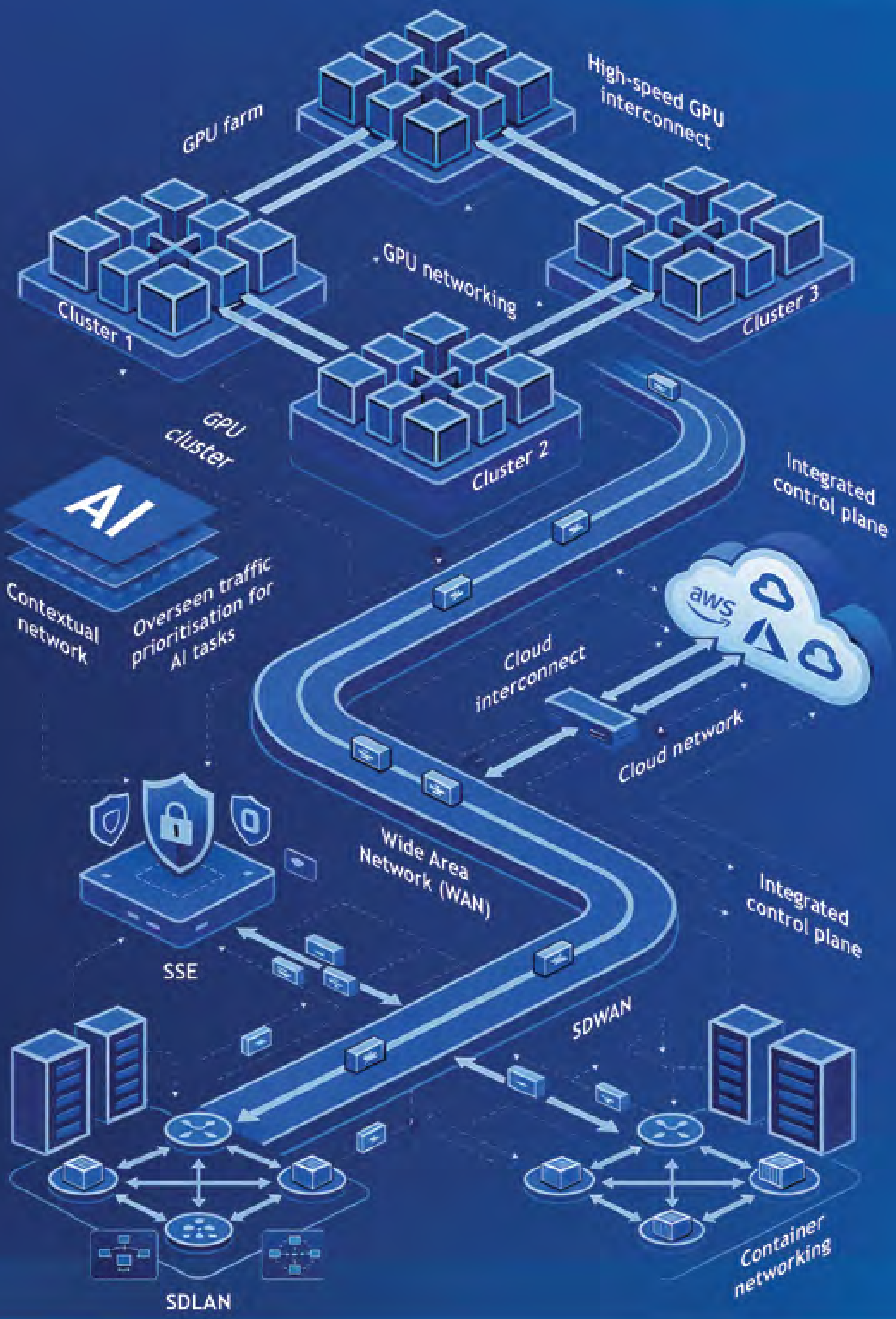
Evolution of Enterprise Network



#InnovatewithDigitalFabric

Table of contents

Executive summary

Enterprise networking is currently undergoing its most consequential architectural transition in a quarter-century. Since the early 2000s, the network has served as the vital nervous system for global business, but its fundamental purpose is shifting. We are moving from a history of **connecting humans to applications** to the future of **networking in the Agentic Era.**

The three acts of network evolution

To understand the current pivot, one must view it through the lens of three distinct architectural eras:

1. **The connectivity era (2000-2005):**
Defined by static MPLS circuits and centralised hardware, the goal was reliable "pipes" connecting branch offices to the corporate data center.

2. **The cloud & mobility era (2005-present):**
The rise of SaaS and hybrid work forced a shift toward internet-first architectures. **SD-WAN** and **SASE** emerged to provide the flexibility needed for users to access applications from anywhere, moving the perimeter to the cloud.

3. **The agentic AI era (now to near future):**
We are now entering an age where the primary "user" of the network will shift from human to dual-**Autonomous AI Agents & humans.** This shift alters the scale, speed, and topology of the fabric, demanding a network that is not just a connector, but an active participant in the inference process.

The agentic pivot: From transactions to inference

As AI agents proliferate, traffic patterns are transitioning from human-paced, sequential transactions to **machine-speed, parallelised inference chains.** These agents invoke multiple models, distributed APIs, and private data lakes simultaneously. This creates unprecedented "East-West" demand within the data center and complex "North-South" flows across the WAN. To sustain this, the network must evolve into a **programmable, inference-aware fabric.**

The new architectural requirements

The move toward an Inference Economy requires a total re-tooling of the network stack:

1. **LAN & Edge:**
   Transitioning to "Inference-at-the-Edge" with high-bandwidth density and extreme microburst resilience.

2. **The intent-driven WAN:**
   A convergence of public internet and deterministic private circuits, utilising real-time intent to steer traffic based on GPU availability, data sovereignty, and cost.

3. **AI-native data centers:**
   The emergence of specialised AI Fabrics—utilising 800G/1.6T speeds and low-tail-latency transport (RoCEv2 / Ultra Ethernet) to maximize GPU utilisation and **Token Throughput.**

4. **Cloud networking (MCN):**
   Evolution toward seamless Multi-Cloud Networking that treats disparate cloud providers as a single, programmable computing resource.

