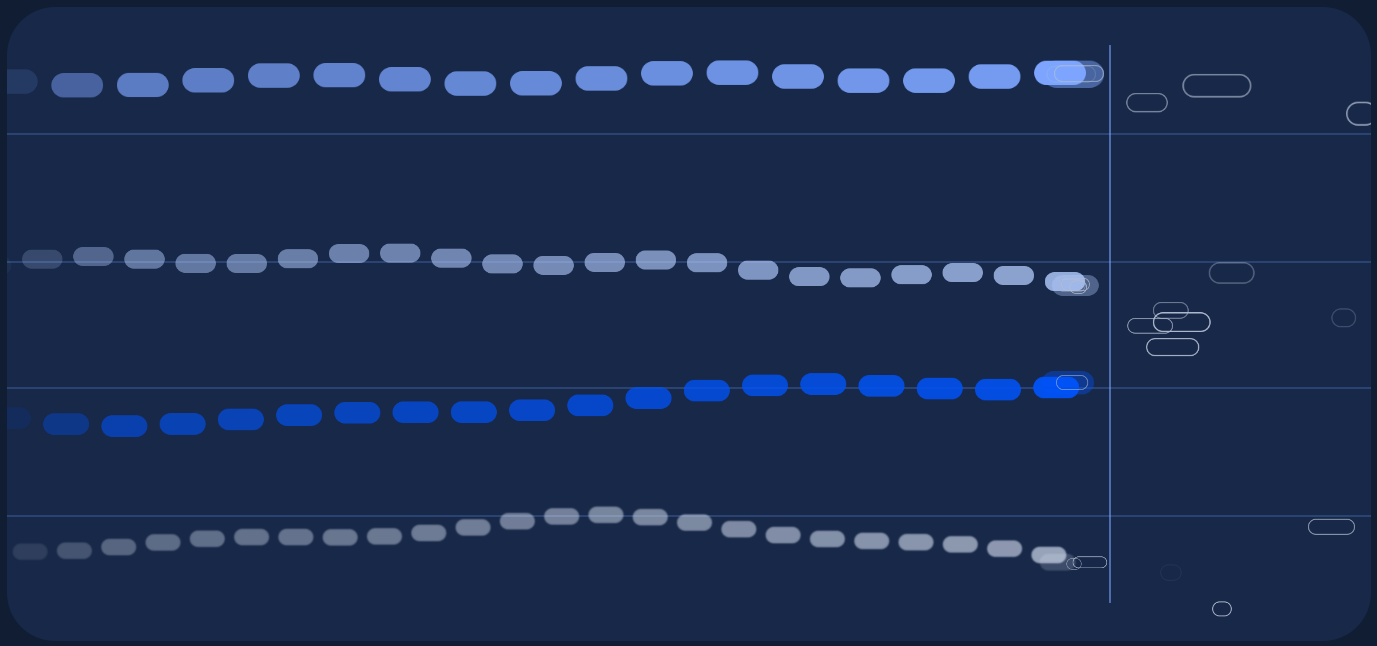
Rillor

METHODS NOTE

Rent-to-value: a simple way to read GPU payback

Annualized rent divided by hardware price gives a quick read on GPU payback, and this note works the ratio through public prices to show what moves it.



