The Lean AI-Ready Data Center

A Phased Blueprint for Modernizing the Data Center Network

For IT and Business Leaders of Growing Organizations

How to read this paper: This blueprint is architecture-first and vendor-neutral. Each phase opens with the capability you need and what to look for in any solution, then a clearly marked Cisco recommendation for teams who want a single integrated stack. If you run another vendor's data center networking today, the phases still apply: adopt the operating model, keep what works, and treat the Cisco sections as one fully integrated way to deliver it. The paper is also deliberately front-loaded. Phases 1 through 3 deliver a complete, secure, modern data center for a single site. Phase 4 extends the same model to additional locations, for organizations that have them. Phases 5 and 6 are optional and are intended for organizations that are ready to adopt AI and are planning to implement AI in the near future.

Executive Summary

Operationalizing AI is now one of the highest priorities on the CIO agenda, and it is not the same thing as having an AI strategy. Gartner's 2026 research separates the two: developing an AI strategy is about choosing direction and use cases, while operationalizing AI is about embedding it into workflows and scaling it into production. Operationalizing AI moved to the number-two CIO priority for 2026, behind cybersecurity and risk management. The distinction matters for infrastructure leaders, because your job is not to pick the use cases. It is to make sure the foundation can carry them.

Most AI initiatives never get that far. Gartner finds AI investment accelerating while a majority of initiatives still fail to reach production. The reason is rarely the model. It is the foundation underneath it: the data, the governance, and the network the workloads actually run on. For a growing organization the exposure is sharpest, because you carry enterprise expectations on the leanest IT team of any segment.

Here is the part that is easy to get wrong. For most growing companies, AI runs in software-as-a-service and the cloud, and it will stay there. A speculative on-premises GPU build is gated by power, cooling, and economics, and is the wrong move for the majority. But the opposite mistake is just as costly: a like-for-like network refresh that quietly forecloses AI-readiness and forces a disruptive rebuild later. The way through is neither. You are refreshing the data center network anyway. Done once, on a modern cloud-managed fabric, that refresh makes the data center AI-ready by default, pays for itself on entirely non-AI merits, and keeps the on-premises option open for the day a workload earns it.

The architecture that makes this work is straightforward: an AI-ready data center is defined by its network foundation, not the compute accelerators themselves. The fabric is what decides whether compute is reliable, whether AI in production behaves, and whether on-premises is even an option later. And the fabric is the part a lean team cannot hand-build. Make it cloud-managed and validated, and readiness stops being a project and becomes a property.

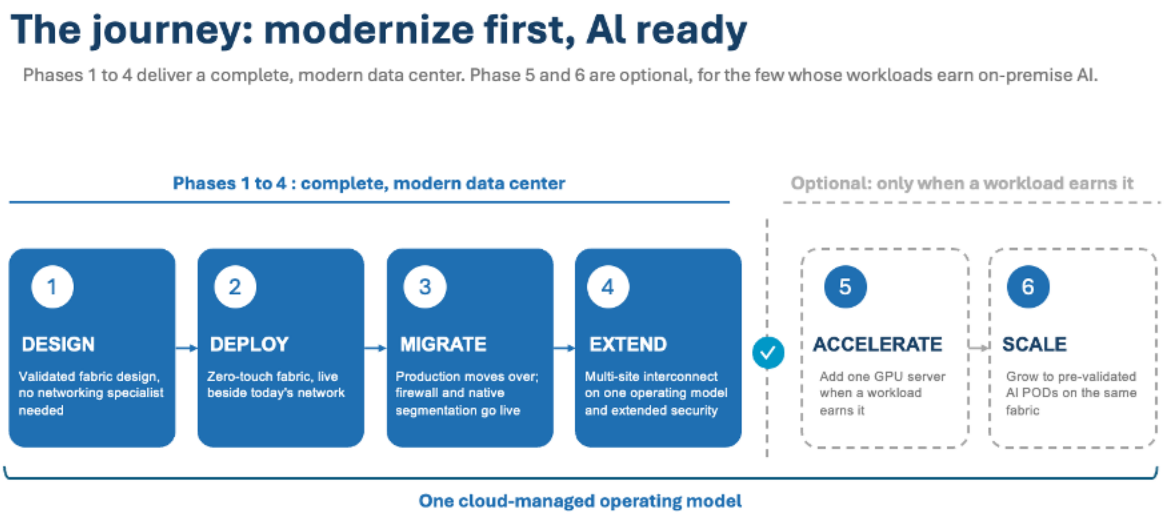
The blueprint runs in six phases. The first three are the destination for most readers, with the fourth for those who run more than one site:

- **Design.** Produce a validated, AI-ready fabric design with IT generalists on staff.
- **Deploy.** Stand up the first modern fabric beside what you run today, with zero-touch provisioning and no disruption.
- **Migrate.** Move production onto the modern fabric, secured at cutover with a perimeter firewall, segmentation, and detection and response, and retire legacy on your schedule. For a single site, the modern data center is complete here.
- **Extend.** For organizations with more than one location or edge locations, run the same operating model and the same security across every site.

