



Colocation Lease Terms for GPU Racks: kW, Ramp, Cooling

Plan and build private AI compute with GPU Smith. We connect workload requirements with GPU hardware, networking, deployment and operating decisions so your infrastructure fits the work it must perform.

Published by GPU Smith · 2026-10-29 · colocation contract high density gpu clauses · committed kw vs metered power colocation · colo lease escalator power pass-through ai · data center colocation for liquid cooled gpu clusters · liquid cooled gpu colocation · managed ai cloud vs colocation gpu infrastructure - what are the cost differences?

Original article: <https://gpusmith.com/articles/en/colocation-lease-terms-for-gpu-racks>



In this article

1. What decides the bill
 2. Why standard colo paper breaks for GPU racks
 3. Committed vs metered power, and how pass-through works
 4. Density caps, circuits and the 80% rule
 5. Liquid cooling: who owns the loop
 6. Ramp schedules and take-or-pay exposure
 7. SLAs and service credits
 8. Escalators, term and renewal
 9. Cross-connects and remote hands
 10. Insurance, liability and exit
 11. Negotiation checklist
 12. FAQ
 13. Sources
-

What decides the bill

A colocation agreement for GPU racks is mostly a power contract; space is close to free next to the kilowatts. Most overruns and disputes come from five places: how committed power is billed, when billing starts, who owns the liquid-cooling loop, what the SLA leaves out, and what it costs to leave. This guide takes those clauses in order, with the arithmetic shown.

The clause patterns below are typical market practice, not norms, drawn from law-firm commentary, operator filings, two published provider contracts and trade guides. Several sources sell colocation; that bias is noted where it matters. Price ranges quoted from [datacentres.com](#) come from one undated trade guide and are indicative, not market data. This is not legal advice; have counsel review any agreement before you sign.

Why standard colo paper breaks for GPU racks

Retail colocation contracts were written for enterprise IT. One trade guide puts the "standard allocation" at 5-8 kW per cabinet ([datacentres.com](#)). Nautilus Data Technologies, itself a colocation operator, says the engineering assumptions behind standard rates are "typically 8 to 12 kW per cabinet", while "rack densities of 40, 60, even 100+ kW are now common in serious AI deployments" ([Nautilus, 2026-03-19](#)). Uptime Institute's 2026 survey reported a modal rack density of 11 kW, and an average of 7.8 kW when facilities designed above 30 kW are excluded ([Data Center Knowledge, 2026-07-31](#)).

A single NVIDIA DGX B300 draws up to 15 kW ("15 kW system max."; 14.5 kW power consumption) in 10U, per the [DGX B300 user guide](#) (updated 2026-10-21); the product page rounds this to "~14 kW". Four of them in one rack come to about 60 kW of IT load at system max (derived) before switches. Equinix's 10-K for 2025 warns that in older sites "the current demand for power may exceed the designed electrical capacity" and that "power, not space, is a limiting factor" ([Equinix 10-K, filed 2026-02-11](#)). For rack-level electrical sizing, see the [rack power requirements planning guide](#).

Know which kind of document you are signing. Retail colocation is usually a master services agreement with orders for cabinets, cross-connects, remote hands and installation placed under it ([Gowling WLG, 2025-04-09](#)). One provider's schedule states outright that an order does not "create any landlord and tenant relationship" ([Viatel colocation terms](#)). Wholesale deals are real leases, commonly triple-net or modified gross, "often with initial terms of 10 years or more" ([Core Scientific 10-K, FY2025](#)). Remedies and exit differ between the two.

Committed vs metered power, and how pass-through works

Committed kW is the number everything else hangs off. It "defines your maximum IT load and determines your monthly base charge" ([datacentres.com](#)). Electricity is then billed in one of two ways. Csquare's 2026 IPO filing describes both clearly:

- **All-in (bundled):** "customers pay a single recurring charge that includes power". Nearly all of Csquare's all-in contracts still contain "power indexation, utility rate pass-throughs, or extraordinary cost adjustment clauses".
- **Metered:** "a fixed facility and capacity fee plus electricity as a separate, metered charge", with "the customer bear[ing] all of the electricity price volatility".

