COLOCATIONQUOTES

AI and High-Density Colocation Checklist

32 checks for GPU clusters, neoclouds and other high-density deployments: power ramps, liquid cooling, structure, network, operations and contract terms.

How to use this. Tick each item once the provider has shown you evidence: documents, a live deployment or a site tour, not a sales deck. Anything you cannot tick is a question for the provider or a risk to price into the deal. For a full facility evaluation with scoring, use our data center checklist.

Power and ramp

	Check	Why it matters
☐	Maximum kW per cabinet in the space offered, today, with the cooling installed in that space today	Rack-scale GPU systems can draw about 120 kW per rack and newer systems more. Halls designed for air-cooled racks cannot host them.
☐	Power energized now versus planned, and for planned power, a signed utility agreement and an energization date	Capacity that exists only on a utility queue is the biggest delay risk in AI deployments.
☐	A ramp schedule that matches your GPU deliveries, with contractual ready-for-service dates	Paying for power you cannot use yet, or waiting for power while hardware sits in boxes, both burn money.
☐	Remedies if a ramp date is missed: credits, rent abatement or a right to exit	Without remedies, a missed date costs you and nothing else.
☐	Power delivery to the rack: busway or remote power panels, 415V availability and whip types	Rack-scale systems often use power shelves fed by several high-amperage whips per rack.
☐	Redundancy design for GPU halls, and whether it differs from the rest of the building	Some AI halls use less redundancy to lower cost. Know what you are getting and price it accordingly.
☐	How power is billed: committed kW, metered use or a take-or-pay commitment	Training loads swing hard. Overage and reset rules matter more than for steady enterprise racks.

Liquid cooling

	Check	Why it matters
☐	Liquid cooling running today, with a live customer deployment you can see	A plan to add liquid cooling is not liquid cooling.
☐	Which types are supported: rear door heat exchangers, direct-to-chip and immersion	Each has different facility water, floor and maintenance needs.

☐	Who owns, supplies, maintains and monitors the coolant distribution units (CDUs)	Unclear ownership is where outages turn into disputes.
☐	Facility water supply temperature, flow per rack and redundancy of the loop (ASHRAE liquid classes W17 to W45)	Your hardware vendor specifies an inlet range. The facility has to meet it at full load.
☐	Secondary loop ownership, water quality, filtration and the treatment program	Poor water quality clogs cold plates and voids warranties.
☐	Leak detection and automatic shutoff at the rack, and who is liable for a leak	One leak can damage far more than the rack it starts in.
☐	Residual air cooling for the heat liquid does not capture	Even liquid-cooled racks reject part of their heat to air.
☐	Commissioning of the liquid cooling with your actual hardware before go-live	Testing with load banks alone can miss problems your hardware will find.

Building and structure

	Check	Why it matters
☐	Floor or slab rating for rack-scale systems, in pounds per square foot and point load	A fully loaded rack-scale GPU system can exceed 3,000 pounds.
☐	Delivery path for integrated racks: dock height, door widths, ramps, freight elevator size and rating	Racks shipped fully built cannot be taken apart to fit a small door.
☐	Ceiling height and space for overhead busway, cable trays and liquid piping	High-density rows need all three overhead or under the floor.
☐	Staging space to unbox and test racks before they go on the floor	GPU racks arrive in large crates that do not belong on the data floor.

Network

	Check	Why it matters
☐	Space and cabling paths for the back-end GPU fabric between racks	GPU clusters run their own high-bandwidth fabric, often with many fiber runs per rack.
☐	Distance limits between racks in the cluster, and whether your rows can be contiguous	Fabric cable and optic choices depend on rack-to-rack distance.
☐	Carriers, cloud on-ramps and dark fiber for front-end traffic and data movement	Training data has to arrive, and models and results have to leave.
☐	Cross connect pricing and install times for high fiber counts	Per-cross-connect fees add up fast at cluster scale.

Operations and support

	Check	Why it matters
☐	Smart hands experienced with GPU servers, liquid-cooled racks and fiber at scale	General remote hands may not be trained to work on liquid-cooled systems.

☐	Spare parts storage near the cluster and how fast parts can reach the floor	GPU nodes fail. The time to swap one sets your cluster's availability.
☐	Monitoring you can see: power per rack, coolant temperatures and flow, and alarms	You need the same data the operator uses to spot trouble.
☐	Change process for work on shared cooling and power systems	Maintenance on a shared loop can affect every rack on it.

Contract and commercial

	Check	Why it matters
☐	A written responsibility matrix for power, cooling, CDUs, manifolds, hoses and fittings	Every component should have one named owner.
☐	SLA terms for liquid cooling: supply temperature, flow and response times, not just power and air temperature	Standard colocation SLAs were written for air-cooled racks.
☐	Expansion rights for adjacent capacity and the price of future power	AI clusters grow in large steps. Lock in the next step before someone else takes it.
☐	Term length and exit rights matched to the useful life of the hardware	GPU generations turn over every few years. A long term can outlast the hardware it was signed for.
☐	Water use and source for cooling, and any local limits or reporting	Water constraints can limit expansion or draw local attention.

Size power and cooling with our free rack planner, or get high-density colocation quotes.