The last two are optional, for those whose workloads earn on-premises AI:

- **Accelerate (optional).** When a specific workload earns it, add on-premises compute to the fabric you already built, starting with a single server.
- **Scale (optional).** For the few whose AI demand outgrows one node, pre-validated AI infrastructure on the same fabric and operating model.

The core idea in one line: **the fabric is the foundation. Get the network right, and the data center is ready for whatever comes next, AI included.**



1. The Priority: Operationalizing AI, Not Strategizing About It

CIOs have moved past the question of whether to invest in AI. The 2026 Gartner CIO and Technology Executive Survey and Leadership Perspective Survey both show the agenda shifting from experimentation to execution. AI funding is among the fastest-rising line items, productivity and cost optimization are the business outcomes leaders are measured on, and “operationalizing AI” has risen to the second priority overall, just behind cybersecurity and risk management, which has held the top spot for three years running.

Operationalizing is the operative word. Gartner deliberately split it out from “developing an AI strategy” because they are different jobs. Strategy is choosing where AI creates value. Operationalizing is making it work, repeatedly, in production, at acceptable cost and risk. The gap between the two is where value leaks: AI investment is accelerating, yet a majority of initiatives still fail to reach production. When they stall, the cause is usually foundational rather than conceptual. The model was fine. The data, the governance, or the infrastructure underneath it was not ready. For a growing organization, this lands hard. You are expected to operationalize at the pace of a large enterprise, with a fraction of the staff to design, run, and secure the infrastructure underneath. That is the opportunity and the exposure in one sentence.

Why the obvious moves fail. Two common approaches often prove counterproductive. The first is to treat AI-readiness as a compute project and buy GPUs. For most growing companies that is premature: the AI they operationalize runs in SaaS and the cloud, and an on-premises GPU build is gated by power, cooling, and economics that rarely pencil out at this scale. The second reflex is the quiet one: refresh the network like for like, because “we are not doing AI in our data center.” That saves money today and forecloses the option tomorrow, turning the next AI decision into a forklift. Neither reflex addresses the real task, which is to modernize the foundation once, in a way that delivers value immediately and keeps the future open. For instance, organizations that over-provision compute before optimizing the network fabric often see GPU efficiency drop significantly due to I/O bottlenecks.

2. The Core Insight: An AI-Ready Data Center Starts with the Network

GPUs get the attention and the budget line. The network plays a critical role in whether GPUs ever pay off. AI workloads move enormous volumes of traffic between compute nodes, and that traffic is unforgiving: the fabric is only as fast as its slowest component, and a single congested link can leave

expensive accelerators idling. Industry and analyst assessments of AI data centers converge on the same point. Networking is the silent backbone of AI, and it is where performance is won or lost. The same is true, less dramatically, for AI that runs in the cloud. Operationalizing AI puts new load on the data center even when the model lives elsewhere: more data movement, tighter integration between internal systems and AI services, and agents that reach into applications and expect them to respond. A fragmented, manually-run, aging network is itself a bottleneck to AI in production, and it is invisible until something stalls.

Here is what a compute-first framing misses. AI-readiness is often sold as a property of the server. In reality, it is a property of the combined compute and network architecture, and for a lean team it is specifically a property of a cloud-managed network, because the alternative requires data-center networking expertise the team does not have and cannot easily hire. So, the readiness question is not "which GPU." It is "can we operate the fabric." That reframes the category around a cloud-managed, validated fabric, which is also why readiness is cheap to buy now and expensive to retrofit. Choose a fabric that scales from a single switch pair to a full cluster, and the refresh you are already doing carries the option forward at no extra cost.

3. The Blueprint

Sequence is the principle. Modernize the network first, because a dependable, automated fabric is what makes everything above it possible. Three properties make this a blueprint rather than a wish list.

- **Each phase stands alone.** Stop after Phase 1 and you still gained a validated design and a path off manual operations. Stop after Phase 3 and you have a complete, secure, modern data center for a single site.
- **Each phase has exit criteria.** You know when the next phase is earned, and when it is not.
- **Nothing is thrown away.** Fabric, operating model, security, and assurance investments compound rather than reset.

A fourth property is specific to this paper: the value is front-loaded. A single-site data center is complete and secure at Phase 3. Phase 4 extends the same model to additional locations for organizations that have them. Phases 5 and 6 exist for those whose workloads genuinely earn on-premises AI, and entering them is a business decision, not a technical inevitability. It is also a growth path. The same fabric that delivers simplicity at 300 people carries the organization into mid-market and, if the business ever demands it, into enterprise-scale AI, without a change of operating model.

