[ENTERPRISE TECHNICAL BRIEF] | US GPU SOLUTIONS PROVIDER • NHN CLOUD PARTNER

GLHF Cloud - NVIDIA HGX B300 / B200 / H100 Archite

Enterprise GPUaaS for the US Market. Powered by NHN Cloud Infrastructure. Dedicated Frontier AI Supercomputi

01 // FRONTIER COMPUTE ARCHITECTURE (BLACKWELL ULTRA)

NVIDIA HGX B300 Supernode Reference Specification

- Node Configuration: 8x NVIDIA B300 Blackwell Ultra GPUs per validated HGX chassis
- Memory Subsystem: Up to 2.1 TB ultra-high-bandwidth HBM3e per 8-GPU node
- Native Precision Support: FP4, FP8, BF16 for frontier training & long-context reasoning
- Fabric Architecture: Dual-Rail Quantum InfiniBand / RoCEv2 (GPUDirect RDMA + SHARP)
- Non-Blocking Throughput: 800 Gb/s per GPU (3.2 Tbps rail fabric) with GPUDirect RDMA

---

02 // STRATEGIC INFRASTRUCTURE PARTNERSHIP (NHN CLOUD)

Tier-1 Reliability. Engineered for Mission-Critical AI.

- NHN Cloud Certified Plant: Carrier-neutral, Tier-3+ physical data centers with redundant power delivery and direct liquid-to-chip thermal dissipation preventing thermal throttling.
- Zero Egress Predictability: Flat-rate economic predictability with 100% free data movement across internal clusters and zero network egress penalties.
- Enterprise Compliance Posture: End-to-end single-tenant isolation aligned with SOC 2, ISO 27001, and strict institutional data governance standards.

---

03 // SUSTAINED MFU GOODPUT & TAILORED CONSUMPTION MODELS

Strategic Capacity Aligned to Your Engineering Roadmap

- Dedicated Reserved Superclusters: Single-tenant B300/H100 clusters reserved for 3 to 36-month terms with custom SLA commitments and deterministic hardware placement.
- Flexible Capacity Blocks: Dedicated multi-node allocations provisioned for sprint-based post-training, RL rollouts, and evaluation batches without spot instability.
- Co-Engineered Deployment: White-glove cluster burn-in, Slurm/Kubernetes orchestration, and direct escalation to senior infrastructure architects.

---

04 // ARCHITECTURAL REVIEW & CAPACITY RESERVATION

Connect Directly with GLHF Cloud Enterprise Infrastructure Engineering

- Architecture Advisory: peterk@glhfcloud.com
- Enterprise Capacity & Custom Term Agreements: peterk@glhfcloud.com
- NDA Benchmark Telemetry Packs (Sustained MFU, All-Reduce Latency, Burn-In Reports) on request.

Confidential Enterprise Technical Overview. Custom agreements only; no spot-market allocation.