Rillor

Contents

01 Summary

02 Background

03 Method

04 Findings

05 Implications

06 Limits of this analysis

07 Sources

Summary

- Rent-to-value is a year of rent for one GPU divided by one GPU's share of the system price. Methodology v1.1 of the Rillor Compute Index defines its RCI Rent-to-Value series this way, by GPU class [1].
- At on-demand prices shown on 9 October 2026 and one public listing per class, the ratio was 1.55 for the B200 and 1.40 for the H200. At CoreWeave's spot prices it was 0.78 and 0.63 [2, 3, 4, 5, 6]. These figures illustrate the arithmetic. They are not RCI values.
- Posted ratios overstate what operators earn. Utilization, committed-use discounts of up to 60% [3] and contract pricing pull the realized ratio down. A five-year contract with disclosed totals works out to about 0.33 [15].
- Payback follows from the realized ratio, operating costs and the path of rents. With power as the only cost, a realized ratio of 0.5 pays back in about 2.2 years with flat rents and 2.8 years if rents fall 15% a year. At 0.33 the same cases take 3.5 and 5.9 years.
- Depreciation over five to six years [10, 11, 12], power and first-year interest at 6% to 9% on 90% of the price [16] set a hurdle of about 0.26 to 0.32 in realized ratio.
- The ratio is silent on residual value. Accounting lives for servers differ by up to a year across filings [10, 11, 13, 14], and rating analysts treat value after the first contract as a question of re-leasing demand [21].

Background

People who buy, rent or lend against GPUs ask the same first question: how long until the hardware pays for itself? Operators answer it in their own terms. CoreWeave estimated a cash payback of about 2.5 years, measured through adjusted EBITDA and including customer prepayments [10]. IREN puts recent three-year contracts at about two years, measured as GPU capital spending divided by contracted revenue less direct costs [16]. Nebius reports one year and ten months for deals signed in the second quarter of 2026, on a revenue-recognition basis that excludes prepayment [11].

Those figures use private inputs: contract prices, costs and capacity plans. Outside readers have public inputs: published rental rates and published hardware prices. Rent-to-value is built from those. It is a first read, comparable across GPU classes and over time, and it can be adjusted step by step toward a payback estimate as better inputs arrive.

Operators already annualize hourly prices the same way. IREN calculates annualized revenue as the GPU-hour price for commissioned GPUs under contract multiplied by 8,760 hours per year [16]. Operators also quote revenue per megawatt: CoreWeave reported short-dated contracts at about $40 million per megawatt of annualized revenue [22], and Nebius reported $20 million to $25 million per megawatt on its large second-quarter deals [11]. Converting a per-megawatt figure to a per-GPU ratio requires the operator's power per GPU and the hardware price per megawatt, which these disclosures do not give.

Method

Definitions

For one GPU class:

- **Rent-to-value (R).** R = (hourly rate per GPU × 8,760) ÷ (system price ÷ GPUs per system).
- **Realized rent-to-value (V).** V = R × utilization × (1 − discount), where utilization is the share of hours sold and discount is the reduction from the posted rate.
- **Operating cost ratio (c).** Annual operating cost per GPU divided by the price per GPU. In this note c covers power only.
- **Simple payback (flat rents).** Payback in years = 1 ÷ (V − c).
- **Payback with declining rents.** If rents fall at a constant annual rate d, cumulative net rent after T years is V × (1 − e^(−kT)) ÷ k − c × T, with k = −ln(1 − d). Payback is the T at which that sum reaches 1. Some combinations never reach 1.
- **Hurdle.** The realized ratio that covers straight-line depreciation, operating cost and first-year interest: 1 ÷ life + c + interest rate × share financed.

Payback here ignores tax, working capital, the timing of prepayments and any value left at the end. It is a ratio test, and it says so.

Inputs

Input	Value	Source and date
HGX B200 8-GPU system, listed price	$384,228, out of stock	Newegg listing, first listed 29 July 2025, viewed 9 October 2026 [5]
HGX H200 8-GPU system, listed price	$293,354, out of stock	Newegg listing, first listed 4 August 2025, viewed 9 October 2026 [6]
B200 on-demand	Lambda $6.69 per GPU (8-GPU instance), CoreWeave $68.80 per 8-GPU instance ($8.60 per GPU), Nebius $8.50 per GPU-hour	Pricing pages, viewed 9 October 2026 [2, 3, 4]
H200 on-demand	CoreWeave $50.44 per 8-GPU instance ($6.305 per GPU), Nebius $5.40 per GPU-hour	Pricing pages, viewed 9 October 2026 [3, 4]
Spot, 8-GPU instance, North America	CoreWeave B200 $34.11, H200 $20.93 per instance	CoreWeave pricing page [3]