As of 2026-03-31, all-in contracts made up about 70.6% of Csquare's recurring revenue and metered contracts about 29.4%. Both models are "typically based on a committed power capacity (kW)" ([Csquare S-1/A, filed 2026-07-06](#)). Other operators file similar language. Digital Realty negotiates "expense pass-through provisions in customer agreements for operating expenses, including power costs" ([Digital Realty 10-K, filed 2026-02-13](#)). Core Scientific's agreements "include power pass-through provisions", and wholesale deals are "frequently paired with long-term, take-or-pay contractual commitments" ([Core Scientific 10-K](#)).

"All-in" does not mean fixed. Viatel's schedule includes a cost of power, the "Facility kW Cost", in the kW rate. When the facility's cost per kWh rises more than 10% above that baseline, Viatel may bill an Energy Cost Surcharge: $ECS = D \times A \times (C/B - 1)$. In that formula, D is the prior month's average kW, A is the Facility kW Cost, B is the baseline cost per kWh and C is the current cost per kWh ([Viatel §11.5](#)).

Worked example: committed vs metered

All inputs are illustrative, not quotes. Adjust them to your bids.

- Contracted: 4 racks $\times$ 60 kW = **240 kW**
- Metered model: capacity fee \$140/kW/month, plus IT kWh $\times$ 1.30 contractual PUE factor $\times$ \$0.09/kWh
- All-in model: \$210/kW/month, of which \$70/kW is the power component (A in the ECS formula)
- 730 hours per month

Metered energy charge = 240 kW $\times$ load factor $\times$ 730 h $\times$ 1.30 $\times$ \$0.09 = \$20,498.40 $\times$ load factor.

Average draw (load factor)	Metered: capacity + energy	All-in
96 kW (40%)	\$33,600 + \$8,199 = \$41,799	\$50,400

Average draw (load factor)	Metered: capacity + energy	All-in
144 kW (60%)	\$33,600 + \$12,299 = \$45,899	\$50,400
192 kW (80%)	\$33,600 + \$16,399 = \$49,999	\$50,400
240 kW (100%)	\$33,600 + \$20,498 = \$54,098	\$50,400

Break-even load factor = $(\$50,400 - \$33,600) \div \$20,498.40 \approx \mathbf{82\%}$ (derived). Inference fleets far below their cap favour metered; training clusters near it favour all-in.

Now raise the utility rate from \$0.09 to \$0.12/kWh with the load factor at 60%. The metered bill rises to $\$33,600 + 105,120 \text{ kWh} \times 1.30 \times \$0.12 = \mathbf{\$49,999}$. On the all-in side, a Viatel-style surcharge applies: $\text{ECS} = 144 \times \$70 \times (0.12/0.09 - 1) = \mathbf{\$3,360}$, for a total of \$53,760. The bundled price moved as well, just behind a 10% deadband.

PUE multipliers and audit rights

If energy is billed as IT kWh times a PUE factor, the contract has to define that factor. Is it a fixed number or a measured one? Which measurement category applies, how often is it measured, and is there a cap? The reference definition is ISO/IEC 30134-2. Its 2016 edition has been withdrawn and replaced by the 2026 edition, which covers measurement categories, on-site generation and unmeasured energy ([ISO/IEC 30134-2:2026](#)). Ask for the "right to audit metered power consumption, PUE calculations, and SLA compliance data" ([datacentres.com](#)), and for access to the meter readings themselves.

Density caps, circuits and the 80% rule

Caps work at three levels: the circuit, the cabinet and the aggregate. Viatel's terms require that peak draw stay within each cabinet's "Individual Limit" and within the "Committed Aggregate kW". If draw exceeds either, Viatel may require corrective action, "limit the provision of power" and bill overage ([Viatel §4.1](#)). The same schedule limits each circuit to 80% of rated amperage. A redundant A/B pair should be balanced at about 40% per side, so that either side can carry the full load on failover (§11.6).

