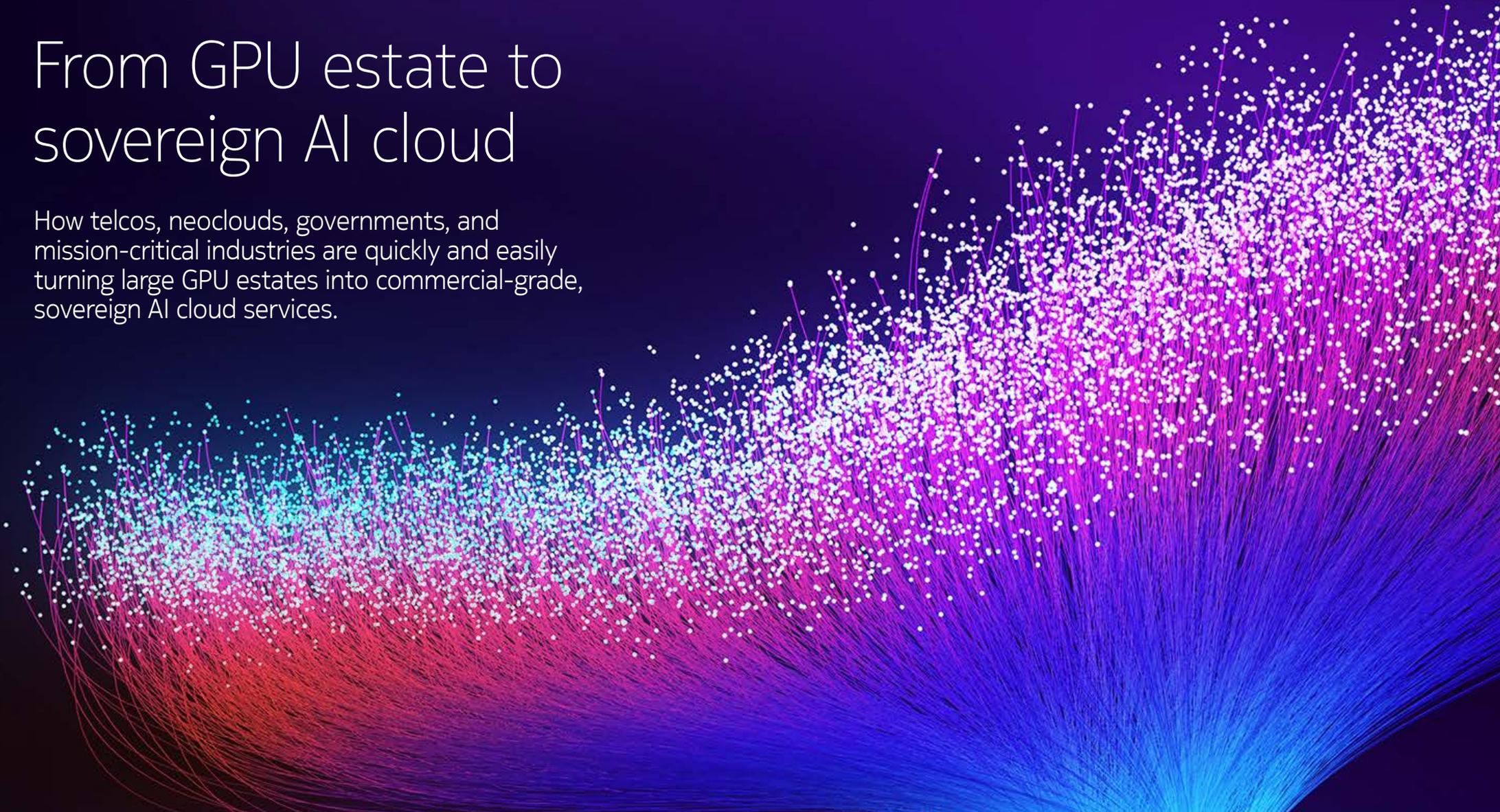


NOKIA



# From GPU estate to sovereign AI cloud

How telcos, neoclouds, governments, and mission-critical industries are quickly and easily turning large GPU estates into commercial-grade, sovereign AI cloud services.



