InfoBites, sponsored by Supermicro and NVIDIA | JULY 2025

Powering AI Factories: Scaling GenAI with Direct-to-Chip Liquid-Cooling

Transforming Datacenters from Cost Centers to Engines of Intelligence

Andrew Buss
IDC IT infrastructure EMEA

Luis Fernandes
IDC IT infrastructure EMEA

Direct-to-chip liquid-cooling is foundational to AI factory design. It supports 100kW+ racks and enables scalable and sustainable intelligence production.

Elements of a direct-to-chip liquid-cooled datacenter that turn it into an AI factory:

- Datacenter facilities-side plumbing, fluid, and heat exchanger
- Facilities-side liquid distribution manifolds and heat exchangers in datahalls
- Coolant distribution units (CDUs) in racks, rows, or datahalls
- Hybrid or 100% direct-to-chip cooling loop with **advanced technical cooling fluid**

The suitability of different air and liquid-cooling approaches based on rack-level power density

The key benefits of direct-to-chip liquid-cooling:

- Dramatically lowers cooling overhead and enhances TCO
- Maximum system performance due to the high heat capacity of the technical cooling liquid
- More efficient datacenter use of water
- Helps meet ESG commitments and reporting requirements
- Supports significantly higher GenAI compute density
- Quieter operation, improving workplace comfort
- Prevents hotspots and thermal throttling
- Processors that run more consistently at peak “boost” clocks

<15kW → **50kW** → **100kW+**

- Air cooling
- Rear-door heat exchanger
- Hybrid direct-to-chip liquid-cooling and rear-door heat exchanger
- 100% direct-to-chip liquid-cooling

InfoBites, sponsored by SUPERMICRO | NVIDIA.

Message from the sponsor

Supermicro and NVIDIA are redefining the economics of deploying AI factories. We offer state-of-the-art infrastructure solutions that address increased power and cooling challenges in modern AI datacenters. Additionally, significant savings can be achieved with direct liquid-cooling (DLC-2) for highly efficient generative AI datacenters.

For more information