The derived arithmetic for one circuit: a 415 V three-phase 60 A circuit is $\sqrt{3} \times 415 \times 60 \approx 43.1 \text{ kVA}$ nameplate, or about 34.5 kVA at 80%. A 60 kW rack with A/B feeds therefore needs at least two such circuits per side. Because GPU training produces synchronized power swings, negotiate the overage trigger (instantaneous peak, 15-minute demand or monthly average) so a compliant average does not breach on peaks.

Ask for:

- Pooling of kW across the cage or suite, rather than per-cabinet caps.
- Stated maximum kW per rack and per row, tied to the cooling design.
- The right to raise density without a renegotiation. Morgan Lewis lists "any limitations on increasing density over time" as a cost driver ([Morgan Lewis, 2026-05-18](#)).
- Floor loading. "A fully loaded DGX rack and associated CDU can exceed 2,000 pounds", and standard leases often say nothing about it ([Nautilus](#)). Attach the slab or raised-floor rating and the delivery path to the order. The [floor loading and delivery worksheet](#) lists what to measure.

Liquid cooling: who owns the loop

Direct-to-chip cooling makes the contract boundary physical. Equinix, which sells this as a product, describes a primary loop from the facility to the CDU and a secondary loop from the CDU to the servers. It calls the quick-disconnect at the rack manifold "the logical place to mark the boundary". It also notes that the coolant "crosses the demarcation", so "neither party has full ownership over the coolant" ([Equinix blog, 2026-05-07](#)). Uptime Institute adds that the industry has not settled how to split ownership and maintenance of DLC components between facilities and IT teams ([Uptime Institute, 2026-04-15](#)). Vertiv lists several workable splits. In one, "Facilities manages up to the rack CDU or manifold; IT handles internal server plumbing" ([Vertiv, 2025](#)).

Write the split down in a cooling rider; Nautilus recommends one that "addresses who designs and pays for any new fluid infrastructure" ([Nautilus](#)). The rider should cover:

- **Ownership and capex.** Who buys, owns and maintains the CDU (in-rack or in-row), the manifolds and the secondary piping ([datacentres.com](#) raises the same question for CDU, piping and coolant).
- **Performance spec.** Supply temperature, flow, pressure and delta-T at the demarcation, per the OEM.
- **Fluid chemistry.** Who samples it, how often, and who may top up or change fluid.
- **Alarms and authority.** Who receives the leak or pressure alarm, and who may close which valve.
- **Leak liability.** "If a tenant deploys direct liquid cooling and a line fails, who owns the damage?" Settle that before deployment, "not during the insurance claim" ([Nautilus](#)).
- **Change control.** Nautilus suggests operator approval before installation and 60-90 days' notice of material workload changes.

Some operators advertise high liquid-cooled ceilings. Csquare, for example, cites air cooling up to about 50 kW per rack and liquid cooling up to about 250 kW ([Csquare S-1/A](#)). Tie any such ceiling to the specific hall in your order. Demarcation tests are in the [cold plate commissioning checklist](#).

Ramp schedules and take-or-pay exposure

Capacity charges usually start at service commencement, servers or not; a ramp schedule bills committed kW in steps. One guide suggests the "right to deploy below committed power without penalty during ramp-up (6-12 month grace period)". It also advises against "locking in more capacity than needed for 12-18 months of projected growth" ([datacentres.com](#)).

Illustrative ramp, using the 240 kW example: bill 60 kW in months 1-3, 120 kW in months 4-6 and 240 kW from month 7. Compared with billing 240 kW from day one, the tenant avoids $180 \times 3 + 120 \times 3 = 900$ kW-months of capacity charges. That is **\$126,000** at a \$140/kW capacity fee, or \$189,000 at a \$210 all-in rate (derived).

The ramp must also protect against provider delay: a calendar-dated ramp is worthless if ready-for-service (RFS) slips. Morgan Lewis points to "tiered credits and termination rights for RFS delays" as ways to align incentives ([Morgan Lewis](#)). Digital Realty notes that "lease commencement dates may significantly impact final delivery schedules" ([Digital Realty 10-K](#)). Tie each ramp step to provider readiness: power energized, cooling commissioned and acceptance passed. Add a right to delay a step if GPU delivery slips.