# Introducing the Gcore and Nokia AI partnership

Organizations that own large GPU estates face a common dilemma: they own cutting-edge GPU infrastructure at scale but lack a platform to use and monetize those GPUs. Building a hyperscaler-grade AI cloud from scratch can take 12–24 months and demands expertise that most enterprises cannot assemble in-house. GPUs sit idle or underused, reducing ROI.

Gcore and Nokia address this together. Gcore AI Cloud Stack transforms raw GPU clusters into fully operational, multi-tenant AI cloud environments in weeks, complete with orchestration, billing, observability, and integrated go-to-market support. Nokia Data Center Fabric solution provides the lossless, automated, AI-scale networking foundation that underlies reliable and scalable performance.



“

There’s a generation of GPU estate owners sitting on some of the most powerful infrastructure available today with no fast path to commercial AI services. We built AI Cloud Stack to change that. Our partnership with Nokia means customers don’t have to choose between speed and reliability: they get a production-grade platform with the networking foundation to back it up.

— Andre Reitenbach  
CEO, Gcore

## The challenge:

# A raw GPU estate without a cloud platform

Owning GPU infrastructure is only part of the equation. To deliver AI services to customers, operators need a complete software platform on top: multi-tenancy, orchestration, billing, observability, and compliance tracing. And beneath it all, there needs to be a network fabric engineered to move petabytes of training data without bottlenecks or packet loss.

The operators most affected fall into three types:

- Telecommunications providers with inter/national GPU deployments, seeking to launch sovereign AI-as-a-service for regulated public-sector and enterprise customers.
- Regional neocloud providers competing for enterprise GPU contracts but lacking the CloudOps tooling to productize their capacity
- Government, defense, and mission-critical enterprises deploying AI in air-gapped, classified facilities where public cloud connectivity is prohibited.

Three shared challenges emerge for large-scale GPU owners:

### Deployment complexity

Building an AI cloud on raw GPU infrastructure typically takes 12–24 months and requires deep expertise across networking, orchestration, storage, security, and operations — expertise most GPU estate owners do not have in-house.

### Sovereignty requirements

Regulated industries and government bodies cannot use hyperscaler clouds. They need on-premises deployment, data residency guarantees, full operational control, and audit traceability that only private infrastructure can provide.

### Monetization risk

An in-house platform build represents a multi-million dollar investment without guaranteed revenue. Without billing infrastructure, go-to-market channels, and tenant management, GPU capacity depreciates without generating return.

## The solution:

# Gcore & Nokia

Gcore and Nokia address these challenges with a unified solution, with clear roles that eliminate integration risk and accelerate time to market:

- Gcore leads on cloud software, AI services, and go-to-market
- Nokia leads on AI-scale networking fabric and lifecycle network automation

## Gcore AI Cloud Stack

### Innovative AI software

Transforms bare-metal GPU clusters into hyperscaler-grade AI cloud environments, proven on estates of 20,000+ GPUs in Europe.

- Full cloudification layer: IaaS, PaaS, GPUaaS, and Model-as-a-Service in a single integrated stack
- Multi-tenant orchestration with Kubernetes, VAST AI OS unified governance, and production-grade inference pipelines
- Built-in billing engine, observability, and white-label options purpose-built for CSPs and regional providers
- Air-gapped training capability for fully offline, secure model development in classified environments
- Integrated GTM support including NVIDIA Cloud Provider reseller channels, so capacity generates revenue rather than sitting idle

## Nokia Data Center Fabric

### Ultra reliable, highly performant networking

AI-scale networking foundation that makes GPU fabrics perform predictably, operate reliably, and scale efficiently.

- High-performance switching: 7220 IXR and 7250 IXR family of switching platforms supporting leaf/spine/super-spine architectures at AI scale
- SR Linux NOS: Open, model-driven network operating system built for programmability, telemetry, and operational simplicity
- Nokia EDA (Event-Driven Automation): Full lifecycle automation covering Day 0 design, Day 1 deployment, and Day 2+ operations, reducing configuration risk and enabling repeatable, auditable deployments
- Lossless, ultra-reliable fabric purpose-built for the all-to-all traffic patterns of distributed AI training and high-throughput inference

