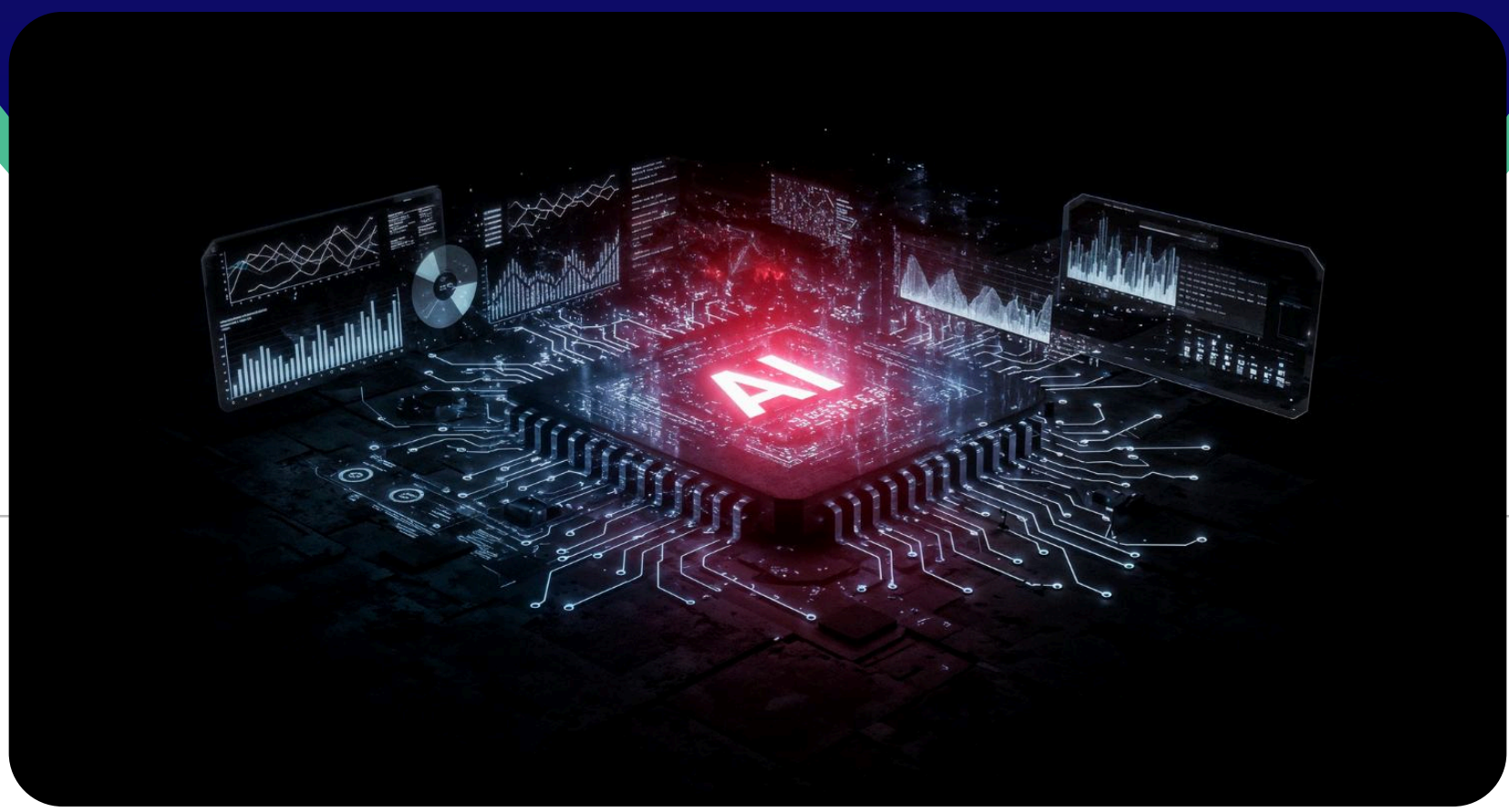
LARSEN & TOUBRO

vyoma

Enabling Sustainable, Sovereign AI Infrastructure for a Large Enterprise

About the Client



The client is a large enterprise operating in a data-intensive environment, supporting a mix of enterprise applications, digital platforms, and emerging AI-driven workloads. With increasing reliance on advanced computing for analytics, automation, and customer-facing services, the organization required infrastructure capable of handling high-density workloads at scale.

