

Accelerating AI Transformation with Torque



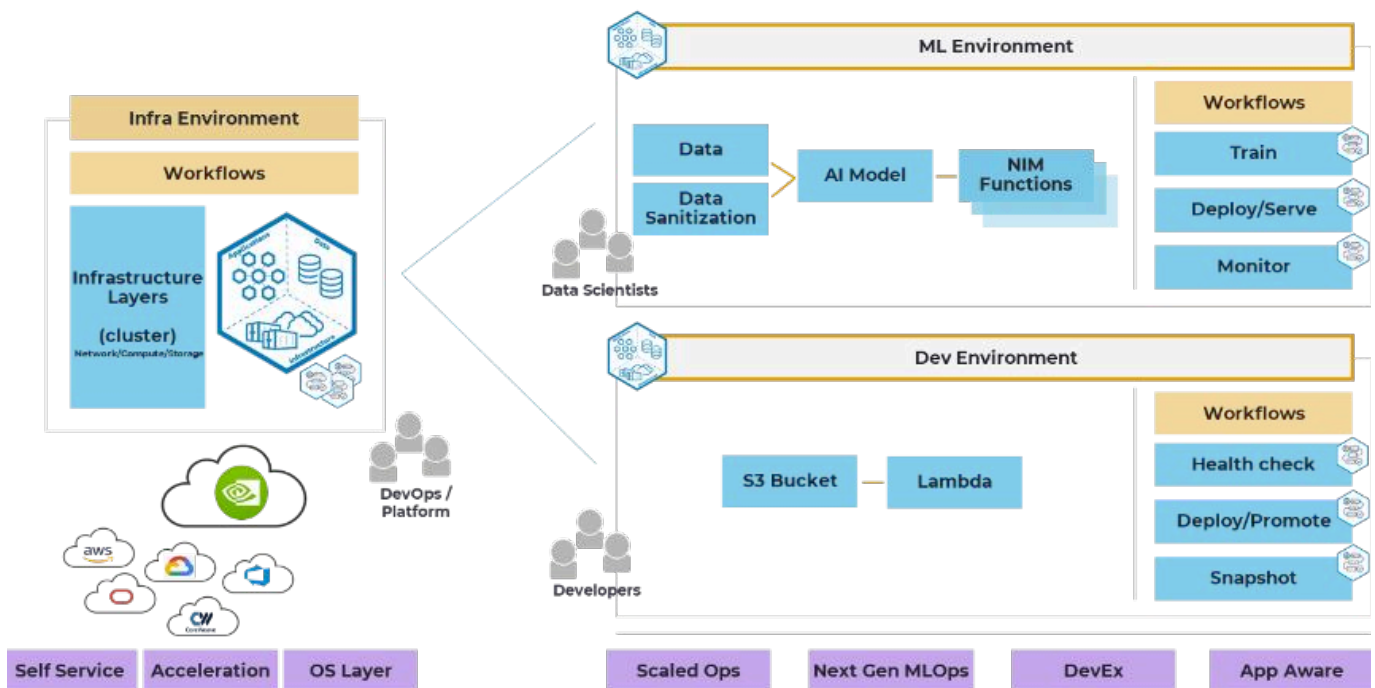
Driving Monetization, Scale, and Collaboration for AI Stacks and GPUs with Quali | TORQUE

In the rapidly evolving AI landscape, organizations face mounting pressure to maximize their AI investments, particularly in GPU technologies and related infrastructure. Torque by Quali is designed to address this challenge by enabling faster monetization, scaling MLOps, and enhancing ecosystem collaboration for AI stacks and GPUs. Torque transforms the complexity of infrastructure management into a competitive advantage, helping you unlock the full potential of your AI ecosystem.

Accelerating Monetization of GPU Investments

Torque accelerates monetization by automating the provisioning and management of AI workloads, from infrastructure setup to model deployment.

By optimizing the allocation and usage of GPU resources and infrastructure, Torque reduces idle time and improves utilization rates, directly addressing the scarcity and complexity challenges associated with GPU investments



- Automation and Elimination of Provisioning Overhead: Automatically provisions and manages GPU clusters, eliminating the delays and costs associated with manual setup and configuration.

- **Smart Allocation Against Policies:** Dynamically allocates GPU resources based on predefined policies and real-time demand, ensuring that GPUs are used where they deliver the highest value.
- **Self-Service and Just-in-Time Provisioning:** Empowers teams with self-service access to GPU resources, enabling just-in-time provisioning that avoids over-provisioning and maximizes the efficient use of infrastructure.