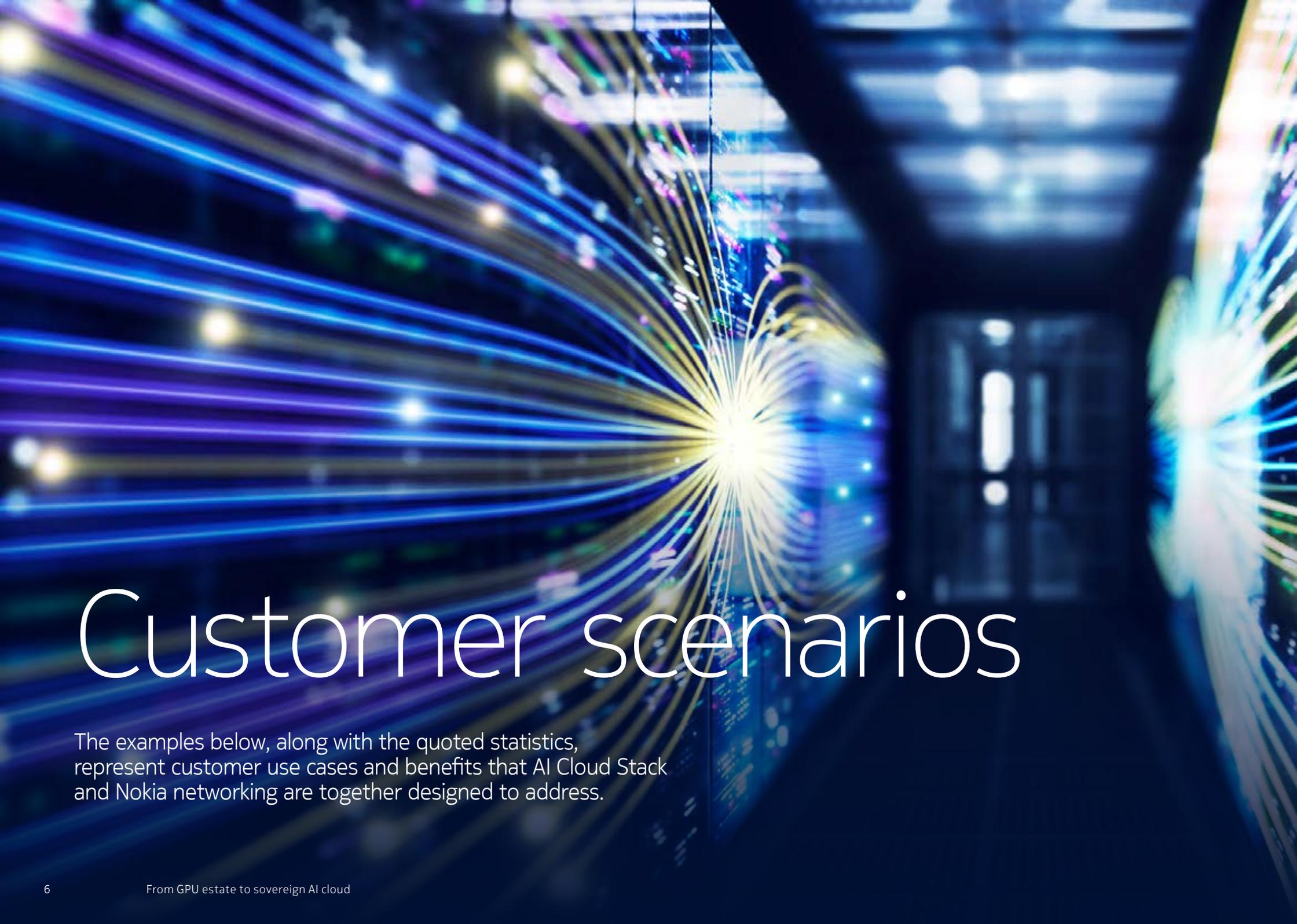


“

AI cloud deployments fail or stall at the network more often than anywhere else, because of congested fabrics, manual provisioning, switching that was never designed for east-west GPU traffic. SR Linux and EDA solve exactly that challenge. Pairing Nokia's data center fabric with Gcore AI Cloud Stack gives operators a platform where the network scales as fast as the business does.

