

Virtana + Nutanix: The Full-Stack Command Center for Your AI Workloads

Product: Nutanix Enterprise AI

Solution Overview

As organizations move from AI pilot programs to full-scale production, infrastructure complexity grows exponentially. AI workloads demand high-performance ation.

Nutanix Enterprise AI (NAI) provides a scalable, AI-ready infrastructure foundation. However, sustaining performance, optimizing GPU investments, and preventing bottlenecks requires deep, AI-aware observability.

Virtana Platform extends Nutanix with intelligent, full-stack visibility across the entire AI stack, from LLM applications and token usage down to GPU thermals and storage IOPS.

This integrated approach accelerates AI adoption while keeping performance high and costs predictable.

The Platform Powering the AI Factory

When you deploy Virtana AIFO on Nutanix, you aren't just getting a monitoring tool; you are gaining access to the full Virtana Platform intelligence layer. This includes:

- **System Dependency Graph:** Our dynamic mapping automatically discovers the relationships between your LLMs, Nutanix Kubernetes clusters, and NVIDIA GPUs.
- **Hybrid Infrastructure Observability:** Get deep visibility that continues beyond the service boundary, covering Nutanix AHV, storage, and network paths.
- **MCP Server Integration:** Seamlessly connect and pull data from across your entire environment without forcing a "rip-and-replace" of your current tools.
- **Agentic AI & Automated Response:** Don't just find problems; fix them. Our platform can approve and trigger actions like scaling GPU resources or restarting pods, keeping you in the loop while closing the gap on downtime.

What You Can Achieve with Virtana and Nutanix

Complete Stack Visibility

Monitor everything from AI workloads and GPUs to storage layers in one place.

GPU Efficiency

Track power, thermal trends, and capacity to ensure you are getting the most out of your hardware.

Token Tracking

Correlate your LLM token consumption directly with infrastructure costs.

Smart Troubleshooting

Use automated intelligence to find the root cause of issues before they impact users.

Future-Proof Planning

Use predictive analytics to forecast exactly when you will need more capacity.

Failure Analysis

Review training job failures across multiple iterations to stop recurring bottlenecks in their tracks.

Common Use Cases

Unified AI Infrastructure Command

- AI workloads create erratic spikes in GPU, storage, and network demand that traditional tools miss. Virtana bridges this gap by correlating GPU health and memory pressure directly with application-level metrics.
- Outcome:** *Faster troubleshooting, reduced alert noise, and a single source of truth for infrastructure and AI teams.*

Maximizing Every Token and Every Watt

- High-performance clusters are too expensive to sit idle. Our platform detects “noisy neighbors” and maps token-to-infrastructure costs to ensure every dollar spent on Nutanix and NVIDIA hardware drives measurable ROI.
- Outcome:** *Higher GPU utilization, lower cost per training job, and sustainable AI scaling.*

Hybrid Topology & Dependency Mapping

- Whether on-prem, at the edge, or in the cloud, Virtana automatically maps your entire ecosystem, ensuring you understand exactly how an infrastructure hiccup impacts a downstream model.
- Outcome:** *Instant impact analysis and the end of “app vs. platform” finger-pointing during critical disruptions.*

Agentic AI & Automated Response

- Move from reactive to autonomous with Agentic AI workflows and MCP Server support. Define operational policies in natural language to trigger automated remediations—like pod restarts or resource scaling—with integrated ITSM approvals.
- Outcome:** *Self-healing infrastructure that resolves incidents before they breach SLAs*