This paper traces the evolution of enterprise networking from its static roots to its autonomous future. It establishes that the re-emergence of **high-performance private networks, edge-localised compute, and observability-driven autonomous operations** are the foundational elements of the next-generation enterprise. In the Agentic Era, the network is no longer just infrastructure—it is the engine of an intelligent enterprise.

Evolution of networks

Understanding how enterprise networking has evolved over the last two decades is essential to understanding the next major shift.

Digitisation began with mainframes in the 1950s, accelerated in the 1990s with the World Wide Web and interconnected business systems, and expanded in the 2000s with mobility, cloud computing, Industry 4.0, Big Data, and IoT. The current phase—driven by GenAI and rapidly moving toward Agentic AI—challenges fundamental assumptions about how networks operate.

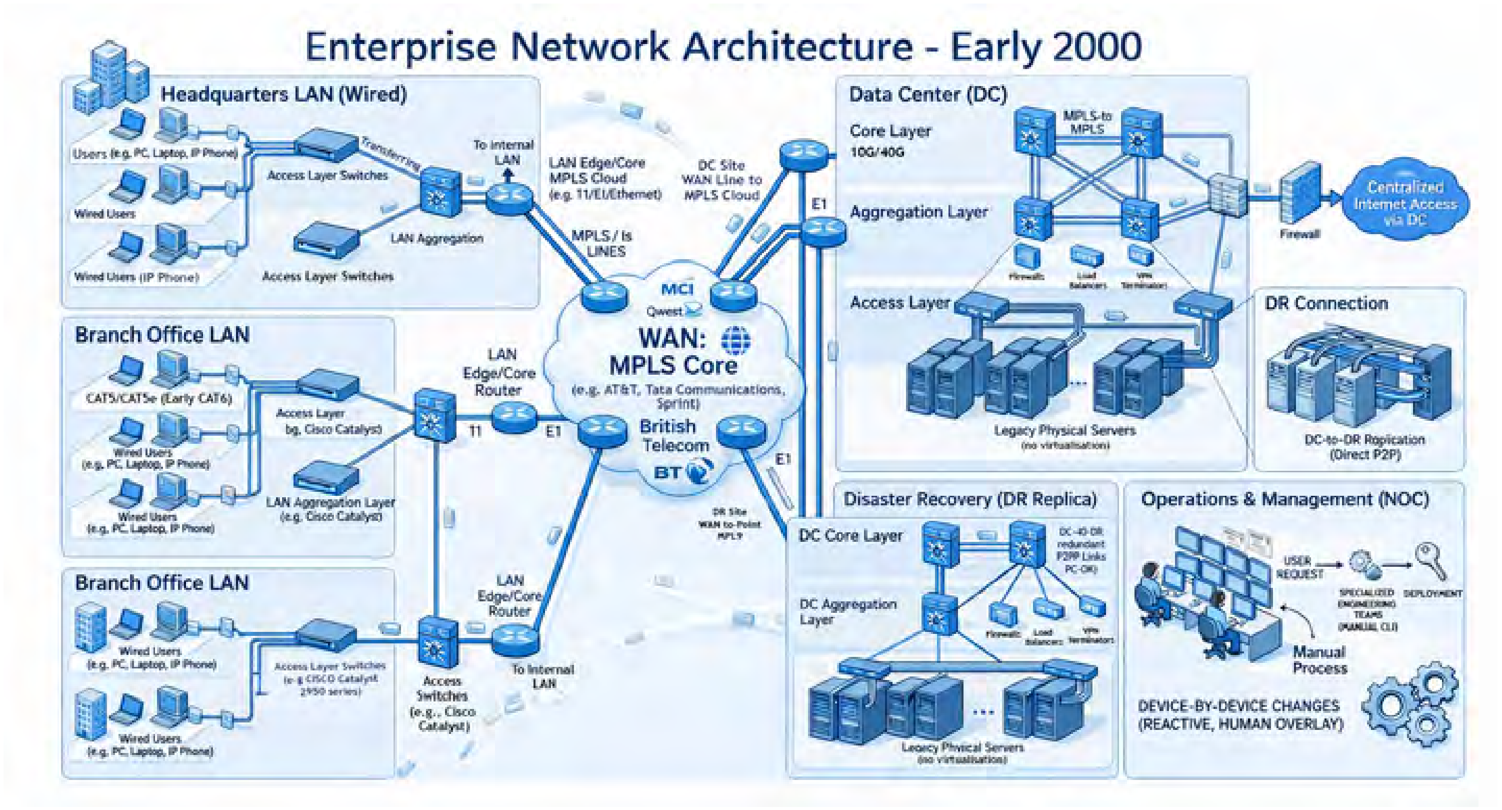
This document examines those macro phases beginning in the early 2000s and explores the changes that Agentic AI will bring in the short to mid term. Future considerations such as Quantum Networking or AGI will introduce further shifts, but those are out of scope for now.

Before exploring the phases, four assumptions remain constant:

1. The network connects users (left) to applications (right)
2. It must remain secure end to end
3. It must meet application parameters (latency, packet loss, jitter)
4. It must meet the uptime requirements of the business

Early 2000

In this phase (2000-2005), enterprise IT was centered almost entirely around data centers, where applications ran on physical server stacks hosted either in enterprise owned facilities or leased from data center providers, and users primarily worked from office locations. "Work from Home" or "Work from Anywhere" existed only as isolated exceptions for specific roles such as executives, field engineers, or support staff. Enterprise networks in this phase consisted of three high level blocks, namely the LAN covering the office/campus/factory network, the WAN connecting offices/factories to data centers, and the data center network as shown in figure below.



The **office network (LAN)** was a predominantly wired environment built on ethernet. Physical connectivity relied on twisted pair copper (CAT5, CAT5e, early CAT6) or coaxial cabling. A typical layout consisted of desktops or laptops connected via ethernet to an access switch—often located on the same floor. Multiple access switches connected in a star topology back to a central distribution switch, which in turn connected to routers that interfaced with the WAN. VLANs were a foundational design component, used extensively to segment departments, floors, and roles. Wireless LANs (Wi-Fi) technically existed but were limited in coverage, speed, and security (WEP/WPA1) and therefore not used/trusted as primary access networks.