Input	Value	Source and date
Committed-use discount	Up to 60% off on-demand	CoreWeave pricing page [3]
System power	About 14.3 kW maximum for 8 GPUs (DGX B200)	NVIDIA [7]
Power usage effectiveness	1.54, 2025 weighted average	Uptime Institute survey [9]
Electricity price	8.51 cents per kWh, U.S. industrial, 2025 (final)	EIA, released 23 October 2026 [8]
Useful life	5 to 6 years	Operator filings [10, 11, 12]
Financing cost	6.0% and 9.0%, with 90% of GPU spending financed	IREN [16]

The rent for each class is the median across providers of each provider's on-demand rate per GPU-hour, following the index's rule of a median per provider and then a median across providers [1]. With one rate per provider, the per-provider median is that rate. For the H200, two providers published a rate, so the median is their average.

The hardware prices are listings in the evidence classes the index uses: what a seller asked in public, at a point in time [1]. Both listings were out of stock. A quote or a reported sale for the same grade could differ.

Findings

1. Rent-to-value at published prices

GPU class	Provider	Rate per GPU-hour	Annual rent per GPU	Rent-to-value
B200	Lambda	$6.69	$58,604	1.22
B200	CoreWeave	$8.60	$75,336	1.57
B200	Nebius	$8.50	$74,460	1.55
B200	Median	$8.50	$74,460	1.55
H200	CoreWeave	$6.31	$55,232	1.51
H200	Nebius	$5.40	$47,304	1.29
H200	Median	$5.85	$51,268	1.40

Sources: [2, 3, 4, 5, 6]. Price per GPU: B200 $48,028.50, H200 $36,669.25. Annual rent uses the unrounded per-GPU rate (CoreWeave's H200 instance price divided by eight is $6.305).

On 9 October 2026, provider rates for the same class differed by up to 29% (B200 from $6.69 to $8.60), so the ratio depends on which providers are counted. The rule for combining them matters as much as the inputs.

Rates also changed shortly before the capture date. Nebius's price page listed two on-demand rates per class, one of them effective 1 October 2026 [4]. At the first set, the Nebius B200 ratio is 1.30 and the H200 ratio 1.08. At the rates effective 1 October they are 1.55 and 1.29. Rates have moved since then as well. Lambda's pricing page, viewed 15 November 2026, lists $7.69 per GPU-hour for its eight-GPU B200 instance from 16 November 2026, which gives a ratio of 1.40 at the same system price [23]. A ratio is a statement about a date.

2. From posted to realized

Basis	B200 rate per GPU-hour	Rent-to-value
On-demand, median of three providers	$8.50	1.55
CoreWeave on-demand	$8.60	1.57
CoreWeave spot	$4.26	0.78
CoreWeave on-demand less the maximum stated 60% committed-use discount	$3.44	0.63
On-demand median at 60% utilization	n/a	0.93
On-demand median at 60% utilization and a 60% discount	n/a	0.37

Sources: [2, 3, 4, 5]. Utilization and discount rows are arithmetic on the stated assumptions.

A contract with disclosed totals gives a check from the other direction. IREN's Microsoft contract, for NVIDIA GB300 systems, is worth about $9.7 billion over five years, including a 20% prepayment, or about $1.94 billion a year. IREN agreed to buy the GPUs and ancillary equipment from Dell for about $5.8 billion, a figure that includes deployment services, servers, InfiniBand, cabling, software and licensing [15]. Annual contract revenue divided by that cost is about 0.33. The denominator includes more than GPU servers, so the ratio is lower than a server-only calculation would give, and the GPU class differs from the B200 example.

The gap between 1.55 and 0.33 is the gap between a posted hourly rate for uncommitted capacity and the price of a five-year commitment for a large cluster. Both are real prices. They describe different products. SemiAnalysis reported in April 2026 that most of the rental market traded on contracts of six months or longer, and that large neoclouds wanted terms of at least one year and preferred two or three years, with five years the usual tenor for long-term offtake [20].

