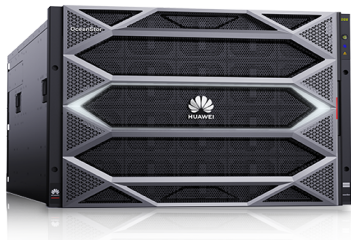


Huawei OceanStor A800 AI Flagship Storage

Vlajková IoT AI infrastruktury pro trénování velkých modelů (LLM) a masivní GPU cluster



Huawei OceanStor A800 je nejvýkonnější distribuovaný souborový systém navržený speciálně pro potřeby generativní umělé inteligence, velkých jazykových modelů (LLM) a autonomních systémů. Zajišťuje nepřetržité krmení desítek tisíc GPU jader bez starvation efektu a dosahuje bezkonkurenční propustnosti až 500 GB/s a 10 milionů IOPS v jediném 8U šasi. Plná podpora NVIDIA GPUDirect Storage (GDS), 200G/400G RoCEv2 a InfiniBandu.

Klíčové technologické přednosti & Architektura

- Rekordní propustnost:** 500 GB/s šířka pásma a přes 10 000 000 IOPS v jediném 8U šasi (2 uzly, 8 CPU Kunpeng 920).
- Nulový GPU Starvation:** Využití výpočetního výkonu GPU clusteru > 95 % a zkrácení doby trénování velkých modelů.
- Bleskový Checkpointing:** Zkrácení ukládání a obnovy mezipřímých modelů z desítek minut na jednotky minut (-80 %).
- Podpora moderních AI datových struktur:** Nativní handling pro vektorové indexy (RAG), tenzory a Key-Value cache.
- Škálovatelnost:** Až 512 uzlů v clusteru s podporou AI výpočetních klastrů sítí až 100 000 xPU.

Podrobná technická specifikace

Parametr / Vlastnost	Technická specifikace
Formát šasi	8U high-density šasi (integruje 2 vysoce výkonné úložné nody)
Šířka pásma (Propustnost)	Až 500 GB/s na 8U šasi
Maximální IOPS (Random 4K)	Přes 10 000 000 IOPS na 8U šasi
Procesory & RAM	4x Kunpeng 920 per uzel (8 CPU / 8U), až 2 TB systémové RAM na šasi
Disková média	Až 64x Palm-sized NVMe SSD (kapacity 15.36 TB, 30.72 TB a 61.44 TB na disk)
Škálovatelnost	Až 512 uzlů v plně symetrickém clusteru
Síť & Konektivita	PCIe 5.0, 25/100/200 GbE, 100/200/400 Gbps RoCEv2, 200G InfiniBand, NVIDIA GDS
Souborový systém	Distribuovaný paralelní souborový systém OceanFS s architekturou POSIX / MPI-IO
AI akcelerace	AI Data Fabric Omni-Dataverse, akcelerace Time-To-First-Token (TTFT) o 78%