The spine: **Design** → **Deploy** → **Migrate** → **Extend**, then optionally **Accelerate** → **Scale.**

What to look for: one cloud-managed operating model.
Whatever products deliver the phases, the operational win comes from running them from as few consoles as possible, with as little specialized networking expertise as possible. A lean team cannot absorb a new data center across several dashboards and a new skill set. Look for cloud-managed design, deployment, and lifecycle automation, validated designs that remove the need for deep networking expertise, and unified operations that span the fabric, its performance, and eventually the compute attached to it.

The Cisco recommendation: Cisco Nexus Hyperfabric.
Cisco Nexus Hyperfabric is its cloud-managed delivery: a fabric-as-a-service in which a Cisco-operated cloud controller designs, deploys, monitors, and scales fabrics, with the switches managed entirely from the cloud rather than a command line. The experience is closer to cloud-managed networking than to traditional data center engineering, which is precisely why a lean team can run it. Cisco Nexus Hyperfabric is one domain within Cisco Cloud Control[1], the unified operating platform that sits above the individual domain controllers and brings them under one interface. For this

blueprint the domains that matter are networking (Hyperfabric), security (Security Cloud Control), and, if you ever add on-premises compute, Intersight. The point is that the fabric, its security, and its compute are operated as one estate rather than three islands.

You likely already run this operating model.
There is a good chance your team already knows how Cisco Nexus Hyperfabric works, because Cisco Meraki made cloud-managed, zero-touch, subscription networking the norm in the campus and branch for exactly these teams. Cisco Nexus Hyperfabric brings that same operating model to the data center. If you run Meraki today, the dashboard philosophy, the plug-and-play onboarding, and the subscription model will be familiar, and both Meraki and Hyperfabric sit under the same Cisco Cloud Control umbrella, so the data center joins the way you already operate rather than adding a separate world. Important distinctions regarding the operating model: This approach leverages the operational philosophy of Meraki, rather than a direct product-line extension. Cisco Nexus Hyperfabric delivers the cloud-managed experience on dedicated data center hardware. And Cisco Nexus Hyperfabric is a newer platform than Meraki, so treat the comparison as a guide to how it operates, not a claim to the same track record.

Security is designed in and matures with the fabric.
Security is not a phase bolted on at the end. It is sequenced the way a data center is actually secured. The management plane is hardened from the first switch. The moment production first rides the fabric, a north-south perimeter firewall, an east-west firewall, native network segmentation go live together. Security policy across these controls is expressed in Security Cloud Control, the security domain within Cisco Cloud Control, so it is operated alongside the network rather than in a separate world. Deeper east-west microsegmentation, identity-centric zero trust, and full security operations are the domain of the companion security blueprint; this paper embeds the security that lives in the network itself and hands off the rest cleanly rather than duplicating it.

4. Phase 1 — Design

The promise: A validated, AI-ready fabric design, with IT generalists on staff.
Business outcome: a target architecture you can trust, produced in hours, with the bill of materials right the first time.
The refresh starts on paper, not in the rack. A cloud controller takes what you actually know, your host and port counts, oversubscription, cabling, and power, and produces a validated fabric design. You decide where to start: a rack, a row, a new colocation footprint, an edge site or a disaster-recovery site. No rip-and-replace is committed. Two decisions are made cheaply here and expensively later. First, platform: most SMB and mid-market fabrics run at 1 to 100 Gb, while only high-density or AI-class fabrics need 400 or 800 Gb, so the design should pick the right speed band rather than over-buying. Second, the security edge: where the perimeter firewall sits at the fabric border, and how the network is segmented, belong in the topology from the start, not as a later addition. Choosing a fabric that scales from a single switch pair to a spine-leaf cluster means the refresh you do now does not box you in later, and it costs nothing extra to keep that option open at the design stage.

What to look for (any vendor)

- **Validated, templated fabric design** that takes capacity, oversubscription, cabling, and power as inputs and produces a known-good topology.
- **A platform and speed range matched to the workload**, so a typical 1-to-100 Gb data center is not forced onto AI-class hardware it does not need.[1]

[1] Cisco Cloud Control is the unified operating platform that brings Cisco domains under one interface as a layer above their existing controllers (it does not replace them). For this blueprint the domains are predominantly Nexus Hyperfabric (networking), Security Cloud Control (security), and Intersight (compute). Cloud Control is in Controlled Availability (US, 2026; global to follow);

- **The security edge designed in**: planned firewall placement at the border and a segmentation model, not a retrofit.[2]
- **A design range** that scales down to a couple of switches and up to a full spine-leaf, so the same model serves a modest data center, AI clusters at core data center and AI at edge sites.
- **Bill-of-materials generation** tied to the design, so the order is correct the first time without manual translation.
- **Low expertise required**, because the goal of this phase is a trustworthy design without a dedicated network architect.