3. Operating cost

Power is the one operating cost that every GPU carries and that public data can estimate. NVIDIA lists the DGX B200 at about 14.3 kW maximum for eight GPUs [7], or about 1.79 kW per GPU. The Newegg system is a different build, so this is a proxy. Multiplying by the 2025 weighted average power usage effectiveness of 1.54 reported by Uptime Institute survey respondents [9] gives about 2.75 kW of facility power per GPU. Over 8,760 hours at 8.51 cents per kWh [8], that is about $2,052 a year per GPU at full draw, or 4.3% of the B200 price per GPU.

The figure is an upper bound at the stated electricity price, because it assumes maximum system draw for every hour of the year. Uptime also reports lower averages for newer and larger facilities: 1.48 for facilities commissioned since about early 2020 and 1.44 for sites of 20 MW and above [9]. Colocation, networking, staff, software, insurance and maintenance are excluded and vary widely by operator. Together they make c larger than 0.043.

4. Payback by realized ratio and rent path

Payback in years, power as the only operating cost (c = 0.043), rents falling at a constant annual rate:

Realized rent-to-value	Flat rents	Rents fall 15% a year	Rents fall 30% a year
0.25	4.8	Not reached	Not reached
0.33	3.5	5.9	Not reached
0.50	2.2	2.8	Not reached
0.75	1.4	1.6	2.0
1.00	1.0	1.2	1.3
1.55	0.7	0.7	0.8

"Not reached" means cumulative net rent never equals the price under that path.

Two features stand out. Below a realized ratio of about 0.5, the rent path dominates: a steady decline can stop the hardware from ever paying for itself. Above about 0.75, payback arrives before declines have much effect. That is the logic behind take-or-pay contracts with prepayment: they fix the rent path for the contract term and shorten the time at risk.

Rents have moved in both directions over the period covered by public data.

Date or period	Observation	Source
Aug to Dec 2023	H100 hyperscaler median $7.76 per GPU-hour	Silicon Data [18]
Jun to Dec 2024	H100 medians: hyperscaler $9.34, neocloud $2.99, marketplace $2.58	Silicon Data [18]

Date or period	Observation	Source
From 1 Jun 2025	AWS cut On-Demand prices for H100-based P5 instances by 44%	AWS [17]
Jul to Dec 2025	H100 medians: hyperscaler $6.26, neocloud $3.33, marketplace $1.95	Silicon Data [18]
Oct 2025 to Mar 2026	One-year H100 contracts rose from $1.70 to $2.35 per GPU-hour	SemiAnalysis [20]
1 Oct 2026	Nebius on-demand H100 from $3.85 to $4.50, B200 from $7.15 to $8.50	Nebius [4]
9 Oct 2026	Silicon Data H100 rental index, neocloud reading, $2.85 per GPU-hour	Silicon Data [19]
14 Nov 2026	Silicon Data H100 rental index, neocloud reading, $3.13 per GPU-hour	Silicon Data [24]
From 16 Nov 2026	Lambda on-demand B200, eight-GPU instance, $7.69 per GPU-hour, against $6.69 on 9 Oct 2026	Lambda [2, 23]

The table also shows why segments must be kept apart. In the second half of 2025, Silicon Data's hyperscaler median for the H100 was more than three times its marketplace median [18]. A ratio built from one segment's rents says little about another's.

5. The hurdle set by depreciation and financing

Useful life	Interest rate on 90% of price	Depreciation	Power	First-year interest	Hurdle
6 years	6.0%	0.167	0.043	0.054	0.264
6 years	9.0%	0.167	0.043	0.081	0.291
5 years	6.0%	0.200	0.043	0.054	0.297
5 years	9.0%	0.200	0.043	0.081	0.324

Sources for the inputs: useful lives [10, 11, 12], rates and share financed [16], power from Finding 3.

