

Yotta Shakti Cloud – Pioneering Sustainable AI Infrastructure

About Yotta

Yotta Shakti Cloud, recognized as the first NVIDIA® Cloud Partners (NCP), has emerged as a leader in delivering high-performance and sustainable AI cloud solutions. In today's fast-evolving AI landscape, Yotta has set a new benchmark by combining cutting-edge GPU technology, optimized networking, and energy-efficient power and cooling systems into an innovative infrastructure. This case study examines how Yotta's holistic approach to AI infrastructure, supported by the WEKA® Data Platform, delivers unparalleled performance while driving sustainability.

The Challenge:

High Performance with Sustainability

AI workloads, particularly in machine learning (ML) and deep learning, demand significant computing power and data throughput. Yotta recognized that traditional cloud infrastructure was ill-suited for managing these intensive demands efficiently, often requiring extensive physical resources, high power consumption, and complex operations. The challenge was to create an infrastructure that could balance **speed, performance, and sustainability** while being flexible and scalable for a wide variety of AI applications.

The Solution:

Yotta's Holistic AI Infrastructure

Yotta's solution involved designing an ecosystem where every component—GPUs, networking, power, cooling, and storage—was optimized for performance and sustainability. By leveraging **NVIDIA GPUs**, a high-speed network architecture, and energy-efficient data center designs, Yotta reduced the physical footprint, enhanced power usage efficiency, and minimized cooling requirements, all while delivering peak AI performance.



Challenges

- High power consumption and resource demands for AI workloads.
- Inefficient traditional infrastructure for scalable AI applications.
- Complex operations with high energy and cooling needs.

Solution

- WEKA Data Platform provided fast, efficient data access for AI workloads.
- Optimized infrastructure with NVIDIA GPUs and high-speed networking.
- Reduced data center footprint with optimized power and cooling.

Benefits

- **High performance:** Accelerated AI workloads with reduced bottlenecks.
- **Sustainability:** Lower power consumption and environmental impact.
- **Simplified operations:** Seamless scaling and easier management.

A key element of Yotta's infrastructure is its **storage layer**, where the **WEKA Data Platform** plays a vital role. WEKA provides the high-speed, low-latency data access needed to process AI workloads efficiently. Integrating the WEKA Data Platform's intelligent data tiering and zero-copy capabilities into Yotta's design further streamlined data movement, improving performance while keeping operational costs and complexity low.