At the same time, the client operated under strict regulatory and data governance requirements, with growing expectations around sustainability, cost control, and infrastructure transparency.

The Challenge

As AI adoption accelerated, the client encountered multiple infrastructure and operational challenges:

1. Rising Infrastructure and Energy Demand

AI workloads required compute-intensive environments, leading to increased energy consumption, higher operational costs, and pressure on existing infrastructure.

2. Sustainability and Compliance Pressures

The organization needed to align its infrastructure with environmental standards while ensuring compliance with evolving regulatory requirements.

3. Limited Control in Traditional and Public Cloud Models

Existing infrastructure models, including dependence on external cloud and virtualization platforms, created concerns around data sovereignty, vendor lock-in, and pricing predictability.

4. Evolving Virtualization Landscape

Changes in enterprise virtualization licensing models introduced cost uncertainty and operational risks, making long-term infrastructure planning more complex.

5. Scalability Constraints for AI Workloads

The client required a platform capable of supporting GPU-intensive AI/ML workloads while maintaining consistent performance and flexibility.

Client Requirements

The client's priority was to build a controlled, scalable, and sustainable infrastructure environment that could support long-term AI adoption. Key requirements included:

- ***Full control over data location and infrastructure (data sovereignty)***
- ***Support for high-performance AI/ML workloads using GPU infrastructure***
- ***Predictable and optimized cost structure***
- ***Compliance with environmental and regulatory standards***
- ***Flexibility to scale infrastructure without operational disruption***
- ***Reduced dependency on external vendors and proprietary platforms***

Solution Implementation

1. Sustainable Data Center Design

The infrastructure was designed to optimize resource usage while supporting high-performance workloads:

(i) Water Management System

- ***100% wastewater treated and reused through a ZLD-based system***
- ***Rainwater storage supporting operational requirements***
- ***Air-cooled chillers eliminating cooling tower water dependency***

(ii) Energy Optimization

- ***Integration of renewable energy contributing approximately 30% of total consumption***
- ***Deployment of rooftop solar for on-site energy generation***
- ***Use of energy-efficient systems including VFD-enabled pumps and EC fan technology***

(iii) Environmental Compliance

- ***IGBC Green Data Center - Platinum Rating***
- ***ISO 14001 and ISO 45001 certifications***

(iv) Climate-Conscious Infrastructure

- ***Use of environmentally safer refrigerants***
- ***Inert gas-based fire suppression systems with no ozone depletion or global warming impact***

2. AI-Optimized Infrastructure Enablement

To support AI initiatives, the environment incorporated:

- ***GPU-enabled compute infrastructure for training and inference workloads***
- ***High-speed interconnects (InfiniBand/NVLink) for efficient data processing***
- ***Integrated ML/Ops capabilities for model training, deployment, and lifecycle management***

This allowed internal teams to develop and scale AI workloads without dependency on external platforms.

Execution Approach

The implementation was carried out in a phased manner:

Phase 1: Deployment of core infrastructure and migration of critical workloads
Phase 2: Validation of performance, stability, and compliance alignment
Phase 3: Optimization of energy usage, cost structures, and workload efficiency
Phase 4: Enablement of internal teams to manage and scale the platform independently

This ensured minimal disruption while establishing a stable and scalable operating environment.

Measurable Results

The deployment delivered clear operational and financial outcomes:

- ***Up to 53% cost savings over three years, with additional long-term reductions***
- ***Break-even achieved in approximately 16 months for initial deployments***
- ***Improved energy efficiency, supported by renewable integration and optimized systems***
- ***Reduced water consumption, enabled by recycling and conservation mechanisms***
- ***Predictable cost structure, improving financial planning and scalability***
- ***Enhanced security and compliance, with full control over data and infrastructure***
- ***Improved AI workload performance, supporting faster development and deployment cycles***

Business Impact

The transition strengthened the client's overall infrastructure strategy:

- ***Enabled greater control over data, infrastructure, and compliance requirements***
- ***Supported scalable AI adoption without compromising operational efficiency***
- ***Improved alignment with sustainability and environmental goals***

Conclusion

By implementing a sustainable hyperscale infrastructure integrated with a sovereign cloud platform, the client addressed key challenges related to scalability, cost control, and governance.

LARSEN & TOUBRO

vyoma

Start Your Journey With Us

Contact Us