The lives come from filings: CoreWeave depreciates technology equipment over six years [10], Nebius uses five years for server and network equipment [11], and Nscale uses five to six years [12]. The rates are IREN's reported GPU financings: 6.0% with an investment-grade customer and 9.0% fixed for non-investment-grade customers, the latter funding 90% of the related GPU spending [16].

A realized ratio near 0.26 to 0.32 is needed to cover depreciation, power and first-year interest, before any other cost and before any return to equity. The posted on-demand ratio of 1.55 clears the upper end at any utilization above about 21%. The contract-level 0.33 clears it narrowly, on a cost base that includes more than servers. Operators in the filings reviewed also use prepayment and investment-grade debt [16], and aim to extend customer relationships beyond the first contract [12].

6. Residual value context

Rent-to-value describes the first years. Value after the first contract depends on what the hardware can earn then.

Accounting lives give one reference, and they differ. CoreWeave uses six years [10]. Nebius moved from four years to five for 2026 [11]. Amazon shortened the useful life of a subset of its servers and networking equipment from six years to five, effective January 2025, citing "an increased pace of technology development, particularly in the area of artificial intelligence and machine learning" [13]. Meta extended most of its servers and network assets to 5.5 years from the same date [14].

Market evidence gives another. SemiAnalysis reported in April 2026 that some H100 contracts were renewing at the rates signed two to three years earlier [20]. KBRA wrote in June 2026 that the central credit question for these financings goes beyond whether GPUs or TPUs remain physically functional after the original contract term, to whether customers would keep using the capacity at an attractive price [21]. The methodology pairs the ratio with payback and residual-value context [1]. For residual value, the relevant evidence is the observed price of the same hardware later in its life. Under the methodology, RCI Hardware records system prices by comparable grade [1].

7. How RCI Rent-to-Value is built

Methodology v1.1 of the Rillor Compute Index, published 7 November 2026, defines four series: RCI Rental, RCI Availability, RCI Hardware and RCI Rent-to-Value [1].

- **Definition.** RCI Rent-to-Value is "a year of rent for one GPU at the published rate, divided by one GPU's share of a comparable system's price", reported by GPU class as a ratio, "with payback and residual-value context" [1].
- **Rent input.** Under the methodology, RCI Rental uses published prices from providers' official pricing APIs and pricing pages, by GPU class and variant, region, term and pricing type (on-demand, spot or preemptible, and reserved with its term), with hyperscaler and neocloud segments shown separately [1].

- **Price input.** RCI Hardware uses GPU system prices by comparable grade from listings, quotes and reported sales. Each observation carries its evidence class, the classes stay separate in calculation, and air-cooled and liquid-cooled builds are separate grades [1].
- **Calculation.** Each series uses a median for each provider, then the median across providers [1].
- **Conflicts.** Rillor's own quotes and agreements are excluded from every published series [1].
- **Publication.** The methodology sets the coverage a series needs before a value is published, and records every change with its date and reason [1].

The worked example in this note follows the same definition with a single listing per class and three or fewer providers. It shows the arithmetic and is not an RCI value.

Implications

For buyers of compute. The posted ratio shows how much room a provider has between list price and the cost of the hardware. The examples here run from a posted on-demand ratio near 1.5 in October 2026 to a contract ratio near 0.33 on a five-year contract signed in November 2025. Buyers with long, committed demand negotiate within that range.

For builders. The payback table shows where risk concentrates. Below a realized ratio of about 0.5, the path of rents decides whether the hardware pays back at all. Contract terms that fix the rent for most of the payback period, and prepayments that shorten it, move a project out of that zone.

For lenders. The hurdle table links the ratio to debt service. Depreciation over five to six years, power and interest on 90% of the price take 0.26 to 0.32 of the price in the first year. Collateral analysis also needs the residual-value side: what the same grade sells or rents for later.

For researchers. State the inputs with their dates, sources and evidence classes. Keep hyperscaler, neocloud and marketplace rents apart, and keep listings apart from sales. Report the posted ratio and the realized assumptions separately, so a reader can redo the arithmetic.