— Mark Vanderhaegen

Head of Business Development, Data Center Networks for Nokia



# Customer scenarios

The examples below, along with the quoted statistics, represent customer use cases and benefits that AI Cloud Stack and Nokia networking are together designed to address.

## Scenario 1

# European telecommunications provider

Sovereign AI-as-a-service for government ministries

### Background

A major European telecommunications provider holds 2000 GPUs across two national data centers, acquired as part of a 5G infrastructure investment program. The provider wants to launch a sovereign AI-as-a-service (AIaaS) platform for government ministries and critical national infrastructure operators, but faces strict requirements: full data residency within national borders, multi-tenant isolation between agencies, and operational auditability mandated by the EU AI Act's phased compliance framework.

### Challenge

The engineering team lacked the time and expertise to build a multi-tenant AI cloud platform, particularly the orchestration, billing, and compliance-tracing layers required for government contracts. An internal estimate placed the DIY build timeline at 18–22

months. Meanwhile, government procurement windows were already open. The legacy Ethernet fabric was not engineered for the lossless, high-throughput demands of distributed AI training across racks of H100s, risking GPU idling and degraded performance during large model training runs.

### Solution

The provider deployed Gcore AI Cloud Stack on its existing GPU estate, replacing the months-long platform build with a structured deployment of the cloudification stack. Nokia simultaneously replaced the legacy switching layer with a purpose-built AI fabric using the 7250 IXR series in a leaf-spine architecture, operated via SR Linux and automated end-to-end via Nokia EDA.

6 weeks

Contract to first government tenant live on platform

91%

Average GPU utilization across 800 nodes within 90 days

4 agencies

Government ministries onboarded in first quarter of service

Zero

Data egress events; 100% residency within national borders



“

Sovereign AI has moved from a compliance checkbox to a core business requirement. Telecommunications providers need private AI, and they need it deployed, operational, and generating value within a quarter, not a fiscal year. This Gcore & Nokia collaboration is the first time we've seen customers genuinely able to deliver on that timeline.

— Seva Vayner

Product Director of AI and Cloud, Gcore

## Scenario 2

# Regional neocloud provider

## GPU monetization at scale

### Background

A regional neocloud provider acquired 1000 GPUs through a strategic partnership with a national carrier. The infrastructure was operational, but the company's platform was bare-metal rental: no tenant isolation, no usage metering, no managed AI services. The provider was losing enterprise deals to better-equipped competitors and facing board pressure to demonstrate AI revenue within two quarters.

### Challenge

Without a CloudOps layer, the provider could not offer the multi-tenant isolation, usage-based billing, or managed inference services that enterprise buyers required. Each new customer required manual, custom provisioning, which was slow, error-prone, and unscalable. The internal network was built for traditional workloads, not for the all-to-all traffic patterns of large-scale AI training.

Network congestion during heavy training runs caused GPU stalls, degrading job performance and generating customer complaints that threatened early relationships.

### Solution

Nokia redesigned the fabric with 7220 IXR leaf switches running SR Linux, with Nokia EDA automating fabric operations: node provisioning time dropped from hours to minutes, and continuous telemetry delivered per-tenant SLA data directly into the provider's observability stack. Gcore AI Cloud Stack was deployed as the full control plane for the neocloud's GPU estate, instantly providing multi-tenancy, Kubernetes orchestration, a billing engine, and a model-as-a-service catalog, transforming bare-metal rental into an enterprise-grade AI cloud product. Gcore's integrated GTM support accelerated the sales pipeline alongside the technical launch.

7 weeks

Time to first billable enterprise tenant (vs. 9-month internal build estimate)

3×

Revenue per GPU vs. bare-metal rental pricing, enabled by managed AI services

Zero

Network congestion events in the 6 months post-deployment

18 tenants

Active enterprise customers onboarded within 6 months of platform launch

## Scenario 3

# International airline

## Sovereign AI for fleet-wide predictive maintenance

### Background

A major international airline wanted to deploy a fleet-wide predictive maintenance system to anticipate engine component failures and optimize repair schedules. The model required continuous analysis of high-volume sensor telemetry and maintenance records, but the airline's proprietary engineering data and safety logs could not be processed on public, multi-tenant clouds.

