LIQID

UltraStack 30 - AMD MI350P Datasheet

Why Liqid

Delivers the highest AMD GPU density available, up to 30 AMD MI350P GPUs in a single server, for superior AI inference tokenomics: more tokens per dollar and more tokens per watt.

Key Advantages

» **Breakthrough Scale-Up Density:** Up to 30× AMD Instinct MI350P GPUs in one server, up to 4× the GPU density of legacy servers.

» **Superior Tokenomics:** Up to 3.7× tokens/second, 2.1× tokens/dollar, and 1.8× tokens/watt versus legacy designs.

» **Lower Cost & Power:** Up to 65% lower deployment cost and 50% lower power for large-model deployments.

» **CXL memory-pooling ready:** Offers a clear path to terabyte-scale shared memory for KV cache and other memory-intensive workloads.

Key Features:

» 30× AMD Instinct™ MI350P PCIe GPUs
» 4.3 TB aggregate HBM3E memory
» 69 PFLOPS FP8 AI performance
» Liqid GPU pooling over PCIe fabric
» Native Kubernetes support
» CXL memory-pooling ready
» Peer-to-peer GPU communication

Contact Information

Liqid Inc.
11400 Westmoor Circle, Suite 225
Westminster, CO 80021
office: +1 303.500.1551 email: sales@liqid.com

Liqid UltraStack 30 AMD MI350P

Massive AI Scale-Up with AMD Instinct™ MI350P GPUs

AI inference workloads are growing relentlessly in size and complexity, while the servers meant to run them haven't kept pace. Most traditional servers top out at 8 GPUs, and roughly 80% of the market is limited to just 2–4, forcing organizations to spread large models across many systems. The result is stranded capacity, added networking and software overhead, and a rising cost for every token generated.

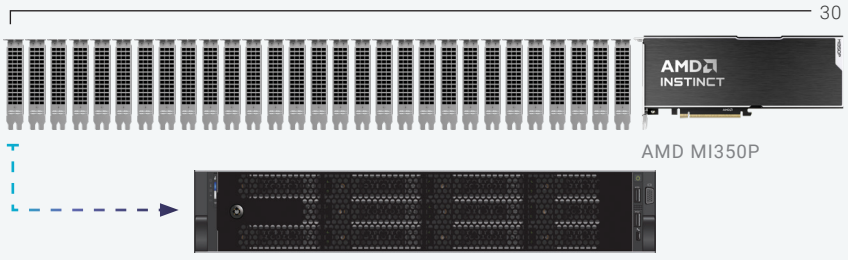
Liqid UltraStack 30 AMD is built on Liqid's GPU pooling platform and powered by AMD Instinct™ MI350P Series GPUs. It scales up to 30 GPUs to a single server, creating a large, pooled accelerator fabric with 4.3 TB of aggregate HBM3E and 69 PFLOPS of FP8 performance. Users that were forced to deploy large or multiple LLMs across four or more servers can now deploy on a single AMD EPYC CPU-based server.

Liqid Matrix software attaches GPUs and orchestrates them across workloads in real time over a high-speed PCIe fabric, keeping every accelerator inside a single node. Liqid UltraStack AMD eliminates multi-node overhead, delivers predictable linear scaling, and drives GPU utilization toward 100%, all with native Kubernetes support for multi-model serving.

The outcome is breakthrough inference economics: up to **3.7× more tokens per second, 2.1× more tokens per dollar,** and **1.8× better tokens per watt** than legacy multi-server designs, with up to 65% lower deployment cost and 50% lower power.

AI Inference Scale-Up Node - On a Single Server

30x AMD MI350P GPUs **4.3 TB** HBM3E **69 PF** FP8



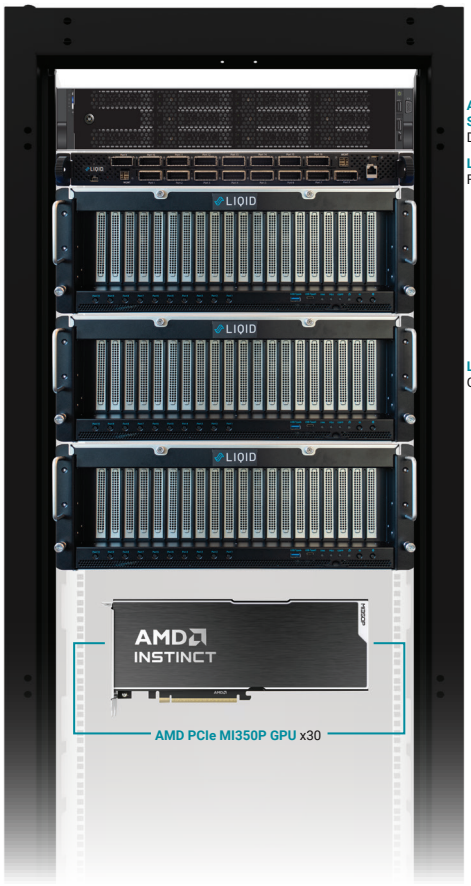
Built For

Large LLM Inference ▪ Long-Context AI ▪ Multi-Model Serving ▪ Agentic AI ▪ RAG & Vector Search ▪ Scientific Computing (HPC)

Ideal for NeoCloud and AI service providers, enterprise AI teams, and HPC and research institutions deploying AI at scale on-premises.

UltraStack

Specification	UltraStack 30 (AMD Instinct™ MI350P)
Description	30-GPU AI Inference Scale-Up Node
Host Server	Dual-socket AMD EPYC™ 9005 Series
Management Appliance	1× Liqid Director 1U
GPUs	30× AMD Instinct™ MI350P PCIe
AI Performance	69 PFLOPS (FP8)
GPU HBM Memory	4.3 TB HBM3E
GPU Pooling Fabric	Liqid PCIe Gen 5 fabric with Liqid HBA
Expansion Chassis	Liqid EX-Series expansion chassis
NVMe Storage	Liqid NVMe SSD (recommended)
Networking	High-speed Ethernet / InfiniBand NICs & DPUs (recommended)
Orchestration	Liqid Matrix® software — native Kubernetes support
Memory Expansion	CXL memory-pooling ready
Total System Power	~22 kW



Storage and networking devices are recommended and can be customized or omitted. Power estimate reflects typical AI inference load; actual consumption may vary by configuration and workload.

Efficiency Proof Points

Key Metric	Value
Performance per kW	~3.14 PFLOPS/kW
HBM per kW	~195 GB/kW
HBM per PFLOP	~62 GB/PFLOP

Contact Information

Liqid Inc.
11400 Westmoor Circle, Suite 225
Westminster, CO 80021
office: +1 303.500.1551 email: sales@liqid.com