The Cisco recommendation
Log in to Cisco Nexus Hyperfabric and build a validated design tailored to your desired host and port capacity, oversubscription, and environmental constraints such as cabling and power. Select the platform to match: the Cisco silicon-based N9300 and Cisco 6000 Series deliver a wide range of performance options, from mainstream speeds of SMB and mid-market (1-100G) to AI-scale (400-800G). The NVIDIA silicon-based N9100 Series is well suited for customers looking to build a fabric that follows the NVIDIA Reference Architecture and leverages NVIDIA Adaptive Routing technology. The controller is integrated with Cisco's ordering tools, so the design converts into a bill of materials without transcription errors, and you can begin designing at no cost before any commitment. The fabric is built on Ethernet VPN and VXLAN, the same standards-based foundation used in large data centers, so the design is sound at any size you grow into[2].

What you deploy
- **A Cisco Nexus Hyperfabric account** (cloud controller access) to design and generate the bill of materials. No hardware is committed in this phase.

Exit criteria
Phase 1 is complete when a validated fabric design exists, sized to your starting footprint, with a confirmed bill of materials and a chosen location to deploy first.

5. Phase 2 — Deploy

The promise: Your first modern fabric, live beside what you run today, in days rather than months.
Business outcome: a working cloud-managed fabric with no disruption to the existing network.

Hardware arrives, connects to the internet, and is claimed and provisioned by the cloud controller with a zero-touch, plug-and-play approach. The process takes minutes and results in a working fabric. You can start as small as a single switch or a switch pair. Critically, the existing network is untouched. The new fabric is a greenfield island standing beside what you run today. This is the answer to the most common worry about a data center refresh: no, you do not cut over everything at once. You stand up the new fabric, prove it, and only then begin to move onto it.

What to look for (any vendor)
- **Zero-touch provisioning** from the cloud, so switches come online without manual, per-device configuration.
- **A small starting point**, a single switch or pair, with a clean path to grow, so the first step is low-risk.

[2]Cisco Nexus Hyperfabric runs on the selected Cisco N9300 switches and Cisco 6000 Series switches. N9300 provides a brownfield path onto the same cloud-managed fabric. The Cisco N9100 Series is well suited for customers looking to build a fabric that follows the NVIDIA Reference Architecture and leverages NVIDIA Adaptive Routing technology.

- **No disruption to the existing network**, because the new fabric runs alongside it rather than replacing it on day one.
- **Fast time-to-operational**, since the goal of this phase is a visible, working result quickly.

The Cisco recommendation
Deploy cloud-managed Nexus Hyperfabric switches sized to your fabric. For most SMB and mid-market data centers, the Cisco silicon-based N9300 and Cisco 6000 Series deliver a wide range of performance options, from mainstream speeds of SMB and mid-market (1-100G) to AI-scale (400-800G). The NVIDIA silicon-based N9100 Series is well suited for customers looking to build a fabric that follows the NVIDIA Reference Architecture and leverages NVIDIA Adaptive Routing technology. Either way, the switches arrive, tether to the cloud controller over the internet, and are provisioned into a working fabric in minutes. They are sealed appliances by design: a locked, validated image with a hardware root of trust, manufacturer certificates for hardware-based attestation, software-integrity verification, and a secure tether to the cloud, managed entirely from the cloud rather than a command line. That is a deliberate trade, and it is also the first security control in the blueprint. You give up box-by-box CLI access in exchange for a hardened, attested, cloud-driven management plane that a lean team can actually run.[3]

What you deploy
- The Cisco silicon-based **N9300 and Cisco 6000 Series** for mainstream speeds of SMB and mid-market (1-100G) to AI-scale (400-800G). The NVIDIA silicon-based N9100 Series is well suited for customers looking to build a fabric that follows the NVIDIA Reference Architecture and leverages NVIDIA Adaptive Routing technology. Size to your starting footprint, with a Cisco Nexus Hyperfabric subscription (Essentials, Advantage, Premier).

Included
- **Zero-touch provisioning, the cloud controller, a hardened and attested management plane, assertion-based monitoring, and cloud-delivered software updates**, within the Hyperfabric subscription.

Exit criteria
Phase 2 is complete when your first fabric is live and operational, provisioned from the cloud, with the existing network unaffected.

6. Phase 3 — Migrate

The promise: Production workloads on the modern fabric, legacy retired on your schedule.
Business outcome: a clean transition with no flag-day, and the legacy solution decommission when you are ready when you are ready.

With a working fabric beside the old one, migration is incremental and reversible. You interconnect the new fabric to the existing network, extend the relevant Layer-2 and Layer-3 networks across the boundary, move the first-hop gateway, and migrate workloads application by application. As each group moves and proves out, you decommission the legacy it left behind. This is the brownfield reality of every real data center, and it is a well-documented path rather than an improvisation. The full standalone value of the blueprint lands here: an automated, cloud-managed fabric running real production traffic, operated by a small team.

[3] [3]The Hyperfabric fabric switches run a locked, cloud-managed image and do not host in-switch security enforcement (for example, Hypershield). North-south and East-west protection is delivered by inserted firewalls. Hyperfabric provides native network segmentation but it does not provide native microsegmentation today; that capability is the domain of the companion security blueprint.