SLAs and service credits

Read past the headline uptime number. Verizon's published colocation SLA shows the usual limits ([Verizon, 2025](#)):

- On a dual feed, the commitment "covers power to at least one power feed at any one time". The standards are 100% for dual feed and 99.98% for single feed.
- The credit is 25% of the affected space's MRC, which is also the cap.
- The customer must file a written claim within 15 days of month-end.
- Credits are the "sole and exclusive remedy".
- Measurement is suspended during scheduled maintenance.
- No credit is due when customer usage "trips a breaker or otherwise exceeds any power consumption limits".

That last exclusion matters for dense GPU racks. A trade guide puts typical credits at 5-30% of MRC and argues they should be automatic, with the SLA covering "the full power and cooling chain, not just utility power" ([datacentres.com](#)). Digital Realty discloses that repeated failures can give customers "the right ... to terminate the agreement" ([Digital Realty 10-K](#)).

For liquid-cooled racks, ask for a cooling SLA measured at the demarcation (supply temperature and flow) and a chronic-failure termination right. A 25% credit on a cabinet's MRC is small next to one lost cluster-day.

Escalators, term and renewal

- **Term.** Equinix's IBX colocation contracts generally run one to five years and "thereafter automatically renews in one-year increments" ([Equinix 10-K](#)). Digital Realty's new >1 MW leases signed in 2025 had a weighted average term of 10.0 years, against 4.5 years for 0-1 MW ([Digital Realty 10-K](#)).
- **Escalators.** The trade guide calls 2-3% fixed "most common" and CPI-linked with a cap "generally most favourable for the tenant" ([datacentres.com](#)). Watch for the reverse structure. Viatel may raise charges annually "by the higher of (i) 3 (three) % or (ii) CPI ([Viatel §10.2](#)), which is a floor with no cap. Compounding (derived): over ten years, a 3% escalator yields 14.6% more total rent than a flat rate, and 2% yields 9.5% more.
- **Renewal.** Renewal pricing can reset to market. Digital Realty's >1 MW renewals in 2025 re-signed at cash rates 27.0% above expiring rates ([Digital Realty 10-K](#)). Negotiate a renewal cap or a fixed renewal formula, and watch the notice windows. Morgan Lewis flags "strict notice requirements" on auto-renewal ([Morgan Lewis](#)).
- **Unilateral changes.** Viatel's §10.1 lets charges change "in line with any change in cost incurred by Viatel"; look for such clauses.

Cross-connects and remote hands

GPU fabrics mostly stay inside the cage; cross-connects carry WAN, cloud on-ramps and any links between suites. One guide quotes \$200-500 per month each, with installation at \$500-2,000 ([datacentres.com](#)), and Morgan Lewis calls them "a meaningful expense". If the cluster could split across halls, pre-price bulk fiber

between them.

Remote hands is usually time-and-materials ([Csquare S-1/A](#)); none of the major operators we checked publish hourly rates, so none are quoted here. Viatel's operational-support scope covers reboots and visual checks as remote hands, plus hardware tasks such as hot-swap PSU replacement. Several of those tasks require a signed release of liability, and "other tasks" are performed at the "Customer's own risk" ([Viatel §9](#)). For GPU nodes, list the tasks you expect: reseating, transceiver swaps, cold-plate quick-disconnect handling. Ask for response times, the billing increment and liquid-cooling training.

Insurance, liability and exit

- **Liability caps.** Providers "typically limit their liability, often capping damages and excluding consequential losses" ([Morgan Lewis](#)).
- **Insurance.** The operator usually does not insure your IT equipment ([Gowling WLG](#)). Viatel requires €1,000,000 per occurrence public liability plus all-risk cover on the customer's equipment. A trade guide cites \$5-10M general liability as standard. Insure the GPUs at replacement value, and check subrogation waivers in both directions.
- **Liens.** Ask for lien waivers and clear bailment language, which Morgan Lewis lists as protection against creditor claims on your hardware.
- **Early termination.** It is rare. Where it exists, it "typically requires 12-18 months' remaining rent as penalty" ([datacentres.com](#)).
- **Exit mechanics.** Viatel requires removal within 10 business days of termination, with charges continuing while equipment stays. After that, Viatel may remove it with a 15% administration fee, and anything left after 90 days is removed (§11.10). Liquid-cooled racks add draining, fluid disposal and secondary-loop decommissioning; assign each in the rider and agree a holdover rate.