The **Wide Area Network (WAN)** connected offices to data centers and data centers to each other. This was almost entirely a private-network era. WAN technologies included Layer 1 leased lines (Domestic and International Private Lines), Layer 2 ethernet circuits, and MPLS (Layer 2.5/3) which was still gaining global adoption. Internet access for users was typically centralised and “hair pinned” through the data center, where the security stack (firewalls, IDS/IPS, proxies) was located. Some enterprises enabled direct internet access from offices, but this required local firewalls and was less common due to cost and risk.

Enterprises frequently mixed technologies based on use case: DC to DC links used Private Line/ethernet for predictable performance, while offices connected to DCs via MPLS to enable any to any communication for voice/video without backhauling through the DC.

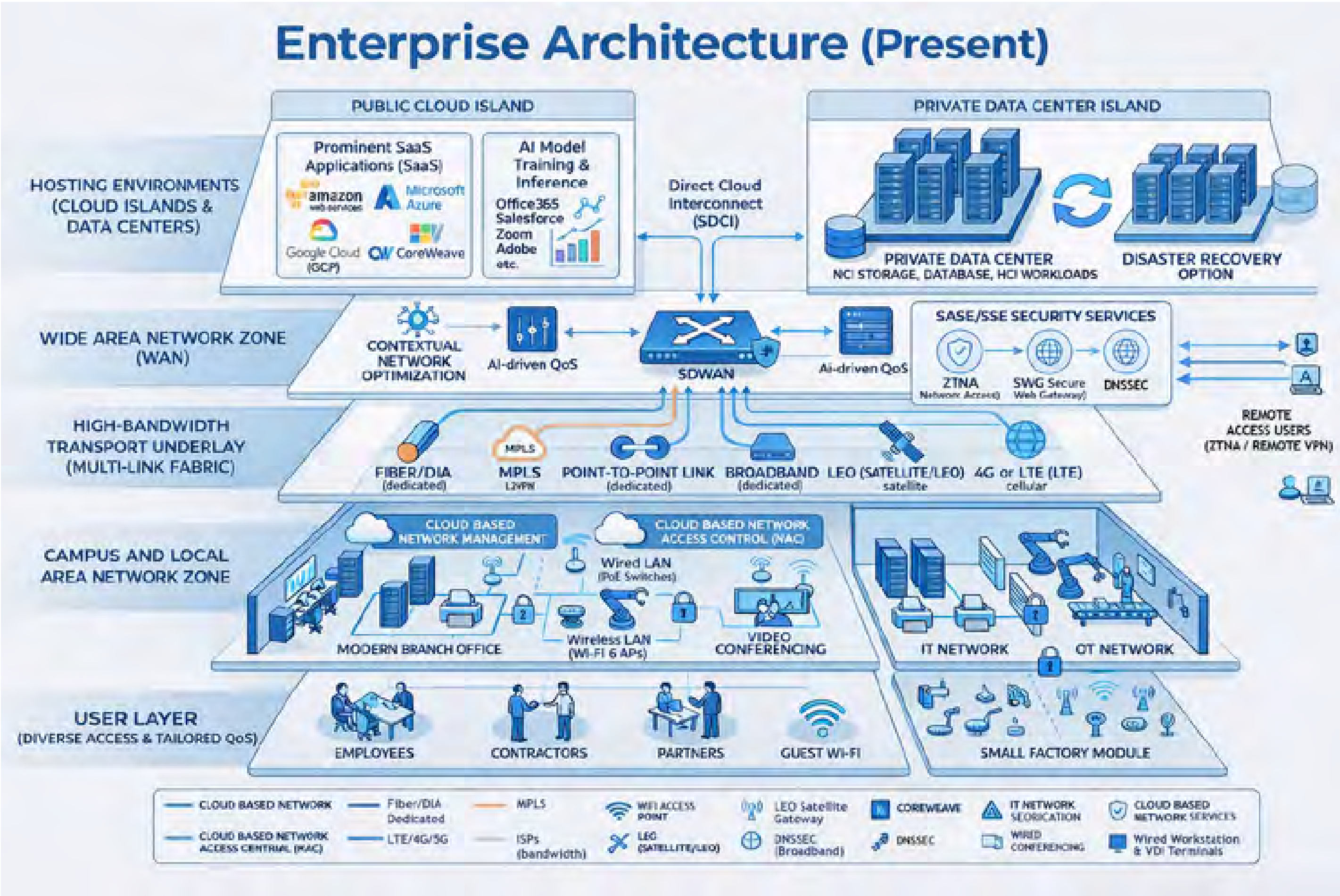
The **Data Center (DC) networks** were built using a three-tier hierarchical architecture—core, aggregation, and access—designed mainly to support North-South traffic (client-to-server) rather than the East-West traffic (server-to-server) that dominates today. Server connectivity typically relied on 10/100 Mbps ethernet with 1 Gbps uplinks, and deployments made heavy use of copper cabling routed under raised floors. Network design within the data center was largely Layer 2 at the rack level, with Layer 3 implemented only at the core, resulting in pervasive use of Spanning Tree Protocol (STP) and VLANs that spanned large portions of the data center. End-of-Row (EoR) switching was common instead of the Top-of-Rack (ToR) model used today, server speeds ranged from 100 Mbps to 1 Gbps, and environments consisted of physical servers with no overlays, no software-defined networking (SDN), and extensive reliance on dedicated appliances. DC to DC replication or DR traffic ran over Private Lines, ethernet circuits, or dark fiber, depending on latency and distance requirements.

To manage this environment, enterprises operated **NOCs (Network Operations Centers) and SOCs (Security Operations Centers).** These could be fully in house or outsourced, but the operations model was centralized, ticket driven, and largely manual, reflecting the stable but hardware centric nature of networks during this period.