This is also the moment the data center is secured, because it is the moment production first rides the fabric. A north-south perimeter firewall sits at the fabric border, governing everything entering and leaving the data center. An east-west firewall sits within the fabric for the traffic among endpoints within the data center. These enable advanced threat protection across perimeters and within the fabric at scale without compromising performance. With these in place, a single-site organization now has a complete, secure, modern data center. Phase 4 is only for those who run more than one location.

What to look for (any vendor)

- **A documented brownfield interconnect**: Layer-2 extension and Layer-3 external connectivity between the new fabric and the existing network.
- **Gateway migration** that lets you move the default gateway with minimal disruption.
- **A north-south perimeter and an east-west firewall**, in place before production rides the interconnect.
- **Native network segmentation**, so workloads are separated as they move onto the fabric, without bolt-on appliances.

The Cisco recommendation

Use Nexus Hyperfabric's documented migration guidance to bridge the new fabric to your existing network. Native Logical Networks carry your VLANs across as VXLAN segments. At the border, place a Cisco Secure Firewall (Threat Defense) for north-south protection. Place another set of Cisco Secure Firewall within the fabric for east-west protection. Security policy for these controls is expressed in Cisco Security Cloud Control, the security domain within Cisco Cloud Control[4], so it is operated alongside the network rather than in a separate console.

What you deploy

- **Cisco Secure Firewall (Threat Defense)** at the fabric border for north-south protection and within the fabric for east-west protection.

Exit criteria

Phase 3 is complete when production workloads run on the new fabric, a north-south perimeter firewall is enforcing at the border, an east-west firewall within the fabric, native network segmentation is in place by the fabric itself, and legacy decommissioning is underway or scheduled. **For a single site, the modern data center is complete here.**

7. Phase 4 — Extend

For organizations with more than one location. A single-site data center is complete at Phase 3.

The promise: The same operating model and the same security across every location.

Business outcome: the same experience across multiple sites, run by the same lean team.

Many growing organizations run a single data center, and for them the work is done at Phase 3. Phase 4 is for those who have, or grow into, more than one location: a second site, a colocation footprint, a disaster-recovery facility, or edge locations. The principle is that another site should be a repeat, not a project. Multi-site interconnect joins fabrics across locations on one operating model, and the security you established in Phase 3, the firewalls and segmentation, extends to the new site rather than being rebuilt. Assurance is built into the fabric itself, so the experience stays provable across locations without adding agents or appliances.

What to look for (any vendor)

[4] Security Cloud Control is the security domain within Cisco Cloud Control, the same umbrella that unifies Nexus Hyperfabric (networking) and Intersight (compute). It is not a separate platform from Cloud Control; it is how security policy is expressed within it.

- **Multi-site interconnect on one operating model**, so additional locations inherit the same design and management.
- **Security that extends, not repeats**, so the firewall and segmentation extend across multiple sites.
- **Built-in assurance** that distinguishes your network from the circuit from the provider, without deploying separate monitoring agents.
- **One operational view** across sites, rather than a separate dashboard per location.
- **Low expertise required**, because of the difficulty of placing skilled technical staff at remote locations such as edge sites.

The Cisco recommendation
Extend with Nexus Hyperfabric multi-site interconnect, using border gateways and VXLAN multi-site to join fabrics across locations. The Phase 3 firewall and native segmentation, simply extend to the new site, with policy still expressed in Security Cloud Control within Cisco Cloud Control. Assurance does not need a new product: the fabric's own assertion-based monitoring, surfaced in the cloud controller, makes the path across sites provable end to end, so a slow application is answerable in minutes rather than a finger-pointing exercise between teams.

What you deploy
- **Additional Nexus Hyperfabric switches** (Nexus 9300 or Cisco 6000 Series) per site, on the same Hyperfabric subscription. No new security or monitoring products are required to cover the new location when those in the main site are being shared. Additional **Cisco Secure Firewall** in a Disaster Recovery site.

Exit criteria
Phase 4 is complete when more than one location runs on one operating model, with the extended firewall and segmentation across sites, and the whole estate visible from a single view.

8. Phase 5 — Accelerate

Optional. Enter only when a specific workload earns on-premises compute. Start with one server, not a cluster.
The promise: When a workload earns it, on-premises compute on the fabric you already built, starting with a single server.
Business outcome: lower cost or latency on the one workload that justifies it, with no new operating model to learn.
Most readers stop at Phase 4, and that is the expected outcome. Enter here only when a specific workload crosses one of three gates: a regulatory data-residency requirement, a sustained volume where cloud cost crosses over the cost of ownership, or a hard latency or edge constraint. Absent one of those, staying in SaaS and the cloud is the right answer, and this phase is a future option rather than a present need. When you do qualify, start small. A single GPU server, hosting one justified workload, attached to the fabric you already deployed. Be honest about the physics: on-premises AI is gated by power and cooling as much as by budget, so size to the workload in front of you, not to a speculative cluster.