Negotiation checklist

Clause	Ask for	Red flag
Power billing	Metered or all-in chosen against your load factor	All-in with uncapped surcharge
PUE factor	Fixed number, ISO/IEC 30134-2 category, audit right	"Facility PUE as determined by provider"
Caps	Aggregate pooling, stated per-rack and per-row kW, defined averaging window	Instantaneous per-cabinet cap with suspension right
Ramp	Steps tied to RFS and acceptance, 6-12 months below commit	Billing from a calendar date
Liquid cooling	Rider: demarcation, spec, chemistry, alarms, leak liability	"Customer responsible for all liquid infrastructure"
SLA	Power and cooling chain, automatic credits, chronic-failure exit	Claim-only credits, breaker-trip exclusion
Escalator	CPI with cap	"Higher of 3% or CPI"
Renewal	Capped renewal rate, short notice window	Reset to market

Clause	Ask for	Red flag
Exit	Defined removal period, drain/decommission scope, holdover rate	Disposal after short deadline

FAQ

Is high-density colocation cheaper than renting GPUs from a cloud? Only above a utilization threshold, once hardware, staff and these contract terms are counted. The [own vs rent GPU TCO comparison](#) works through that break-even. Colocation adds take-or-pay exposure that rental does not have.

Should I pick committed or metered power? Model both against your measured average draw. In the example above, the crossover sits near an 82% load factor.

If you want a second engineering read on the power and cooling exhibits of a colocation bid, [GPU Smith](#) reviews them alongside rack and cluster design.

Sources

- [Data Centre Lease Negotiation: Terms, SLAs, and Deal Points](#) — datacentres.com, undated (© 2026), accessed 2026-10-28
- [What Happens to Your Colo Contracts When a Tenant Upgrades to High-Density GPU Workloads](#) — Nautilus Data Technologies (colocation vendor), 2026-03-19, accessed 2026-10-28
- [AI Drives Data Center Uncertainty in Uptime's 2026 Survey](#) — Data Center Knowledge, 2026-07-31, accessed 2026-10-28
- [NVIDIA DGX B300](#) — NVIDIA, accessed 2026-10-28
- [Equinix, Inc. Form 10-K for fiscal 2025](#) — SEC EDGAR, filed 2026-02-11, accessed 2026-10-28
- [Colocation data centre services: key points to consider in drafting colocation services agreements](#) — Gowling WLG, 2025-04-09, accessed 2026-10-28
- [Viatel Co-Location Specific Terms and Service Schedule \(Schedule F\)](#) — Viatel Ireland, accessed 2026-10-28
- [Core Scientific, Inc. Form 10-K for fiscal 2025](#) — SEC EDGAR, 2026, accessed 2026-10-28
- [Csquare, Inc. Form S-1/A](#) — SEC EDGAR, filed 2026-07-06, accessed 2026-10-28
- [Digital Realty Trust Form 10-K for fiscal 2025](#) — SEC EDGAR, filed 2026-02-13, accessed 2026-10-28
- [ISO/IEC 30134-2:2026 Power usage effectiveness \(PUE\)](#) — ISO, published 2026-01, accessed 2026-10-28
- [Activating the Colo Model: What Really Matters in Data Center Colocation Agreements](#) — Morgan Lewis, 2026-05-18, accessed 2026-10-28
- [The Anatomy of a Direct-to-Chip Liquid Cooling System](#) — Equinix (provider blog), 2026-05-07, accessed 2026-10-28
- [Liquid cooling will not outgrow its high-density niche](#) — Uptime Institute, 2026-04-15, accessed 2026-10-28
- [Liquid cooling services: Why the separation of IT and facilities is becoming increasingly fluid](#) — Vertiv (vendor), 2025, accessed 2026-10-28

- [Colocation Service Level Agreement — Verizon, 2025, accessed 2026-10-28](#)

Source links

Links cited in this article, in order of first appearance. Full addresses are provided for readers using extracted text.