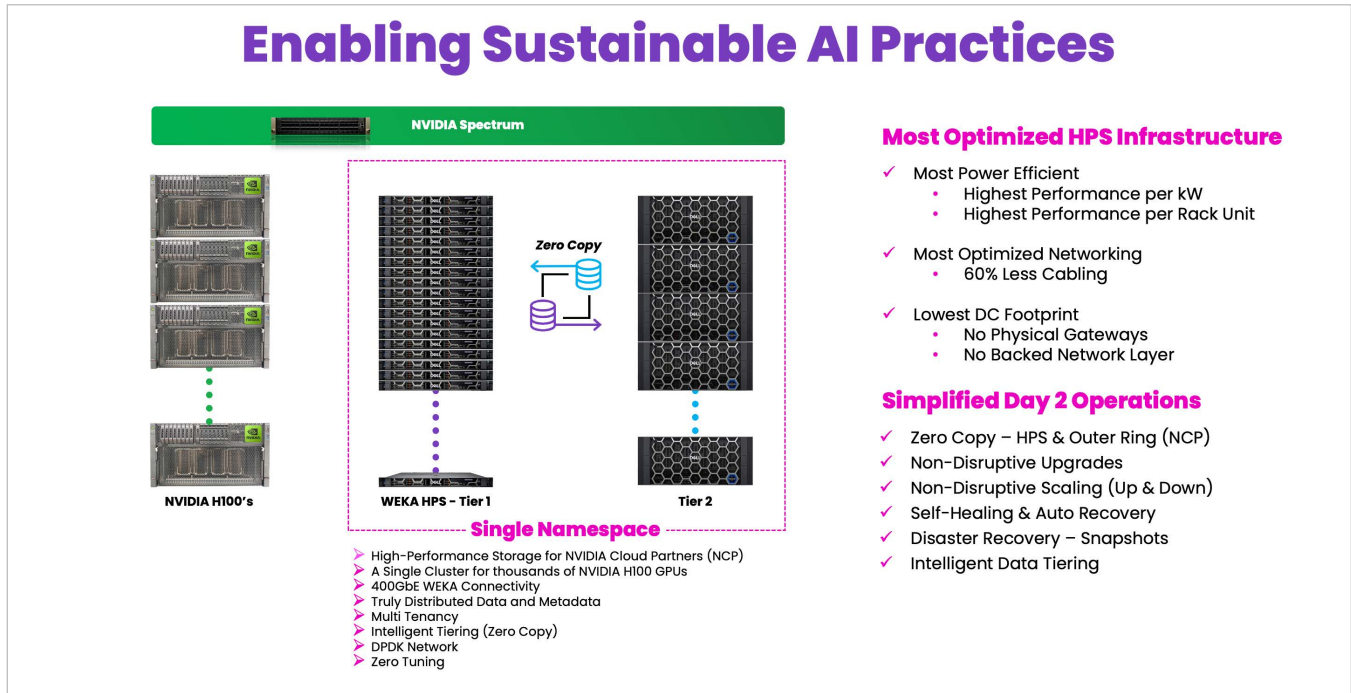


FIG. 1 Enabling Sustainable AI Practices

Sustainability Achievements

Yotta's commitment to sustainability is evident in several key design decisions:

- Power Efficiency:** Yotta's infrastructure maximizes performance per kilowatt, reducing energy consumption while ensuring high performance for AI workloads. This focus on power efficiency aligns with Yotta's sustainability goals, minimizing environmental impact.
- Optimized Networking:** By utilizing 60% less cabling in its infrastructure to decrease network complexity, power needs, and cooling requirements, Yotta enhances operational efficiency and sustainability, while WEKA's L6 networking, combined with DPDK for flow control and congestion control, optimizes network performance with low latency and high throughput.
- Reduced Data Center Footprint:** Eliminating additional storage gateways and backend network layers allowed Yotta to minimize the amount of physical hardware needed, creating a smaller, more sustainable data center environment.
- Linear Scaling:** Yotta's infrastructure scales proportionally as GPU clients increase, ensuring that AI services can expand without significantly increasing power or cooling demands. This efficient scaling avoids overprovisioning and conserves energy.

WEKA's Role in Yotta's Success

The **WEKA Data Platform** is a cornerstone of Yotta's infrastructure, providing the speed and agility needed to support AI and ML workloads. By enabling seamless data transfers between storage and AI processing units, WEKA ensures that Yotta's infrastructure can handle even the most data-intensive tasks without performance bottlenecks. The platform's compatibility with bare metal and **Platform-as-a-Service (PaaS)** environments provides flexibility, allowing Yotta to streamline its operations and deliver a simplified, cost-effective experience for its tenants.

Additionally, WEKA's use of optimized x86 hardware servers enables Yotta to avoid the limitations of fixed, appliance-based architectures, allowing them to offer high-performance AI services without driving up costs. This combination of performance, flexibility, and operational simplicity supports Yotta's long-term strategy of providing sustainable AI cloud infrastructure.