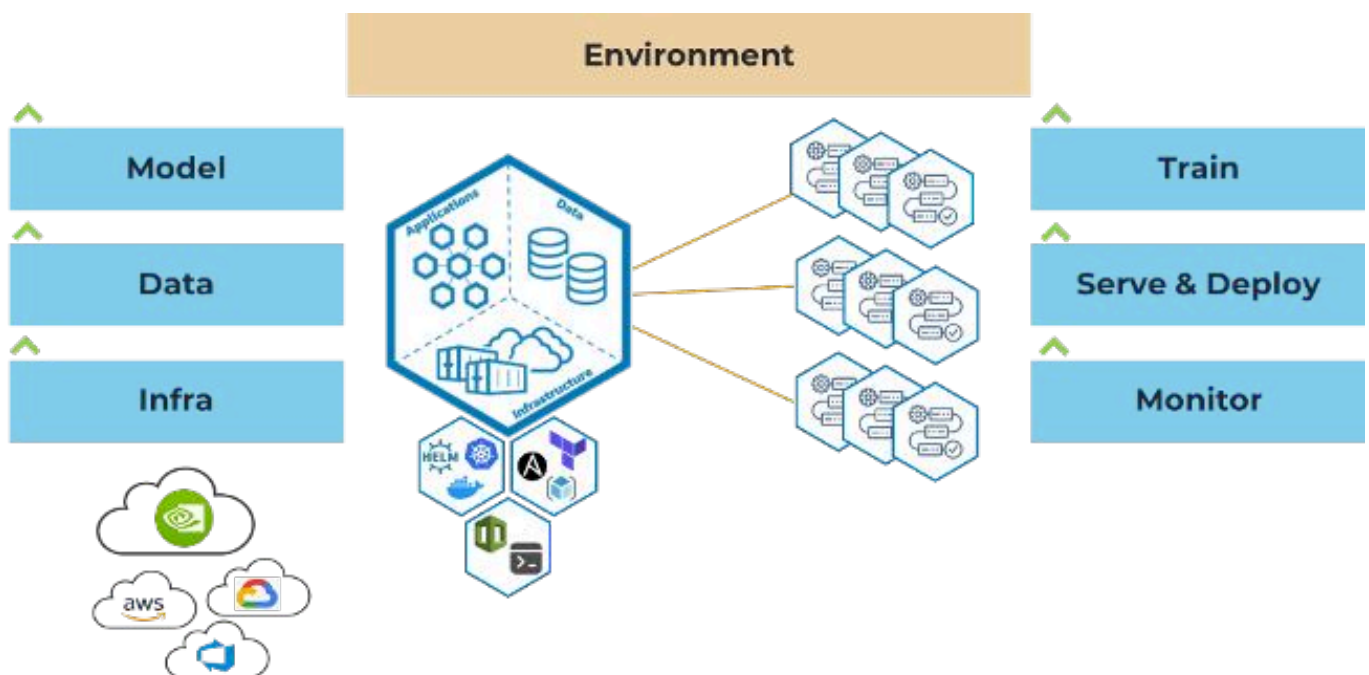
By streamlining GPU allocation and usage, Torque enables organizations to quickly capitalize on their GPU investments, reduce time-to-market, and maximize returns through smarter, more efficient infrastructure management.

Torque empowers businesses to scale AI workloads, driving higher demand for AI stacks and GPU infrastructure, which in turn accelerates AI innovation and market leadership.

Scaling MLOps for the Entire ML Lifecycle

Torque scales MLOps by providing managed infrastructure for the full machine learning lifecycle—from development and testing to serving and retraining models. It simplifies the infrastructure management needed to support complex AI workloads, allowing data scientists and developers to focus on innovation rather than infrastructure.

- **End-to-End Lifecycle Management:** Supports the entire ML pipeline, from model training to serving and maintenance, accelerating development and reducing operational overhead.
- **Hybrid and Multi-Cloud Flexibility:** Enables consistent operations across hybrid and multi-cloud environments, ensuring scalability and adaptability to diverse infrastructure needs.



Enhancing Ecosystem Collaboration




Torque facilitates collaboration across the AI ecosystem by providing access to pre-configured, full-stack environments, integrating applications, modules, and services from partners. This approach is similar to the value Torque delivers to Cisco DevNet, enhancing the partner ecosystem and driving new opportunities for growth.




- **Ready-Made Environments:** Offers out-of-the-box environments pre-configured with ecosystem tools and applications, reducing setup time and promoting faster adoption.
- **Partner Integration:** Enables seamless integration with partner technologies, creating new pathways for collaboration and joint go-to-market strategies.

Accelerating Development over GPU and Hybrid Stacks

Torque bridges the gap between complex infrastructure and the development teams dependent on it, streamlining the process of developing over GPU and hybrid stacks. It simplifies the deployment, monitoring, and scaling of environments, enabling faster and more efficient AI-driven development.

- **Developer-Centric Approach:** Provides intuitive, self-service tools for developers to quickly provision, test, and deploy environments without waiting for IT support.
- **Optimized Resource Utilization:** Dynamically allocates resources based on real-time demand, ensuring optimal use of GPU and hybrid infrastructure and preventing resource bottlenecks.

Category	Benefits of Torque	Impact on Organizations
 Scalable Infrastructure Management	Facilitates dynamic adjustment of environment configurations based on demand.	Supports scalable growth without compromising on performance or reliability.
 Control Over Multi-Cloud Environments	Provides centralized management across various cloud platforms and on-premises environments.	Enhances consistency and control, reducing complexity and risk.
 Resource Optimization	Ensures precise allocation, minimizing over-provisioning and resource waste.	Increases operational efficiency and maximizes IT investments.

Category	Benefits of Torque	Impact on Organizations
 <p>Cloud Cost Management</p>	Offers detailed insights into cloud usage and costs, enabling effective financial oversight.	Helps in reducing unnecessary expenses by optimizing resource sizing and runtimes.
 <p>Integrated Development & Operations</p>	Integrates with CI/CD pipelines for seamless development and operational workflows.	Supports continuous delivery and agile practices, accelerating time-to-market.
 <p>Automated Governance & Compliance</p>	Automates compliance with security, network, and data policies, reducing manual governance efforts.	Ensures environments are secure and compliant with minimal manual intervention.

Combining Infrastructure Platform Engineering and Torque’s Environments as Code offers a powerful solution for infrastructure complexity, development velocity, and operational efficiency. By addressing core IT challenges such as scalability, control, resource optimization, and cost management, Torque not only enhances operational efficiency but also provides a strategic advantage in a competitive market.