Huawei Atlas 950 SuperPoD

1024x Ascend 950DT | Da Vinci v3 | pod | 2026

OVERVIEW

1,000	98.3 TB	4096.0 TB/s	100.0 kW
FP8 PFLOPS Dense	Aggregate Memory	Aggregate Bandwidth	Aggregate Power

PERFORMANCE METRICS

Precision	Peak (PFLOPS)	Per GPU (TFLOPS)	TFLOPS/W
FP8	1000.0	976.6	–
FP4	2000.0	1953.1	–

All figures are dense. Vendors commonly headline the sparse number, which is twice the dense one.

CONFIGURATION

Accelerator	Ascend 950DT
Count	1024
Memory per GPU	144 GB
Bandwidth per GPU	4.0 TB/s

POWER & EFFICIENCY

Aggregate Power	100.0 kW
TDP per GPU	—
FP8 Efficiency	10.00 TFLOPS/W
Aggregate Bandwidth	4096.0 TB/s

SYSTEM DETAILS

Vendor: Huawei
Family: Atlas
Architecture: Da Vinci v3
Form Factor: pod
Release Year: 2026
Process Node: —

INTERCONNECT

UnifiedBus 2.0 (Lingqu) all-optical, 1.72 PB/s aggregate scale-up, 64 x 1.68 TB/s bidirectional per cabinet

NOTES

Huawei's current product specification: 16 compute cabinets plus 4 UnifiedBus interconnect cabinets (44OU each), up to 1,024 Ascend 950DT, 1,024 x 96 GB on-chip memory at 4.0 TB/s, 1 EFLOPS mxFP8/FP8/HiF8 and 2 EFLOPS mxFP4, 1.72 PB/s total scale-up bandwidth, 256 TB globally addressable memory, 100 kW, fully liquid cooled. Offered in 64-card and 1,024-card configurations. Huawei's HC2025 keynote of 18 September 2025 described an 8,192-card configuration delivering 8 EFLOPS FP8 and 16 EFLOPS FP4; that figure does not appear in the current product table and is recorded here for reference. Huawei states no dense or sparse basis for any of these figures.