What to look for (any vendor)
- **A right-sized, single-node entry point**, one GPU server for one justified workload, not a cluster you do not need.
- **Bring-your-own-AI flexibility**, so you can host the model or framework the workload requires.
- **Cloud-managed compute lifecycle** on the same lean-operations model for compute hardware operation.

- **AI-specific security**, model validation before deployment and runtime guardrails, carried over from your security posture.

The Cisco recommendation
Attach a Cisco UCS GPU server, managed by Cisco Intersight, to the fabric using Nexus Hyperfabric's bring-your-own-AI option. Cisco Intersight is itself a domain within Cisco Cloud Control, the same umbrella that already runs your fabric (Hyperfabric) and your security (Security Cloud Control), so adding compute does not add an operating model. It surfaces in the same unified platform you already use. Cisco AI Defense provides model validation before model deployment and runtime guardrails, extending the discipline you already apply elsewhere to the AI you now host. Because the fabric is already in place, this phase is genuinely additive: you are connecting a server, not building a data center.

What you deploy
- **One Cisco UCS GPU server** with a Cisco Intersight subscription (a Cisco Cloud Control domain).
-

Recommended / Optional
- **Cisco AI Defense** for model validation and runtime guardrails (recommended whenever you host a model).

Exit criteria
Phase 5 is complete when one on-premises workload runs on a right-sized node on the existing fabric, managed from the same operating model, with AI-specific security applied. There is no obligation to proceed to Phase 6.

9. Phase 6 — Scale

Optional. For those whose on-premises AI demand outgrows a single node.
The promise: When demand outgrows one server, pre-validated AI infrastructure that scales without a redesign.

Business outcome: predictable scale-up/out on the same fabric and operating model, from a pod to a factory if the business ever needs it.

For those whose on-premises AI workload sustains and grows, the path forward is pre-validated infrastructure rather than custom engineering. Move from a single node to a validated, full-stack configuration of compute, network, GPU, and storage, then to pre-built AI pods as workloads multiply. At the far enterprise edge, a full-scale AI factory pattern is available on the same architecture. The point of this phase is that scale does not mean a redesign. The fabric, the operating model, and the governance are the ones you already run.

What to look for (any vendor)
- **Pre-validated AI infrastructure** aligned to a recognized reference architecture, so design time does not balloon as you scale.
- **Automate** full-stack deployment of production-ready workloads at cloud speed.
- **One operating model from node to cluster**, so growth is additive rather than a rip-and-replace.
- **Security embedded at every rung**, carried from the earlier phases rather than bolted on at scale.

The Cisco recommendation

Adopt the Nexus Hyperfabric full-stack AI Infrastructure option, a cloud-managed configuration of Cisco UCS GPU servers and high-performance networking compliant with the NVIDIA Enterprise Reference Architecture, then Cisco AI PODs as workloads multiply, and at full scale the Cisco Secure AI Factory with NVIDIA. Cisco AI Defense rides every rung, validating models and guarding them at runtime, so the governance from Phase 5 extends across the cluster. This is where the blueprint meets the enterprise: an organization that reaches it has grown into enterprise-scale AI with its fabric, operating model, and security already in place.

What you deploy

- **The Hyperfabric full-stack AI Infrastructure option or Cisco AI PODs**, sized to demand.

Optional

- **Cisco Secure AI Factory with NVIDIA** at full scale.
- **Cisco Stack Automation** operationalizes full-stack AI environments by automating infrastructure configuration, AI tooling, software layers, security, and observability from rack to application.

Exit criteria

None. Phase 6 is not a destination the blueprint pushes toward. It is a capability the architecture makes available when, and only when, the business case arrives. Most organizations never enter it, and that is the expected and correct outcome.

10. Where to Start

The blueprint meets you where you are, on whatever infrastructure you run today.

- **Refreshing data center networking soon?** Start Phase 1 with the design, sized to whatever is up for renewal. One modern fabric island puts you on the operating model without a forklift.
- **Running a brownfield Nexus or other-vendor estate?** Use the interconnect path. Stand the new fabric up beside the old, keep the legacy gear during the transition, and migrate workloads on your own schedule.
- **Already on a modern fabric?** Focus on the security at Phase 3, the firewall and segmentation, and add Phase 4 only if you run more than one location.
- **Already run a firewall?** Keep it. The blueprint places the new fabric behind your existing security edge.
- **Already run Cisco Nexus 9300?** You may not need all-new hardware. Selected N9300 switches with an active subscription can be onboarded directly onto Hyperfabric, so existing gear joins the same cloud-managed fabric. Confirm your specific models against current Cisco documentation.
- **Single site, and likely to stay that way?** Then Phase 3 is your finish line. A complete, secure, modern data center sits at the end of Migrate; Phase 4 is there if you ever add a location.
- **No AI plans at all?** That is the common case, and the paper is built for it. Phases 1 through 3 deliver a complete, secure, lean-team data center on their own. Phases 5 and 6 wait until a workload earns them, if it ever does.
- **Running lean?** The blueprint assumes it. Lead with cloud-managed design and operations so a small team is not buried in consoles or blocked on specialist skills and use a Cisco partner or managed service for delivery. Every phase boundary doubles as a natural review point.