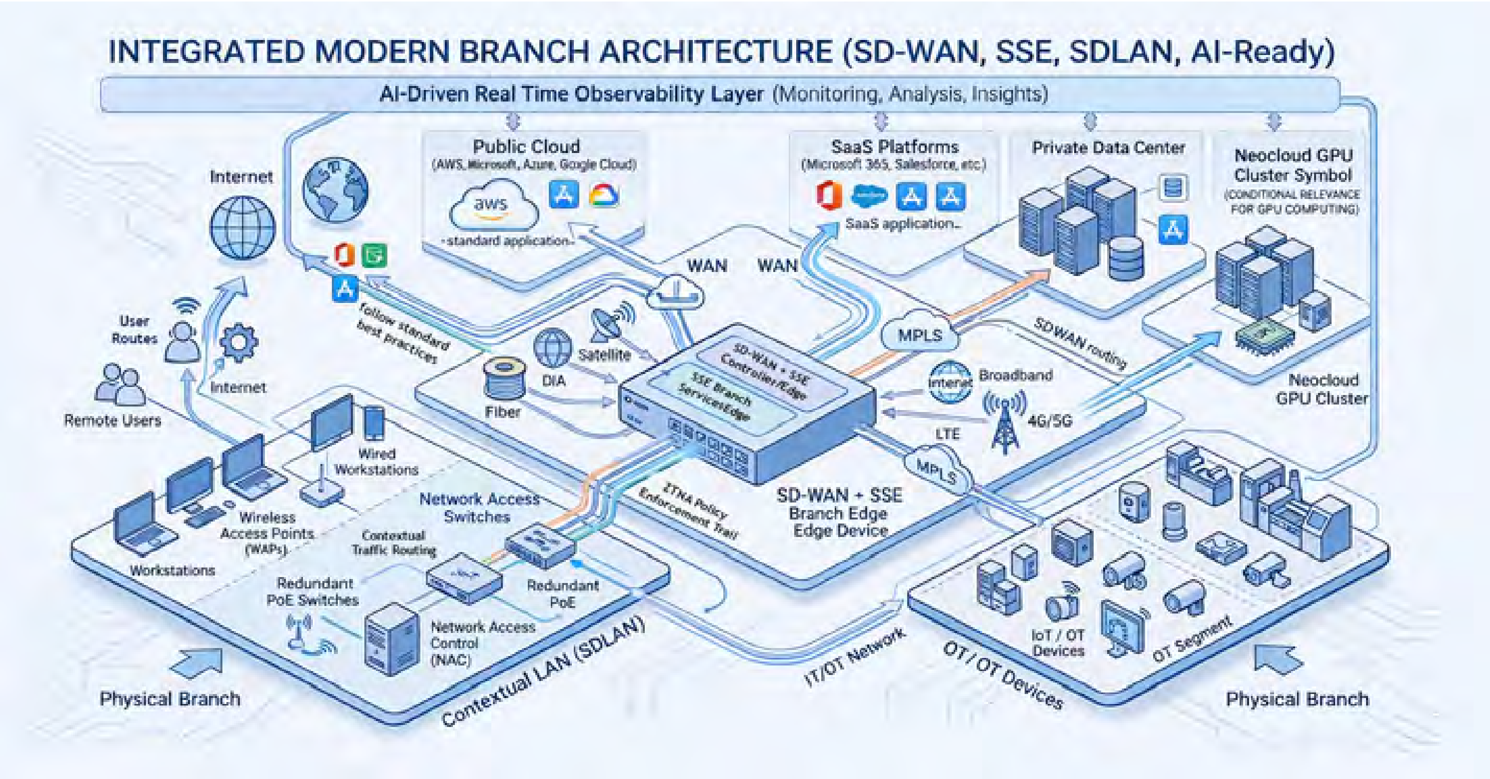
Next phase of network- Starting 2005 to present

With public cloud becoming mainstream during this period—and with the explosion of data and new application architectures—enterprises began moving workloads from the data center to cloud, marking the start of large-scale cloud migration. As public cloud ran on public networks, the natural choice for connecting users and applications became the internet.

While the early transition years saw a hybrid state of public cloud + data center, which created temporary complexity in networking, it is more important to focus on the steady state enterprise network that emerged after this transition. In this period, the enterprise network evolved into four primary building blocks: the Office Network, the Work from Anywhere Network, the Wide Area Network, and the Cloud Network. For enterprises retaining both DC and Cloud, a fifth block—Data Center Network—remained relevant. Enterprise architecture below shows high level view of enterprises today which ensures users (in User Layer) to connect to applications (in DC and cloud layer) with focus on user experience, cost and security.



The **Office Network (LAN)** shifted from predominantly wired to predominantly Wi-Fi. Ethernet drops were replaced by Access Points (APs), which connected to switches, then to a firewall and router or SD-WAN appliance. Over time, LAN design moved from port based to identity based access, with SSIDs mapping to VLANs and NAC systems enforcing user or device profiles. AP density increased to support video, IoT, and collaboration workloads, while switching backbones evolved to CAT7 and fiber uplinks to carry rising east west traffic. Cloud managed LAN controllers, AI enabled RF optimisation, improved PoE budgets, and segmentation for corporate, guest, and IoT networks became standard, transforming the LAN into a wireless first, policy driven access fabric.