### Challenge

Analyzing petabytes of raw engine telemetry demanded substantial compute capacity close to where data was generated. At the same time, aviation safety regulators required end-to-end auditability: every AI-driven maintenance recommendation needed a verifiable, human-reviewable decision lineage to ensure accountability in the event of incident investigation. The airline needed strong data residency controls,

local processing, and regulator-ready evidence trails—without slowing down hangar operations.

### Solution

The airline deployed the Nokia-Gcore sovereign AI solution to turn the hangar into a localized, secure AI environment. As aircraft taxied into the hangar, they connected to a Nokia Private 5G network to offload engine telemetry (vibration, temperature, pressure) to on-site infrastructure. Gcore AI Cloud Stack processed the data locally by comparing each flight's telemetry against "golden models" to detect subtle anomalies early. For every recommended action, the system generated an auditable decision lineage report linking the alert to the specific sensor patterns and thresholds that triggered it, enabling mechanics to validate recommendations and providing regulators with a clear, traceable rationale.

25%

Reduction in AOG  
(Aircraft on Ground)  
time

Zero

Safety data egress  
events

100%

Audit compliance for  
maintenance  
recommendations

5G

Private hangar  
connectivity for rapid  
telemetry offload



## Why Gcore & Nokia

Gcore brings the software that turns GPUs into cloud products. Nokia brings the network that makes those products perform and scale. Natively integrated, together Gcore and Nokia deliver unparalleled reliability and scalable performance, designed with sovereign AI cloud deployments in mind.

# Business outcome

## Gcore contribution

## Nokia contribution

### Weeks to revenue

AI Cloud Stack eliminates the platform build; tenants can be onboarded immediately after deployment

EDA-automated fabric deployment reduces network setup from days to hours, removing the networking bottleneck from the critical path

### Sovereign deployment by design

Air-gapped training, on-premises orchestration, and local billing, with no external cloud dependencies at any layer

SR Linux and EDA operate entirely on-premises, with no external management plane or telemetry requirements

### Multi-tenant commercial operations

Built-in multi-tenancy, Kubernetes orchestration, billing engine, and white-label support for CSP go-to-market

Model-driven network operations and per-tenant telemetry provide the observability data required for SLA reporting and chargeback

### Predictable, high-throughput AI networking

Integration-tested against Nokia fabric for optimized GPU-to-GPU traffic flows across the full AI training and inference lifecycle

7220/7250 IXR platforms with lossless switching purpose-built for the all-to-all traffic patterns of large-scale AI workloads  
Nokia SR Linux is built with the latest in congestion avoidance techniques and is compatible with the Ultra Ethernet Consortium

### Compliance and audit readiness

Decision lineage reporting, VAST AI OS governance layer, and traceability artifacts aligned to risk-based regulatory frameworks (e.g., EU AI Act)

EDA-automated configuration management and change records provide network audit evidence for internal and regulatory review

### Maximize ROI

Gcore AI Cloud Stack accelerates ROI by replacing years of platform development with a weeks-long deployment of commercial-grade AI cloud services and integrated go-to-market support

With Nokia EDA, operational issues are discovered quickly and easily and resolutions are immediately identified with its AIOps and automation capabilities.

“

Gcore gives operators a path from GPU estate to cloud product. Nokia gives them the certainty that the network underneath will never be the reason their platform fails to scale. Together, our work is to make AI infrastructure and networking invisible to the end customer, so they can focus on the services they're building.



— Marika Mentula

VP Network Infrastructure, Nokia

Ready to turn your GPU estate into a sovereign AI cloud?

Gcore AI Cloud Stack

Find out more:

<https://gcore.com/ai-cloud-stack>

Nokia AI data center networking

Find out more:

<https://www.nokia.com/data-center-networks/ai-data-center-networking/>

Reference architecture

Learn more about our design:

<https://nokia.ly/ReferenceArchitecture>

Nokia OYJ  
Karakaari 7  
02610 Espoo  
Finland

Tel. +358 (0) 10 44 88 000

CID:215330

nokia.com

NOKIA



#### About Nokia

Nokia is a global leader in connectivity for the AI era. With expertise across fixed, mobile, and transport networks, we're advancing connectivity to secure a brighter world.

Nokia is a registered trademark of Nokia Corporation. Other product and company names mentioned herein may be trademarks or trade names of their respective owners.

© 2026 Nokia