The Cisco Recommendation

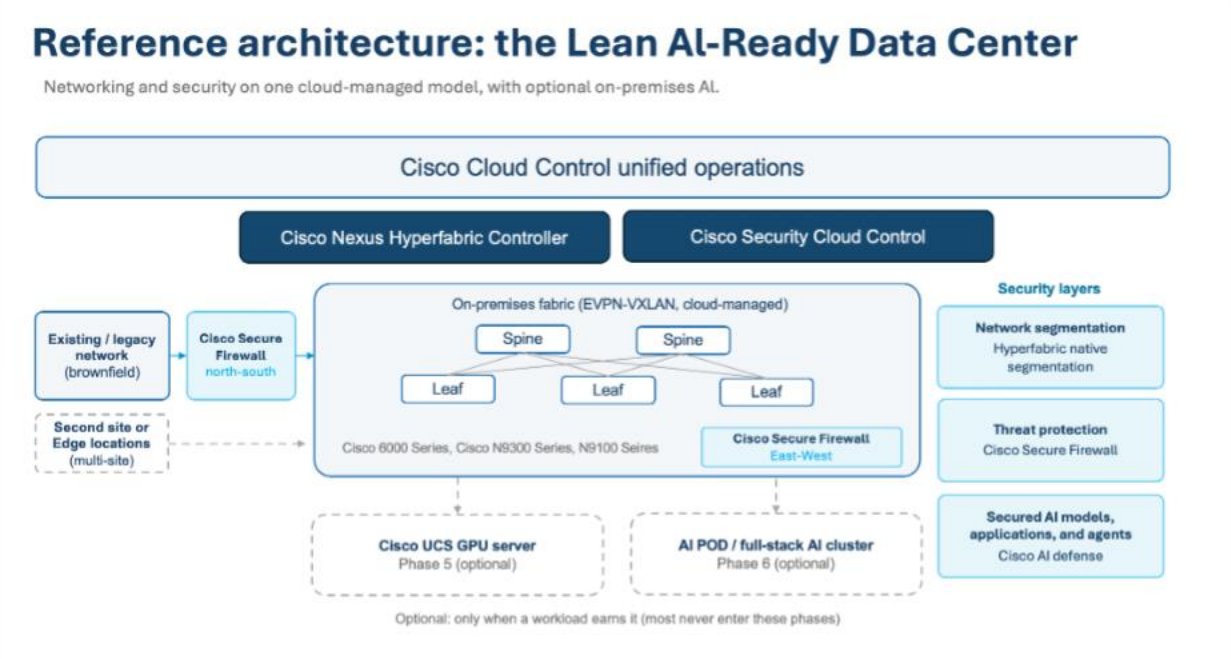
If you would rather not assemble this from parts, Cisco delivers the phases as one cloud-managed operating model, Nexus Hyperfabric under Cisco Cloud Control, available directly or through a Cisco

partner as a managed service. The advantage is the fewest consoles, the least specialized expertise required, and a single model that carries you from a first fabric to enterprise-scale AI if you ever need it.

11. Getting Started

Start by designing, not buying. A short fabric-readiness assessment, run against your current footprint and refresh schedule, should return three things: an inventory of what you run today and what is up for renewal, a validated target design produced by the cloud controller, and a phase-readiness review against the exit criteria in Appendix B. You can begin a design at no cost before any commitment. Nothing is disrupted, and you finish with a clear picture of the first, most valuable step.

From there, the sequence is one this paper has already written. To scope the assessment or see the fabric in action, contact your Cisco account team or Cisco partner.



Appendix A — Capability-to-Cisco Mapping

The capability you need first; the Cisco way to deliver it on the right.

Phase	Capability you need (any vendor)	Cisco recommendation
1 — Design	Validated, cloud-assisted fabric design that scales from a switch pair to a cluster, right-sized speed band, security edge designed in, accurate BoM	Cisco Nexus Hyperfabric cloud controller (Nexus One operating model)
2 — Deploy	Zero-touch fabric deployment beside the existing network, hardened/attested management plane, no disruption	Cisco silicon-based N9300 and Cisco 6000 Series (1-100G, 400-800G) · NVIDIA silicon-based N9100 Series (400-800G) · Nexus Hyperfabric subscription (Essentials / Advantage / Premier)
3 — Migrate	Brownfield interconnect, gateway migration, north-south and east-west firewall, native segmentation. Completes a secure single-site data center	Logical Networks · Cisco Secure Firewall · Security Cloud Control (within Cloud Control)

