and computer vision systems enable predictive maintenance, quality inspection, and operational efficiency directly on the factory floor.

In **healthcare**, localized AI supports diagnostics, imaging analysis, and patient monitoring, where latency and data control are critical.

In **financial services**, edge-based systems help detect fraud, process transactions, and manage risk in real time.

Across these environments, the requirement is the same: run AI and transactional workloads together – securely, consistently, and at scale – where data is created and decisions must be made.

Assessing Your Edge Readiness

A practical path from pilot to scale

For many organizations, the challenge is no longer whether to deploy AI at the edge, but how to do so consistently at scale.

The first step is assessing readiness across three dimensions:

Infrastructure:
Can your environment support distributed workloads with consistent performance and lifecycle management?

Operations:
Do you have a unified model for deployment, visibility, and control across sites?

Governance and Security:
Are policies, compliance, and protections embedded from the start or applied after the fact?

This matters now. While only a portion of organizations have scaled AI into production, **54% expect to do so within the next 3–6 months.**[1] This accelerates the need for a scalable foundation.

Organizations that align these elements can move beyond isolated deployments and establish a repeatable model for edge AI.

[1] Deloitte, State of AI in the Enterprise, January 2026

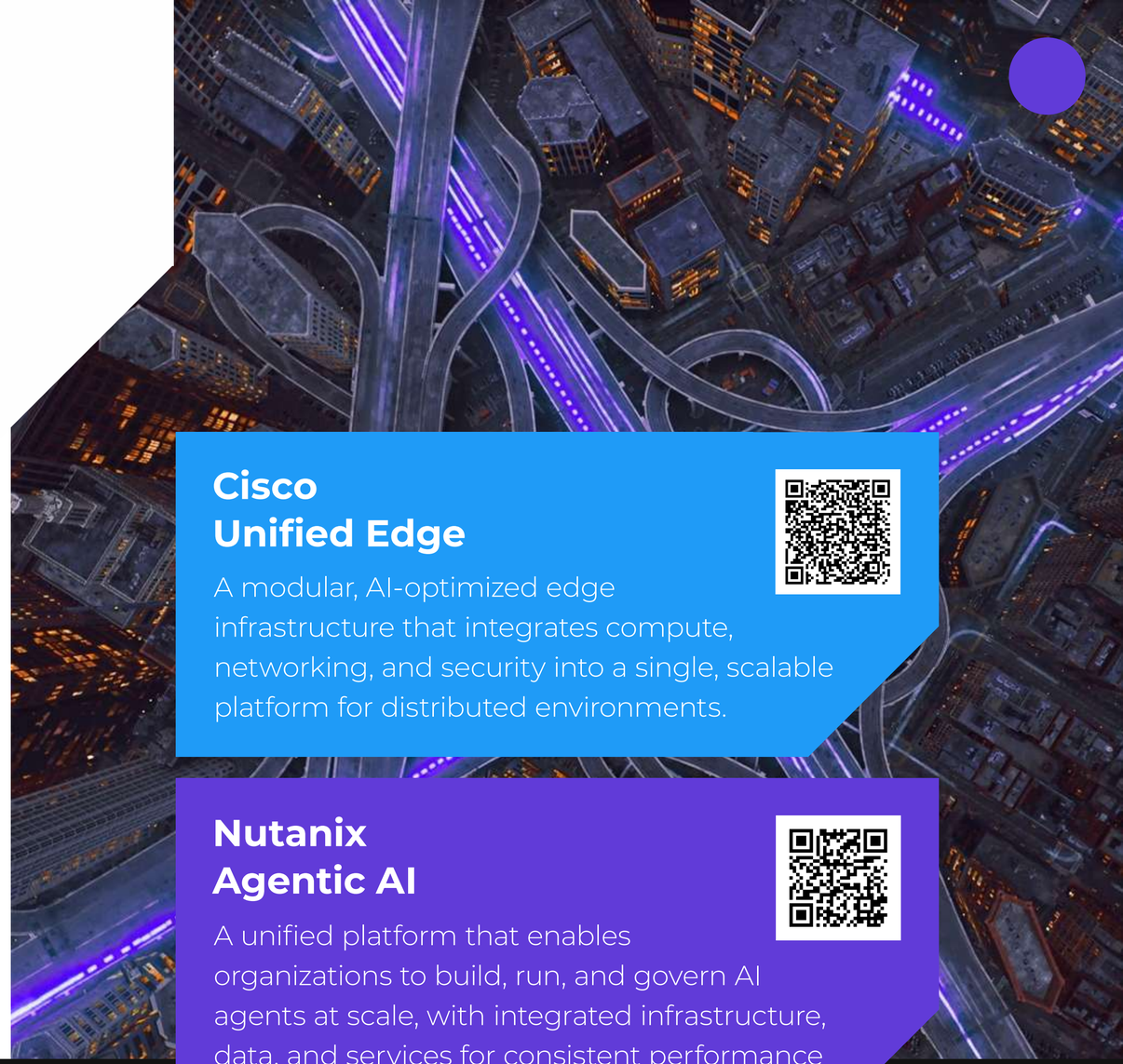
Conclusion: The Edge, Unified

From edge complexity to scalable execution

AI is shifting to the edge, where data is created and decisions happen. Organizations that unify infrastructure, operations, and security can scale AI consistently across distributed environments, transforming the edge into a reliable, governed execution layer for both AI and transactional workloads.

Key Takeaways

- **Edge is now a key part of enterprise AI architecture**
- **Scaling AI is an infrastructure and operations challenge**
- **Standardization is the foundation for repeatability**
- **Governance and security determine long-term viability**



Cisco Unified Edge

A modular, AI-optimized edge infrastructure that integrates compute, networking, and security into a single, scalable platform for distributed environments.

Nutanix Agentic AI

A unified platform that enables organizations to build, run, and govern AI agents at scale, with integrated infrastructure, data, and services for consistent performance across edge and core environments.

NUTANIX

info@nutanix.com | www.nutanix.com | @nutanix