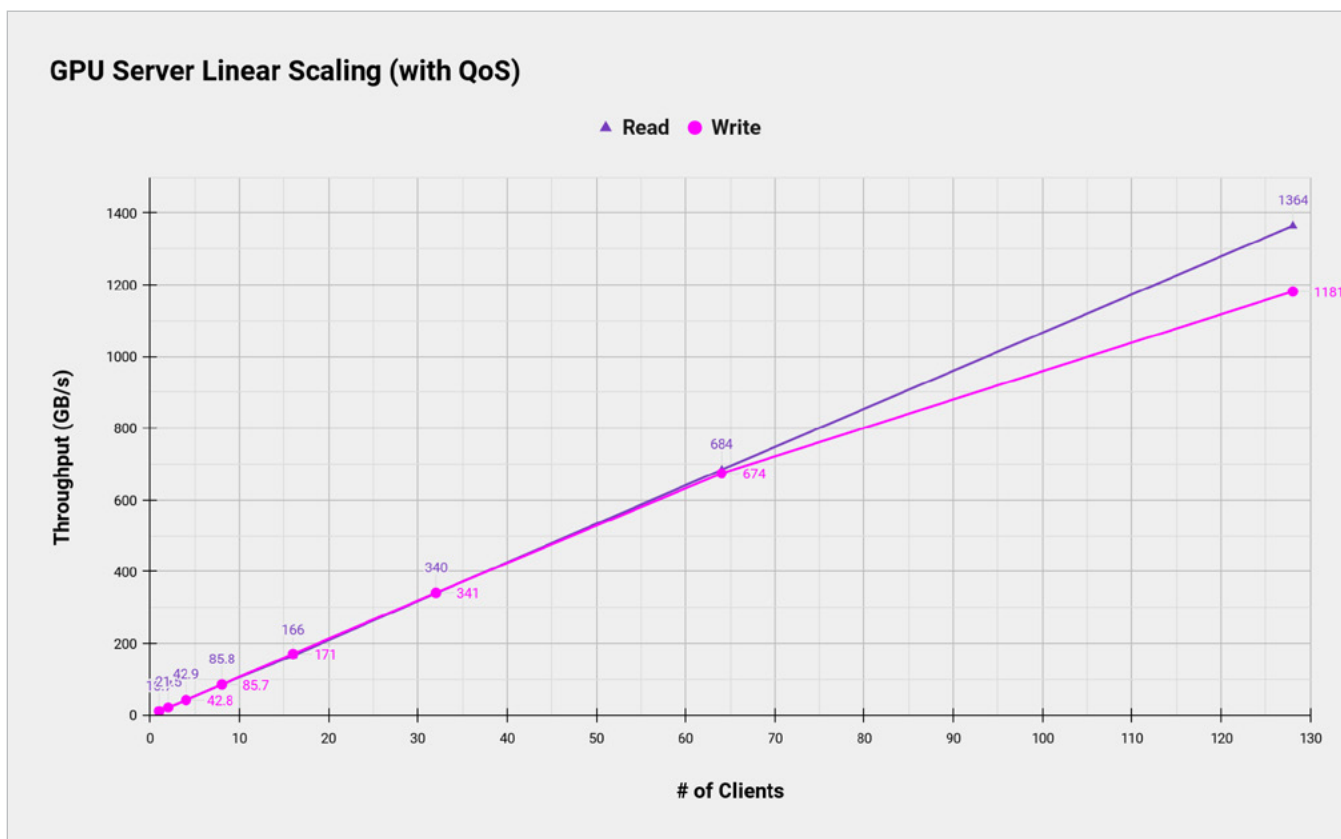


FIG. 2 GPU Server Linear Scaling (with QoS)

Results: A Sustainable, High-Performance AI Cloud

Through its innovative design, Yotta Shakti Cloud has achieved the following:

- **Peak Performance:** Yotta's infrastructure accelerates AI workloads, delivering high performance through NVIDIA GPUs, efficient networking, and optimized storage with WEKA.
- **Sustainability Leadership:** Yotta's power-efficient design, reduced data center footprint, and streamlined networking have significantly lowered its environmental impact while maintaining high operational efficiency.
- **Simplified Operations:** WEKA's integration has helped Yotta offer non-disruptive upgrades, seamless scaling, and automated storage provisioning, simplifying daily operations and reducing complexity for tenants.

Setting a New Standard for Sustainable AI Cloud Infrastructure

Yotta Shakti Cloud's success stems from its holistic approach to building a sustainable, high-performance AI cloud infrastructure. As an NVIDIA Cloud Partner, Yotta has demonstrated excellence in combining cutting-edge technology with intelligent design to optimize performance while minimizing environmental impact. The strategic integration of WEKA's Data Platform has played a crucial role in ensuring Yotta's infrastructure can meet the demands of modern AI workloads efficiently, sustainably, and cost-effectively.

By addressing the challenges of power consumption, data center complexity, and scaling, Yotta continues to pioneer the future of sustainable AI infrastructure, setting a new standard for performance and environmental responsibility in the cloud.



About the WEKA Data Platform

The WEKA[®] Data Platform removes the barriers to data-driven innovation through its advanced software architecture optimized to solve complex data challenges and streamline the data pipelines that fuel AI, ML, and other modern performance-intensive workloads.

The design philosophy behind the WEKA[®] Data Platform was to create a single architecture that runs on-premises or in the public cloud with the performance of all-flash arrays, the simplicity and feature set of network-attached storage (NAS), and the scalability and economics of the cloud. Whether on-premises, in the cloud, at the edge, or bursting between platforms, WEKA accelerates every step of the enterprise AI data pipeline – from data ingestion, cleansing, and modeling to training validation or inference.

Mind-bendingly fast. Seductively simple. Infinitely scalable. Sustainable. Spanning edge, core, hybrid, and cloud. The WEKA Data Platform helps to overcome complex data challenges and accelerate next-generation workloads to unleash your organization's imagination, creativity, and potential.

