Tata Communications Vayu AI Cloud

Secure, governed, and customer-first AI for BFSI sector.

AI's potential in BFSI is massive — but deploying it securely, at scale, and in compliance with regulations is complex. You can't just "swipe a credit card" and point AI at your core systems. The real question isn't if you should use AI, but how you can do it without compromising security, sovereignty, or cost control.

Key Imperatives for BFSI

Hybrid Cloud Reality

Sensitive workloads stay on-premises for a reason, but AI model training needs massive, burstable GPU compute — impractical to maintain in-house.

Absolute Data Sovereignty

Regulatory compliance demands that data residency and computation are controlled and auditable, even when bursting to the cloud.

Predictable Costs

AI workloads are unpredictable, but your budget shouldn't be. Avoid billing shocks with transparent, predictable pricing.

Unified Control

Fragmented infrastructure and platforms create risk and complexity. BFSI needs a one-stop shop for governance and security.

Freedom from Lock-in

Your AI strategy should evolve—leverage the best open-source tools, not just a proprietary ecosystem.

vayu | AI Cloud

The Vayu AI Cloud Advantage

Private & Secure
Enterprise-Grade Security
Private, low-latency
connections you can trust.

Flexible Cloud Options
**Solution that best fits your
business needs**
From fully customized, air-gapped
private clouds to multi-tenant
public clouds.

Sovereignty

High-Speed Fabric
Always-On Performance
Low-latency fabric linking your
data center with our cloud.

Hybrid Made Simple
Seamless Hybrid Experience
Operate everything under a
single operational umbrella.

Scalable, High-Performance AI Infrastructure

GPU Power
- **NVIDIA-Certified Infrastructure**
  Access cutting-edge GPU clusters when you need to scale.
- **Latest Architectures**
  H100 & H200 SXMs purpose-built for large-scale training.

Specialized Inference
- **Massive Model Inference**
  H200 NVL configurations designed for demanding
  inference workloads.

Versatile Options
- **General Purpose Flexibility**
  L40S & L4 GPUs for a wide range of everyday tasks.

Backbone Infrastructure
- **Blazing-Fast Interconnects**
  High-speed InfiniBand networking for seamless GPU communication.
- **No Bottlenecks**
  Parallel file system engineered to keep GPUs fully fed.

Vayu Cloud PaaS

Orchestration
CNCF-Certified Kubernetes
Enterprise-grade container orchestration for hybrid and multi-cloud.

Data Foundation
Robust Data Platform
Ingest, process, and manage data at scale for any workload.

Agility & Automation
Integrated DevOps Services
Streamlined CI/CD pipelines to accelerate innovation.

AI Empowerment
AI Studio (The Star of the Show)
Give your data science teams powerful tools to build and deploy AI.

AI Studio

Workspace: Your AI Workbench
- **Exploratory Data Analysis** — Managed Superset with secure data access.
- **Notebooks on Demand** — Pre-configured VSCode, JupyterHub, RStudio with major AI frameworks.
- **Heavy-Duty Data Prep** — One-click Spark, Dask, or H2O.ai clusters.
- **Dataset Management** — Version, share, and catalog datasets in Vayu Object Storage.
- **Experiment Tracking** — Managed MLFlow ensures reproducibility.

Serverless AI: Accelerating Innovation
- **Model-as-a-Service** — Instant access to leading GenAI models.
- **Prompt Engineering Playground** — Experiment & test RAG workflows interactively.
- **API Integrations** — Plug into Document AI, Text-to-Speech, etc.
- **One-Click Deployments** — Vector DBs & RAG frameworks on-demand.

MLOps / GenAI Ops: Productionizing AI
- **Model Registry** — Govern and version models securely.
- **CI/CD for Models** — Pipelines integrated with MLFlow for automated deployments.
- **Real-Time Serving** — High-performance inference + interactive AI apps.
- **Fine-Tuning Service** — Adapt models to proprietary datasets.
- **Responsible AI** — Guardrails + Explainability tools (critical for finance).

Partner Ecosystem: Ready-to-Deploy AI
- **Marketplace** — Pre-built, partner-validated AI solutions under a single SLA.
- **Agentic CX Platform** — Hyper-personalized campaigns, intelligent chatbots, next-gen customer service.

Case study Scenario

An insurance firm wants to build a generative AI-powered claims processing assistant. The core claims data must remain in their on-premise data center. They need to fine-tune a large language model on this data and deploy the final application for internal adjusters.

How Vayu AI Cloud Provides the Hybrid Solution

STEP 1

Securely Connect and Explore (On-Prem + Cloud)

Using the Tata Comms Network Fabric, they establish a secure, private link. Their data scientists, using the AI Studio Workspace in the cloud, can now securely run an EDA notebook to analyze the on-prem data in-site, without migrating it.

STEP 2

Burst to the Cloud for Fine-Tuning

They select a base model from our Serverless AI offering. Using the fine-tuning service, they spin up a cluster of H100 GPUs on-demand. The training job securely pulls batches of anonymized data over the private network, fine-tunes the model, and then shuts down the expensive GPU resources, ensuring a predictable cost.
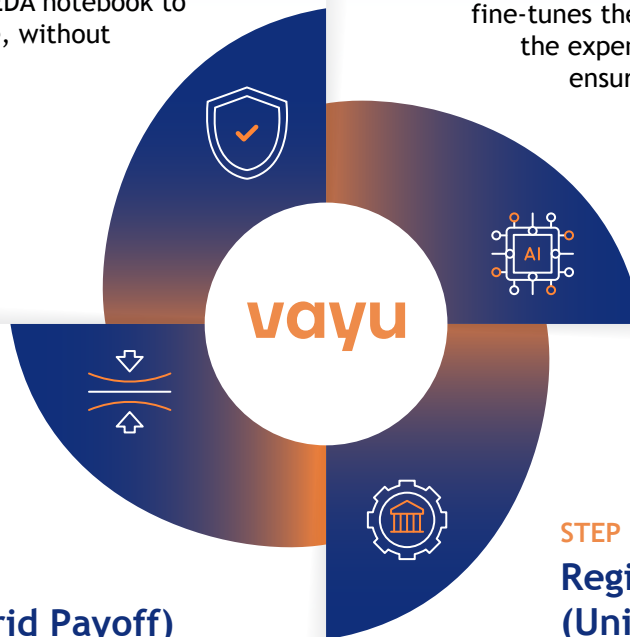
vayu

STEP 4

Flexible, Sovereign Deployment (The Hybrid Payoff)

Using the MLOps pipelines, they now have a choice. They can deploy the inference service in the Vayu Cloud for scalability, OR they can deploy it back onto a smaller GPU server in their on-prem data center for maximum security and low-latency access for their adjusters. The same platform manages both deployment targets. This is the ultimate hybrid flexibility.

STEP 3

Register and Govern (Unified Control)

The newly fine-tuned model is checked into the Managed Model Registry in the AI Studio. It's now a governed, versioned asset, visible through the single pane of glass, even though its training data remains on-prem.

Ready to unlock the next era of AI in BFSI?

Talk to us today about Tata Communications Vayu AI Cloud.

vayu | AI Cloud

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

Contact us