1. <https://www.datacentres.com/guides/data-centre-lease-negotiation>
2. <https://nautilusdt.com/what-happens-to-your-colo-contracts-when-a-tenant-upgrades-to-high-density-gpu-workloads>
3. <https://www.datacenterknowledge.com/ai-data-centers/ai-drives-data-center-uncertainty-in-uptime-s-2026-survey>
4. <https://docs.nvidia.com/dgx/dgxb300-user-guide/introduction-to-dgxb300.html>
5. <https://www.sec.gov/Archives/edgar/data/1101239/000110123926000032/eqix-20251231.htm>
6. <https://gowlingwlg.com/en/insights-resources/articles/2025/colocation-data-centre-services-key-points-to-consider-in-drafting-colocation-services-agreements>
7. <https://www.viatel.com/wp-content/uploads/2026/08/Viatel-Service-Schedule-F-Colocation-Service-Terms.pdf>
8. <https://www.sec.gov/Archives/edgar/data/1839341/000162828026013305/core-20251231.htm>
9. https://www.sec.gov/Archives/edgar/data/2105398/000110465926080556/tm264837-13_s1a.htm
10. <https://www.sec.gov/Archives/edgar/data/1297996/000110465926015365/dlr-20251231x10k.htm>
11. <https://www.iso.org/standard/30134-2>
12. <https://www.morganlewis.com/blogs/datacenterbytes/2026/05/activating-the-colo-model-what-really-matters-in-data-center-colocation-agreements>
13. <https://blog.equinix.com/blog/2026/05/07/the-anatomy-of-a-direct-to-chip-liquid-cooling-system>
14. <https://journal.uptimeinstitute.com/liquid-cooling-will-not-outgrow-its-high-density-niche>
15. <https://www.vertiv.com/en-emea/insights/articles/educational-articles/liquid-cooling-services-why-the-separation-of-it-and-facilities-is-becoming-increasingly-fluid2>
16. https://www.verizon.com/business/service_guide/reg/colocation-service-SLA_2025MAY15_mk.pdf
17. <https://www.nvidia.com/en-us/data-center/dgx-b300>

THE PUBLISHER

About GPU Smith

GPU Smith provides independent engineering and research for private AI infrastructure. We help technical buyers and operators reason about GPU workload sizing, hardware selection, procurement, deployment and inference operations, from the compute system to the networking, power and cooling requirements around it.

Start with the workload

Useful infrastructure decisions begin with the models, data, throughput, latency and operating constraints a team actually has. GPU Smith helps connect those requirements with system design and deployment choices. The goal is a privately controlled AI environment whose capacity and operational demands are understood before a purchasing decision.

Hardware and supplier research

Our public reference library covers GPUs, complete systems and networking components. The server-vendor directory and manufacturing research help teams investigate suppliers, compare options and follow sources behind technical claims. We distinguish manufacturer specifications from measured benchmarks and advertised capabilities from tested configurations. Reference pages are research resources, not a live inventory listing or a binding equipment quote.

Deployment and operations

GPU Smith's engineering scope includes infrastructure integration, networking, power, cooling and the practical operation of inference workloads. Our published articles explain the tradeoffs behind deployment, procurement and ongoing operations so teams can ask better questions and document decisions.

Work with GPU Smith

Explore the [hardware library](#), [GPU references](#), [networking references](#) and [vendor research](#). [Contact GPU Smith](#) with your workloads and deployment constraints to discuss a project and confirm current scope, pricing and availability.

Inclusion of a manufacturer or product in our research does not imply a partnership, certification or endorsement.

Website: <https://gpusmith.com>

This article is educational. Verify consequential medical, legal, financial, regulatory and technical decisions with appropriate professionals and the cited primary sources. Company services are described separately from the article's research findings.