4 — Extend (if multi-site)	Additional locations on one operating model; the Phase 3 firewall, segmentation extend to new sites; assurance built into the fabric	Hyperfabric multi-site · Secure Firewall / segmentation extended · native assertion-based monitoring · Cisco Cloud Control
5 — Accelerate (optional)	Right-sized single-node on-prem compute, bring-your-own-AI, AI-specific security	Cisco UCS GPU server · Cisco Intersight · Cisco AI Defense · Security Cloud Control · Cisco Cloud Control
6 — Scale (optional)	Pre-validated AI infrastructure, one model from pod to factory, security at every rung	Hyperfabric full-stack AI option · Cisco AI PODs · Secure AI Factory with NVIDIA · AI Defense · Security Cloud Control · Cisco Cloud Control · Cisco Stack Automation

Appendix B — Phase Exit-Criteria Checklist

- **Phase 1 — Design complete when:** a validated fabric design exists, sized to your starting footprint, with a confirmed bill of materials and a chosen first location.
- **Phase 2 — Deploy complete when:** the first fabric is live and cloud-provisioned, with the existing network unaffected.
- **Phase 3 — Migrate complete when:** production runs on the new fabric, a north-south perimeter firewall is enforcing at the border, an east-west firewall within the fabric, native segmentation is in place by the fabric itself, and legacy decommissioning is underway or scheduled. For a single site, the modern data center is complete here.
- **Phase 4 — Extend (if multi-site) complete when:** more than one location runs on one operating model, with the same segmentation and firewalls extended across sites, and the estate visible from a single view.
- **Phase 5 — Accelerate (optional) complete when:** one on-prem workload runs on a right-sized node on the existing fabric, with AI-specific security applied.
- **Phase 6 — Scale (optional):** no exit criteria. A capability made available when the business case arrives; most organizations never enter it.

Phase 5 entry gates (any one)

- Regulatory data-residency requirement.
- Sustained volume with a demonstrated cloud-cost crossover.
- A hard latency or edge requirement.

Appendix C — Status Legend and Sources

Status legend

GA: Generally Available. CA: Controlled Availability. Roadmap: as labeled.

Current statuses referenced

- **Cisco Nexus Hyperfabric:** GA.
- **Hyperfabric switch platforms:** GA. Cisco silicon-based N9300 and Cisco 6000 Series for mainstream speeds of SMB and mid-market (1-100G) to AI-scale (400-800G). NVIDIA silicon-based N9100 Series for customers looking to build a fabric that follows the NVIDIA Reference Architecture and leverages NVIDIA Adaptive Routing technology.
- **Cisco Cloud Control:** The umbrella operating platform that unifies domain controllers, including Nexus Hyperfabric (networking), Security Cloud Control (security), and Intersight (compute). In Controlled Availability (US, 2026; global to follow). Security Cloud Control and Intersight are domains within Cloud Control, not separate platforms.
- **Cisco Secure Firewall:** GA. Secure your modern data center and stay ahead of evolving cyber threats. Deploy a north-south perimeter firewall and an east-west firewall and enable advanced threat protection across perimeters at scale without compromising performance. Hyperfabric does not provide native east-west microsegmentation today; network and workload microsegmentation are treated in the security blueprint, not here.
- **Cisco UCS, Intersight, AI PODs, AI Defense, Secure AI Factory with NVIDIA:** GA.

Sources

- Gartner, 2026 CIO and Technology Executive Survey and 2026 CIO Leadership Perspective Survey (operationalizing AI as the number-two CIO priority, distinct from developing an AI strategy; cybersecurity and risk management as the number-one priority; productivity and cost optimization as the leading business outcomes of technology investment).
- Gartner, The Top CIO Challenges (AI investment accelerating while a majority of AI initiatives fail to reach production).
- IDC Perspective (as referenced in Cisco materials): by 2027, a significant share of enterprises will deploy generative-AI network fabrics for AI workloads in their own data centers.
- International Energy Agency, Energy and AI (2025), and industry data center analyses, for the power and cooling constraints on on-premises AI.
- Industry and analyst assessments of AI data center networking (east-west traffic, GPU utilization, and the network as the performance bottleneck).
- Vendor-neutral data center security best practice on perimeter firewall placement and defense-in-depth (north-south perimeter firewall, network segmentation, and detection and response), and the reality that much data center traffic is east-west and bypasses the perimeter.
- Microsegmentation market research (analyst): workload microsegmentation adoption remains concentrated in large and regulated enterprises, with estimates that only a small share of even enterprises have adopted it, which is why this networking blueprint does not assume it for the SMB and mid-market segment.
- Cisco Secure Network Reference Architecture (SNRA) for the defense-in-depth, segmentation, zero-trust, and hybrid mesh firewall principles (note: SNRA is a campus, branch, and distributed-enterprise architecture, not a data center design).
- Cisco product documentation: Cisco Nexus One; Cisco Nexus Hyperfabric and he Hyperfabric full-stack AI Infrastructure option; Cisco 6000 Series, N9100, and N9300 Series switch data sheets; Cisco Secure Firewall; Cisco Security Cloud Control; Cisco UCS and Cisco Intersight; Cisco AI PODs; Cisco Secure AI Factory with NVIDIA; Cisco AI Defense; Cisco Cloud Control.