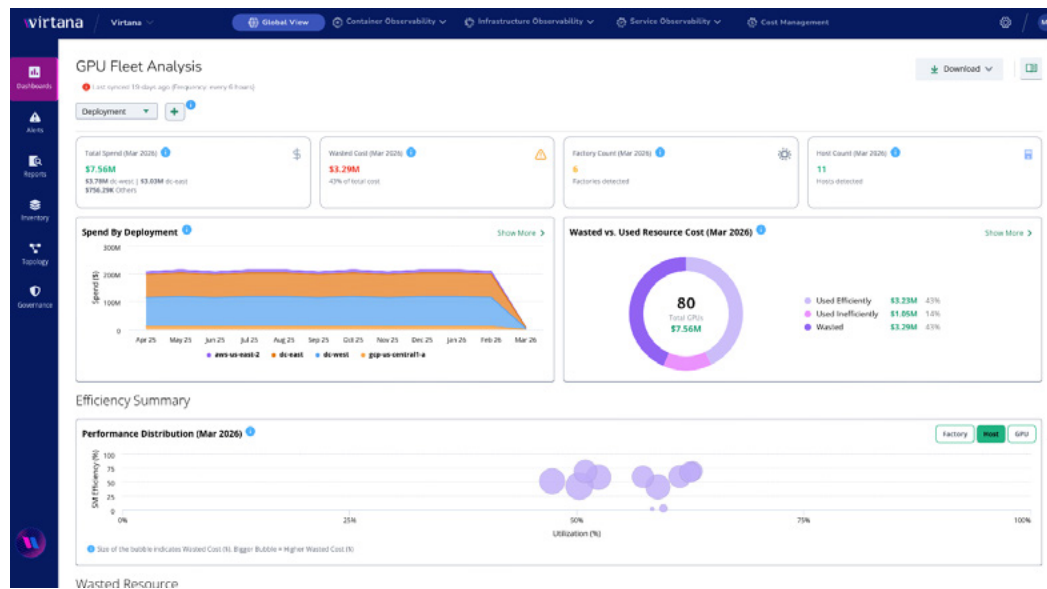


Figure 1: A customizable dashboard of a Unified Nutanix GPU fleet telemetry. Highlighting resource utilization, performance, departmental cost allocation, and ROI optimization.

Business Outcomes: Better Service Assurance

By unifying Nutanix Enterprise AI with the Virtana Platform, organizations can prioritize fixes by their actual impact on the business, not just alert volume.

Traditional Monitoring

Virtana AIFO on Nutanix

Stops at the service trace	Connects traces to Nutanix infrastructure
Manual slow root-cause analysis	Agentic AI and automated remediation
Fragmented point tools	Unified system dependency graph
Unpredictable GPU costs	Precise token and capacity optimization

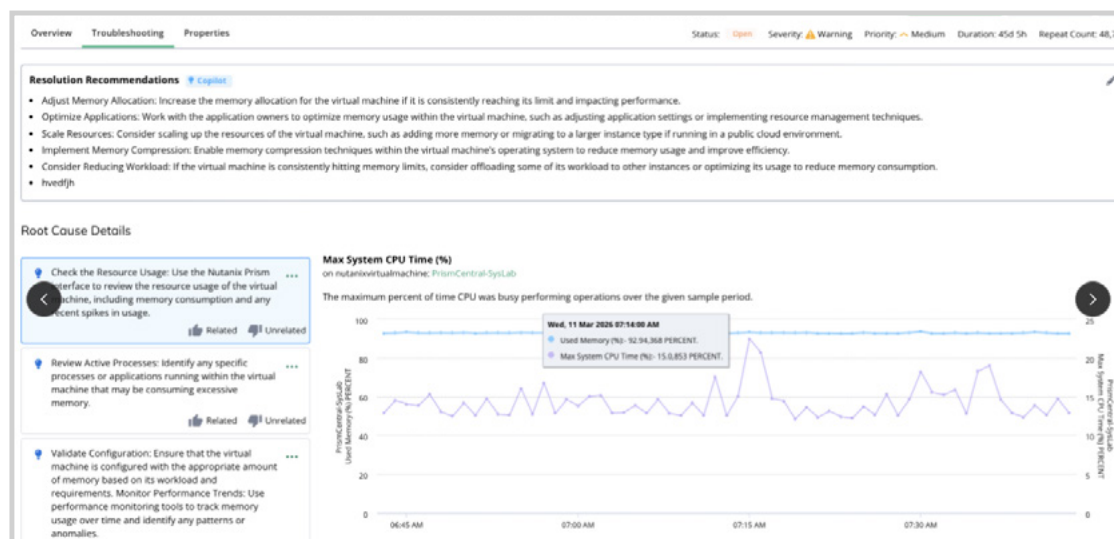


Figure 2: Agentic AI correlating multi-source signals to identify root cause and surfacing automated resolution workflows.

Take Control of Your AI Future

Nutanix Enterprise AI and Virtana Platform provide the visibility and intelligence required to run AI at scale. Stop guessing about your capacity and start optimizing your investment.

Ready to see it in action?

- Explore Nutanix: www.nutanix.com
- Discover Virtana AIFO: <https://www.virtana.com/platform/aifo/>
- Virtana Platform: <https://www.virtana.com/capabilities/>