SENSAKA

sensaka.com

GPU Rack Hot Spot Detection Guide

Why dense racks overheat locally, and the five signals that expose it

Why dense GPU racks develop hot spots

Concentrated power density

An accelerator chassis can draw more in 8U than a conventional rack drew in 42U

Airflow sized for a different rack

Containment, tiles, and CRAC placement designed around an average kW per rack

Synchronized cluster load

Distributed training pushes every node at once — heat arrives as a step change across a row

Mixed cooling in one room

Liquid-cooled accelerators, air-cooled PSUs, memory, NICs — fine on coolant, hot in the aisle

The five telemetry sources (read together, time-aligned)

- Inlet and outlet temperature per chassis — the delta is the first usable signal
- BMC thermal sensors: board, CPU, memory, accelerator temps, fan state, throttling — out-of-band, so data keeps arriving when the OS hangs
- Per-rack power draw from PDUs and branch circuits — where heat is actually produced
- Liquid loop temperatures: coolant supply/return, pressure differential, flow
- Room and zone environmental sensors — is one machine unusual, or the whole row?

How to read the signals

1. **Compare, don't threshold**
   A fixed alarm fires late and everywhere at once. Compare each chassis against its rack neighbors, each rack against its row.
2. **Put thermal and power on one timeline**
   Power first, heat after = load. Heat first without a power change = airflow or coolant.
3. **Watch throttling and fans, not just degrees**
   Sustained max fans and thermal throttling are the hardware already compensating — both appear before temperatures look alarming.
4. **Resolve to a location and an owner**
   Actionable = rack, U position, cooling zone, responsible team, affected jobs.

Separating cooling faults from workload changes

What you see	What it usually means
Rack power rises first, temperature follows	The workload changed
Temperature rises while power stays flat	Heat not being removed: airflow, fan, containment, or loop
Inlet temperature rising	Room or cooling-unit condition, not the server

Public knowledge from Sensaka · free to share with attribution
Source: sensaka.com/resources/gpu-thermal-hotspot-detection
sensaka.com