

Accelerating

Financial AI Innovation

with Velocis

TIFIN takes their AI models into production on Neysa's secure, performant GPU AI PaaS Velocis cloud to serve India's largest mutual funds and asset management companies.

NEYSA SETUP

Velocis AIPaaS Platform

- On-demand Jupyter Notebooks
- On-demand Managed Models

Velocis GPU Virtual Machines & Storage

- NVIDIA H200
- H100 SXM
- NVIDIA L40S
- NVME Storage

OUTCOME

65%

Reduction in total GPU cloud spend

Faster path from experiments

Full ownership

Of model weights & data stays in TIFIN's control

Enhanced system stability

OVERVIEW

TIFIN is a premier global fintech innovator dedicated to engineering AI-powered platforms that are reshaping the wealth management industry. With a significant footprint in India, TIFIN serves a prestigious clientele that includes the nation's largest mutual funds, and wealth management companies.

By leveraging advanced artificial intelligence, TIFIN identifies specific growth vectors for financial advisors and individual investors to deliver hyper-personalized investment journeys. Their mission is to optimize financial decision-making through intelligent, data-driven automation. Ultimately to create better wealth outcomes for more people.

| Case Study

✉ hello@neysa.ai

🌐 www.neysa.ai

CHALLENGE

“ We faced a strategic paradox: scaling our AI required hyperscale resilience, yet global consumption models were commercially inefficient. We needed a partner to bridge this gap, delivering enterprise-grade stability without compromising our unit economics. ”

Kiran Nambiar
Chief Product Officer, TIFIN India & CEO, MyFi

TIFIN faced a strategic dilemma in scaling its AI infrastructure within India’s regulated ecosystem. To maintain optimal unit economics, the company utilized a local neocloud provider, but struggled to meet the reliability and performance standards required for financial platforms.

Conversely, migrating fully to global hyperscalers presented a cost structure that would dilute the ROI of scaling.

01. Economics

Misaligned cost structures

Global hyperscaler consumption models were incompatible with TIFIN’s objective of achieving efficient, sustainable growth at scale.

02. Reliability

Infrastructure volatility

The incumbent local provider failed to deliver enterprise-grade resilience, causing frequent interruptions to critical training workloads.

03. Expertise

Technical deficits

A lack of deep AI expertise from the local provider forced TIFIN to divert engineering focus to troubleshooting.

The Infrastructure Paradox

Every prior option forced a critical trade-off

	Local neocloud	General-purpose clouds	Required for production
Unit economics GPU cost at scale	• Affordable	• Inflated at scale	• Affordable and predictable
Reliability production uptime	• Chronic outages	• Reliable	• Reliable uptime with SLAs
Data sovereignty SEBI • RBI • DPDP	• India-local	• CLOUD Act exposure	• Sovereign by design

TRANSFORMATION

“ Neysa has been a force multiplier. Velocis provides the mission-critical stability and instant scalability of a global hyperscale AI cloud, but with an efficiency that transforms our GPU infrastructure from a cost center into a distinct competitive advantage ”

Uttung Malkanv
Head of Enterprise Business, TIFIN India

To resolve this deadlock, TIFIN integrated Neysa's Velocis AI Cloud into its core development stack, bridging the gap between affordability and performance while replacing the limitations of its local provider with an enterprise-grade environment. Velocis delivers hyperscaler-grade stability, security, and tooling with the economic efficiency needed to support TIFIN's high-growth business model.

01. Hyperscaler-Grade Stability

Velocis eradicated the chronic outages of the previous provider, delivering instant, reliable GPU provisioning that ensures business-critical training workloads run without interruption.

02. Frictionless MLOps Workflows

By offering on-demand Jupyter notebooks and a unified control plane, the platform removes infrastructure overhead. This allows TIFIN's R&D teams to focus solely on model innovation rather than managing compute resources.

03. Streamlined Path to Production

Integrated Inference Endpoints provide a direct, simplified pipeline to deploy models. This seamless transition from lab to commercial deployment significantly accelerates time-to-market compared to the manual processes previously required.

OUTCOME

Dedicated GPUs

GPU VMs with NVMe storage, predictable economics at enterprise scale.

65% reduction

Lower GPU compute spend vs general-purpose clouds.

On-demand

Jupyter Notebooks and Managed Inference, no manual deployment handouts.

Start your production AI deployment with Neysa

Get in touch with our experts.

Request a demo | hello@neysa.ai