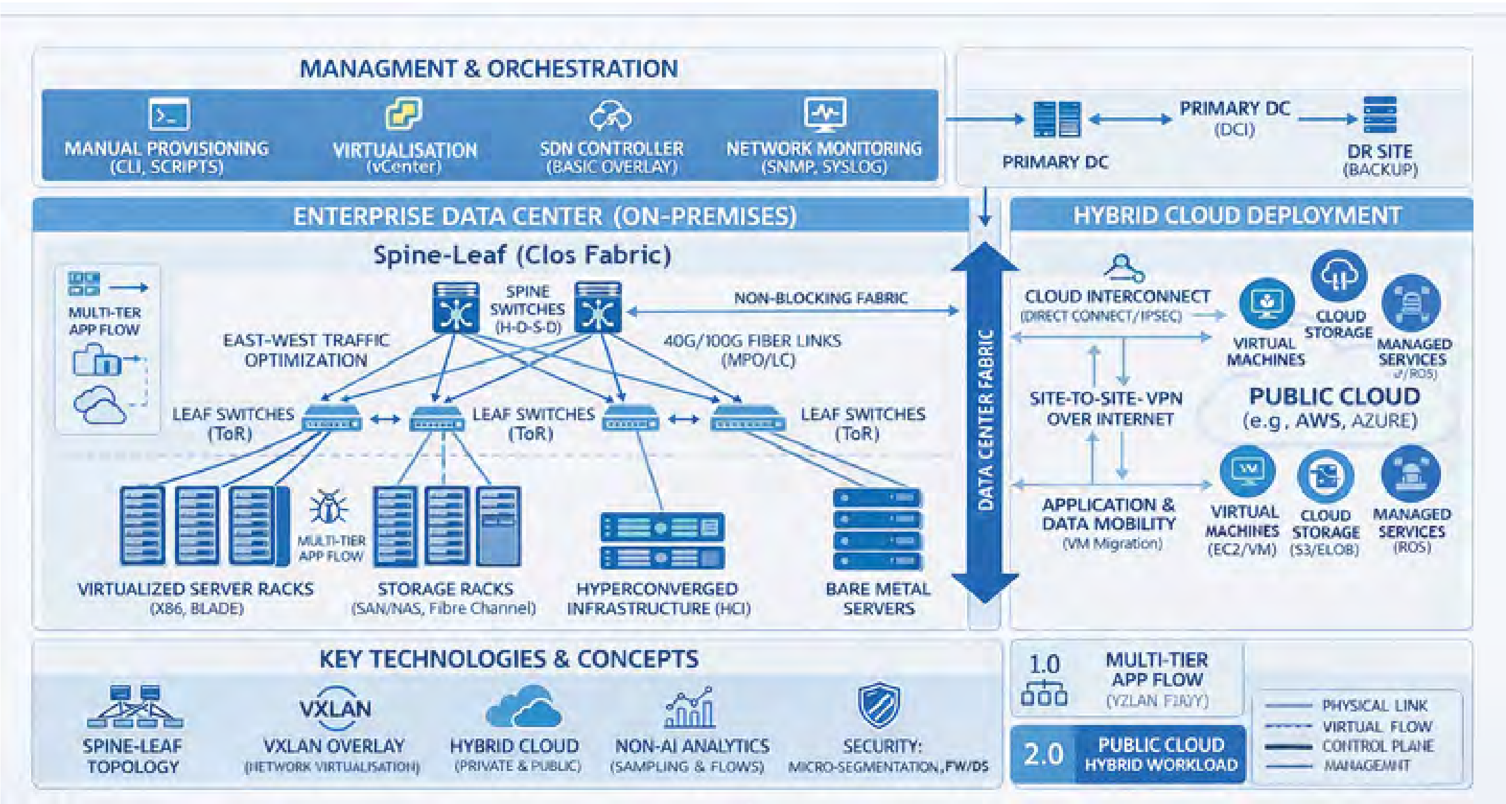
The **cloud network** evolved from a single link to EC2 or S3 in 2006 to a densely interconnected mesh of multi cloud environments. Today, over 90% of enterprises operate at least a two cloud strategy. Connectivity spans both private and public options, with careful design needed to control egress costs. Conceptually, cloud networking now has two major components: "to the cloud" and "within the cloud," structured around standardised landing zones. However, many enterprises have under designed this layer, leading to what we call the "Complexity Tax"—a tangible cost paid partly to the cloud provider and partly as operational inefficiency or opportunity loss. This tax is already significant and will become even more critical as networks evolve to support AI and Agentic AI workloads. If unaddressed, this complexity can severely constrain future cloud network performance.

The **Work from Anywhere network** began modestly wherein remote employees and road warriors connected via IPsec/SSL VPN tunnels. Pre COVID, only ~6% of employees relied on this. COVID accelerated adoption dramatically, pushing enterprises from low single digit usage to nearly 100% remote, stabilizing at ~30% today. As application patterns changed, VPN centric architectures shifted to SSE platforms, with many enterprises using a hybrid approach—SSE for cloud/SaaS and IPsec for legacy or latency sensitive applications.

The **Wide Area Network** pivoted from private networks to public networks as applications migrated to Cloud. This transformed enterprise WAN architectures, replacing MPLS or Private Lines with internet—primarily DIA—except in regulated sectors like Banking & Financial Services, which retained MPLS due to compliance and security needs. For Internet based architectures, SD-WAN became the default mechanism for traffic steering, centralised policy, and path selection. SD-WAN has since evolved into SASE, integrating security and AI assisted automation to create proactive, self healing network behavior. As application volumes and UCC workloads grew, enterprises faced bandwidth inflation, demand for better routing, and operational overhead of multiple regional ISPs. This drove consolidation toward global Tier 1 providers and region specific design patterns (e.g., Dual DIA, DIA + broadband, or single/dual broadband), depending on office type, region, and business criticality.

The **data center network,** the fifth block for hybrid enterprises—underwent its own transformation. From traditional three tier architectures, the shift moved through virtualization, then leaf spine fabrics, then SDN and overlays, eventually standardising on EVPN VXLAN based fabrics supporting 25G/100G/400G. As cloud scale operational models entered the enterprise, networking transitioned from manual configuration to intent based automation and closed loop telemetry. Modern DC networks are now 400G centric, telemetry driven, intent based, and increasingly AI assisted, enabling self operating behavior to manage rapidly growing workloads.

Together, these blocks come together to build the enterprise Network Fabric. It is essential that each block be built modularly—well defined, documented, and designed with clear interdependencies—because issues in one block can cascade across others. Best practice is to first design an overall Network Fabric blueprint or reference architecture, then design each block individually, along with documented operations and interlock procedures for the Network Operations team (NOC) and Security Operations Center (SOC). NOC and SOC continue to manage the entire setup, wherein some of the repetitive work, specifically in the L1 desk, is automated through AIOps to drive efficiencies.



