RAFAY

Telco Use Case

NIM-Powered Model Marketplace as a Service – Delivered by Rafay

Enable telcos to deliver sovereign, revenue-ready AI marketplaces powered by NVIDIA NIM and orchestrated by Rafay.

Turn raw GPU infrastructure into a fully operationalized marketplace where enterprises can select, deploy, and consume AI services instantly, all delivered with governance, sovereignty, and monetization controls.

From Bottleneck to Business Value

Telcos face mounting pressure to transform GPU investments into differentiated AI services. But challenges persist:

- Static model catalogs lack operational depth and enterprise readiness.
- Building portals, pipelines, and integrations in-house takes months.
- Enterprises demand compliant, sovereign AI services that hyperscalers cannot always provide.

Rafay + NVIDIA NIM change the equation. Telcos can stand up branded marketplaces in weeks, offering enterprises instant access to models, blueprints, and compute resources delivered as governed services.

Designed For

Telecom Operators / NCPs: Monetize GPUs with AI marketplaces that bundle compute, models, and blueprints.

Enterprises: Access trusted, production-grade AI endpoints through telco-operated portals.

Sovereign & Regional Clouds: Deliver compliant, in-country AI services with sovereignty, auditability, and consumption-based billing.

Key Capabilities

Capability	Description
Models-as-a-Service	NVIDIA NIM-powered models exposed as governed API endpoints with tenant-level usage tracking.
Blueprints-as-a-Service	NVIDIA-curated blueprints (fraud detection, medical imaging, data flywheels) deployable in one click.
Compute-as-a-Service	On-demand provisioning of Kubernetes clusters, VMs, and GPU slices.
Operationalized Marketplace	White-labeled portals that let enterprises browse, deploy, and consume services instantly.

Business Outcomes

Outcome	Impact
Rapid Time-to-Market	Launch an operationalized, NIM-powered AI marketplace in weeks instead of months.
New Revenue Streams	Offer differentiated SKUs from compute to models to blueprints and maximize GPU ROI.
Trusted Enterprise Adoption	Deliver sovereign, compliant AI services that enterprises trust and hyperscalers struggle to match.

RAFAY

Rafay Systems, Inc. | 530 Lakeside Dr. Ste 220 Sunnyvale, CA 94085 | 669.336.4800 | rafay.co