FULL RACK MICRO-OT-CLOUD CLUSTER

NE0–RACK–31U

The full rack assembles the entire OT-Cloud in one sovereign footprint: 3U Micro-OT-Cloud clusters distribute edge, profile, file, zone and cache services; 1U Pentascale storage nodes hold profiles and logs; and 1U NVIDIA Grace-Hopper coherent-memory systems run OT-AI. Rack fill is dictated by data-center thermal design.

3U cluster · 8 processors | **NVIDIA GH200 Grace-Hopper** | **Pentascale NVMe storage** | **~31U example fill**

SYSTEM COMPONENTS

3U Micro-OT-Cloud Cluster

8 PROCESSORS

- CPU: 8 × single Intel® Xeon® 6300/2400/2300 or AMD EPYC™ 4005/4004 (LGA-1700, up to 95W TDP)
- Memory: up to 128GB ECC UDIMM DDR5-4400 per node (4 DIMM sockets × 8)
- Drives: 2 × front hot-swap 3.5" SATA3 or hybrid 2.5" SAS/SATA3/NVMe U.2 per node
- PCIe: 8 × PCIe 5.0 x8 LP + 8 × PCIe 4.0 x8 Micro-LP
- Power: 2200W redundant Titanium high-efficiency

1U Pentascale Storage

NVIDIA GRACE CPU SUPERCHIP

- Single NVIDIA Grace™ CPU Superchip — up to 144 Arm v9 cores, 600W TDP
- Up to 960GB LPDDR onboard coherent memory
- 16 front hot-swap E3.S (7.5mm) NVMe bays · 8 × E1.S · 2 × M.2
- 2 × PCIe 5.0 x16 FHHL + 1 AIOM (OCP 3.0 SFF)
- Redundant 1600W Titanium; BlueField®-3 / ConnectX®-7 support

1U CPU+GPU Coherent Memory (AI/HPC)

NVIDIA GH200 GRACE-HOPPER

- Up to 2 × NVIDIA GH200 Grace-Hopper Superchips (72-core Arm CPU + H100 Tensor Core GPU)
- Up to 96GB HBM3 + 480GB LPDDR5X per superchip
- NVLink® Chip-2-Chip (C2C) high-bandwidth, low-latency interconnect
- 8 hot-swap E1.S + 2 M.2 · 2 × PCIe 5.0 x16
- Air and liquid cooling options

EXAMPLE RACK FILL — ~31 UNITS

3U Micro-OT-Cloud Cluster	**4 units**	48 cluster processors — OT-Edge, OT-Profile, OT-File, OT-Zone, OT-Cache
1U Pentascale Storage	**5 units**	OT-File and OT-Profile
1U Pentascale Storage (OT-AI)	**6 units**	Logs, profile utilization, OT-Cloud observations
1U CPU+GPU Coherent Memory	**8 units**	OT-AI implementation (load-dependent)