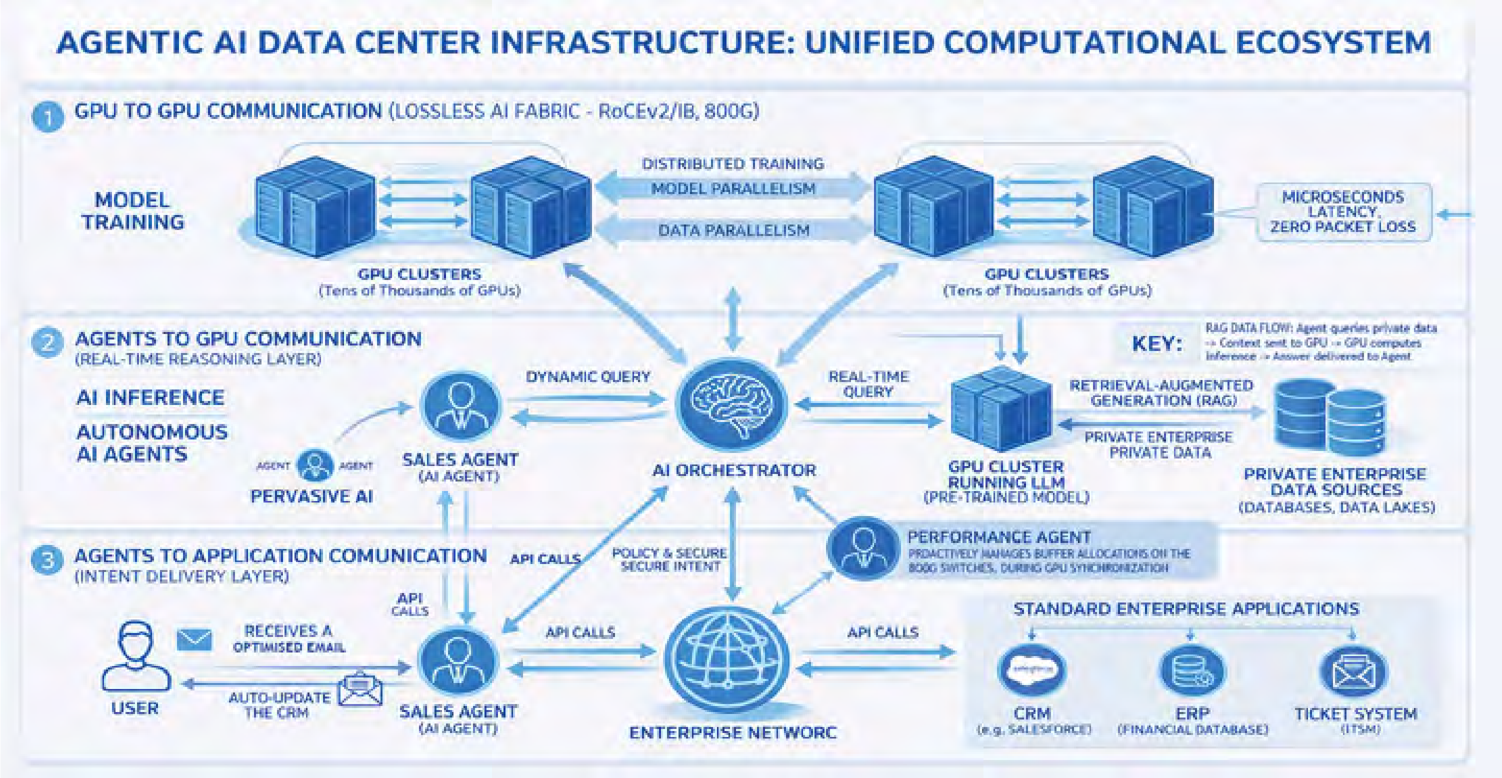
Network in agentic era

The primary change agent in this phase will be “the agents” enabled by Agentic AI. While AI adoption accelerated with the introduction of transformer architectures and GenAI, **the next era is defined by autonomous AI Agents capable of reasoning, planning, collaborating, and completing multi step tasks across systems.** this ushers in an **“Agentic Web” or “Agentic Network”,** where machine-speed agents operate in parallel with humans, dramatically increasing concurrency, API calls, and east–west inference flows.

Historically, enterprise networks were designed predominantly for human users (even as IoT emerged). **Going forward, networks must support a dual population of humans and machine speed agents.** Because agents operate orders of magnitude faster than humans and generate significantly more parallelized interactions, enterprises will face substantial changes in traffic patterns, latency needs, architectural choices, and scaling expectations. Many **enterprises will choose to build or host models and agents in data centers or Neo Clouds instead** of public cloud, driven by cost, regulatory, sovereignty, or operational control reasons along with constraints on space and power.

To support this shift, networks must evolve to handle 50x–200x current workload levels, driven by parallel agent interactions, continuous context retrieval, and cross-model inferencing. As inference becomes distributed, the network becomes the performance control plane. Purpose-built fabrics such as the Arrcus Inference Network Fabric (AINF) exemplify this shift: they provide AI-policy–aware routing that dynamically steers traffic between inference nodes, caches, and data centers to maximise GPU utilisation, increase tokens-per-second (TPS), reduce time-to-first-token (TTFT), and improve end-to-end latency. Such fabrics will integrate with frameworks such as NVIDIA Dynamo, BlueField DPUs, and Spectrum-X Ethernet to deliver sovereign routing, model resolution, and priority classification at the moment a request enters the network.

In this phase, the enterprise network will evolve into **six high level blocks: Office Network, Work from Anywhere Network, Edge Network, Wide Area Network, Data Center Network, and Cloud Network.** Observability becomes a pervasive, foundational layer across all these blocks. Terms like “Intent Based Networking,” “Self Healing Networks,” and “Autonomous Networks” transition from aspirational to mainstream expectations, cutting across every layer and block. Before detailing these blocks, it is important to understand the potential traffic flows and behaviors that agent driven systems will generate, both for general agent interactions and for AI model training/inference workloads that shape overall architecture.

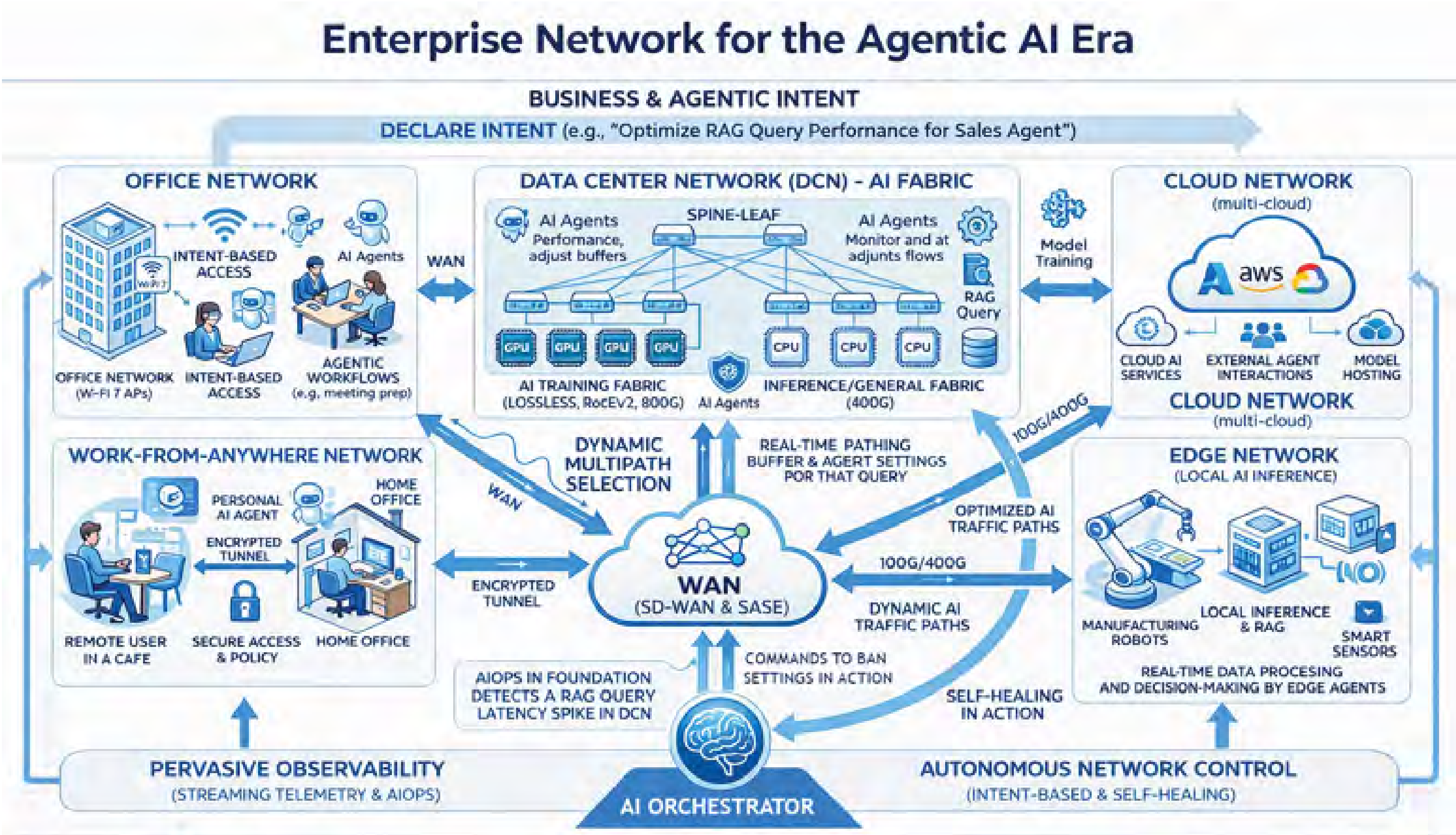


AI models will increasingly be built in distributed data centers housing GPU/NPU/LPU clusters. Today's intra DC interconnects move gradients and activations between GPUs with extremely high bandwidth and sensitivity to latency. Within a node, interconnects like NVLink already offer very high bandwidth. Across nodes, InfiniBand or RoCE based Ethernet fabrics provide hundreds of Gb/s bandwidth, and network efficiency directly determines GPU utilisation. As clusters outgrow a single facility due to space, power, and cooling constraints, enterprises will distribute training infrastructure across multiple data centers. In such scenarios, 800 Gbps to 1.6 Tbps DC to DC interconnects will become standard, with RDMA driven, lossless fabrics optimised for distance-aware synchronization and mitigation of packet loss. These AI fabrics will routinely transport massive "elephant flows," requiring deterministic performance. Much like low latency trading routes in financial networks, specialised premium routes will emerge for AI training clusters, priced and engineered for minimal latency and loss.

These interconnects may exist within enterprise DCs or across Neo Clouds where network services can be consumed as a utility. AI models running in public cloud will rely on multi cloud connectivity frameworks, discussed later in this document. Agent traffic originates on the "left," residing on user laptops, phones, enterprise devices, or IoT/OT sensors. Depending on the assigned task, agents will traverse enterprise boundaries, accessing enterprise applications, external SaaS platforms, public websites, marketplaces, or even interacting with third party agents. For example, an employee instructs an agent to plan an official trip; the agent queries travel sites, corporate booking systems, maps, calendars, weather services, restaurant databases, and more—performing dozens of rapid, parallel lookups before presenting a synthesised output. Such tasks generate high volume, public internet traffic with no strict real time requirements.

Conversely, agents performing tasks like safety monitoring, robotic control, facility automation, or real time decisioning will require tight latency envelopes. Agentic AI will massively increase the volume and burstiness of enterprise traffic—possibly 50× to 200× current levels—due to machine speed parallel interactions, API chaining, and agent to agent protocols. This also drives the need for Edge Compute, where certain tasks must execute within 10–100 ms latencies that traditional cloud or centralised DCs cannot guarantee. Similarly, device capabilities (routers, switches, servers) will need to scale dramatically: API request rates will explode as agent coordination protocols (MCP or enterprise equivalents) become mainstream, and network infrastructure must handle the resulting concurrency load.

With this background, each enterprise network evolves accordingly:



The Office Network

will undergo a major bandwidth and architectural uplift to support agent driven traffic. While the core structure of Wi Fi APs, switches, and SD WAN/routers persists, the density, throughput, and intelligence of these components increase significantly. Edge compute nodes may be deployed directly within office environments to support low latency inferencing for local agent workloads. AI enhanced routers will emerge to interpret network intent, classify human versus agentic flows, apply traffic aware policy routing, and coordinate with centralised intelligence.

The Work from Anywhere Network remains structurally similar

to the present model, with security responsibilities shifting toward the user edge. However, SSE platforms will integrate intent aware capabilities to differentiate between human and agentic¬ traffic, applying policies that optimise performance, security, and compliance. Lightweight client side agent aware routing may be embedded into endpoint security stacks to ensure proper traffic steering.

The Edge Network becomes mainstream

Enterprises will deploy small LLMs, SLMs, or uSLMs at or near user locations to support agentic workflows requiring ultra low latency inference. Edge may exist on premise or be consumed “as a service” from providers. CDN providers will transform existing nodes into AI capable edge locations; telecom and mobile-network operators will evolve base stations into inference points; and data center providers will pair with regional fiber operators to create distributed edge zones.

The Edge Network becomes the aggregation and acceleration layer for agentic flows, ensuring deterministic performance before sending traffic toward DCs, Neo Clouds, or public clouds.

The Wide Area Network now connects human users and agents to DCs,

Neo Clouds, and public clouds. While the public internet remains critical, a resurgence of private networks occurs, particularly for workloads demanding stable characteristics (latency, jitter, packet loss). Intent based networking becomes operational: flows dynamically select paths—to Edge, DC, Neo Cloud, or Public Cloud—based on task type and required performance. The WAN must support on demand bandwidth increases, path provisioning, and dynamic route allocation. Office connectivity at 10–100 Gbps becomes standard, with multi port architectures enabling simultaneous private and public breakout.

Data Center Networking evolves into an AI native,

ultra high bandwidth fabric interconnecting DCs, Edge sites, Neo Clouds, and cloud environments. DC to DC links operate at 800G–1.6T, using Ultra Ethernet class congestion control, RDMA acceleration, and optical interconnects optimised for AI cluster synchronization. DC to Edge links operate at 100G–400G, while DC to Neo Cloud and DC to Cloud interconnects sit between 400G–800G. Path selection becomes workload aware: latency sensitive AI training and inference use premium lossless routes; general compute uses cost optimised paths. Intra DC switching fabrics standardize on 51.2 Tbps silicon with a path defined to next gen 102.4 Tbps platforms. Effective multi die, optical spine architectures will allow aggregate fabric capacities approaching 130 Tbps, particularly in GPU/NPU/TPU clusters. AI native DC fabrics become autonomous, telemetry driven, and resilient by design.

- **Cloud Networking becomes a highly programmable,**

on demand connectivity layer spanning Edge, DC, Neo Cloud, and Public Cloud, with link speeds from 100 Gbps to 1.6 Tbps. Networks are dynamically instantiated or retired based on intent signals, workload triggers, or agent requirements. Multi cloud connectivity becomes continuously reconfigurable, enabling rapid creation of ephemeral networks for AI workflows. One of the main components of cloud networking will be Layer 7 Networking for K8S, which is Networking for AI. In the modern enterprise, container networking has evolved from a simple connectivity layer into a sophisticated, high-performance fabric. This shift is critically important for today's **Agentic and Generative AI** use cases, where applications are no longer static request-response systems but dynamic "reasoning" loops. AI agents often need to pull data from distributed vector databases, call external inference APIs, and coordinate with specialised "tool" containers in real-time. Because these AI workloads are computationally expensive and latency-sensitive, the underlying network must provide near-zero overhead while ensuring that sensitive model weights and proprietary data remain isolated and secure across hybrid environments.

Kubernetes addresses these AI requirements through a "sidecar-less" architectural evolution (sidecars – Routers, Firewall, UTM, IPS, etc). By leveraging **eBPF (Extended Berkeley Packet Filter),** Kubernetes networking (via CNIs like Cilium) moves complex logic out of the standard Linux networking stack and directly into the kernel. This allows AI workloads to bypass slow context-switching between user and kernel space, providing the high throughput necessary for massive data ingestion and model inference. Simultaneously, Kubernetes uses **SPIFFE (Secure Production Identity Framework for Everyone)** to solve the "Identity" problem. In an agentic workflow—where an autonomous AI agent may be dynamically spawned to perform a task—SPIFFE provides a cryptographically verifiable identity (an **SVID**) that is independent of IP addresses or physical location. This ensures that only authorised AI agents can access specific GPUs or sensitive data "brains," even if they are moving across private and public cloud boundaries.

Observability becomes the nervous system of the enterprise network. It continuously measures human and agentic behavior, identifies bottlenecks, predicts failure, and feeds insight into intent based policy engines. Observability systems integrate tightly with distributed AI routers at the Edge and centralised LLM routers operating within service provider or enterprise NOCs. Together, they ensure that both human and machine traffic flows receive the appropriate performance, security, and routing treatment required in an agent driven enterprise environment.

NOC and SOC will remain where we will see concept of "Dark NOC" and "Dark SOC" taking shape where most of the operations will be driven through AI. It will start with "human in the loop" initially and steadily move into more automated and AI driven flows.

TATA COMMUNICATIONS

TATA

Conclusion

The network will become the nerve center and the operational and economic control plane of the AI-powered enterprise. As workloads shift from human-paced interactions to machine-speed agentic workflows, traditional incremental upgrades are no longer sufficient. Enterprises must adopt blueprint-driven architectures built around modular network blocks, AI-policy–aware routing, dynamic multi-cloud fabrics, pervasive observability, and intent-based automation. Success in the agentic era will depend on the ability to build networks that dynamically adapt to inference workloads, route traffic intelligently across edge, data center, and cloud, and ensure sovereignty, latency, and cost-efficiency at scale.

By adopting this architecture, organisations are moving away from the "Security vs. Speed" trade-off. They are building a network that is inherently faster because it’s in the kernel, and inherently more secure because every connection is backed by a verifiable identity. For the business, this translates to **amplified network performance, higher GPU utilisation,** and the ability to meet strict data privacy regulations without slowing down the development of AI-driven features.

At Tata Communications, we are working with the best minds in the industry to help enterprises build this next-generation network foundation and unlock AI as a revenue-generating capability, then a ‘yet another network transformation’. We continue to build on our IZO+ portfolio to bring Dynamic Network as an offering, launched ThreadSpan™ for Observability and building AI in the heart of our offering through Commotion. In addition to these, we are also building for the future, which will help large global companies in their journey to the Agentic Era.

For more information, visit us at www.tatacommunications.com

Contact us

#InnovatewithDigitalFabric