Limits of this analysis

- The worked example uses one listing per GPU class, both out of stock, and the rates of three or fewer providers on one date. It illustrates the method and is not an RCI value or an estimate of any operator's economics.
- Listings are asking prices. Quotes, reported sales and verified transactions for the same grade may differ, and the index keeps those classes apart.
- Power is estimated from a DGX B200 maximum figure applied to a different system, an industry-average PUE and an average U.S. industrial electricity price. Other operating costs are excluded.
- The payback formulas ignore tax, working capital, prepayment timing and residual value. The rent paths are assumptions for illustration.
- The IREN contract ratio uses totals that include networking, services and software, and a contract value that includes a prepayment.
- Prices change often. Inputs here are dated 9 October 2026 or as stated in the table.
- This note is research on methods. It is not investment advice and recommends no purchase or sale of any security or asset.

Sources

1. Rillor Compute Index, Methodology v1.1. Rillor Corporation, 7 November 2026, viewed 15 November 2026. https://rillor.com/markets/compute-index

2. GPU cloud pricing. Lambda, viewed 9 October 2026. https://lambda.ai/pricing

3. Pricing. CoreWeave, viewed 9 October 2026. https://www.coreweave.com/pricing

4. AI Cloud pricing. Nebius, viewed 9 October 2026. https://nebius.com/prices

5. Supermicro Gold Series GPU Server SYS-A22GA-NBRT-G1 (NVIDIA HGX B200 8-GPU), product listing. Newegg, viewed 9 October 2026. https://www.newegg.com/supermicro-gold-series-sys-a22ga-nbrt-g1-rack/p/N82E16859152392

6. Supermicro 8U Gold Series GPU Server SYS-821GE-TNHR-G1 (NVIDIA HGX H200 8-GPU), product listing. Newegg, viewed 9 October 2026. https://www.newegg.com/p/N82E16859152398

7. NVIDIA DGX B200. NVIDIA, viewed 9 October 2026. https://www.nvidia.com/en-us/data-center/dgx-b200/

8. Electric Power Monthly, Table 5.3: Average Price of Electricity to Ultimate Customers. U.S. Energy Information Administration, 23 October 2026, viewed 15 November 2026. https://www.eia.gov/electricity/monthly/epm_table_grapher.php?t=epmt_5_3

9. Uptime Institute Global Data Center Survey 2025. Uptime Institute, July 2025, viewed 15 November 2026. https://datacenter.uptimeinstitute.com/rs/711-RIA-145/images/2025.Annual.Survey.Report.pdf

10. CoreWeave, Inc. Registration Statement on Form S-1. U.S. Securities and Exchange Commission, filed 3 March 2025, viewed 15 November 2026. https://www.sec.gov/Archives/edgar/data/1769628/000119312525044231/d899798ds1.htm

11. Nebius Group Letter to Shareholders, Q2 2026. Nebius Group, 12 August 2026, viewed 15 November 2026. https://assets.nebius.com/assets/4462517b-ce83-41f2-96ed-f2ac1bc06a05/SHLQ226%20(1).pdf-0

12. Nscale Limited Registration Statement on Form S-1. U.S. Securities and Exchange Commission, filed 18 September 2026, viewed 15 November 2026. https://www.sec.gov/Archives/edgar/data/2110365/000119312526395475/ck0002110365-20260918.htm

13. Amazon.com, Inc. Annual Report on Form 10-K for the year ended 31 December 2024. U.S. Securities and Exchange Commission, filed 7 February 2025, viewed 15 November 2026. https://www.sec.gov/Archives/edgar/data/1018724/000101872425000004/amzn-20241231.htm

14. Meta Platforms, Inc. Annual Report on Form 10-K for the year ended 31 December 2025. U.S. Securities and Exchange Commission, filed 29 January 2026, viewed 15 November 2026. https://www.sec.gov/Archives/edgar/data/1326801/000162828026003942/meta-20251231.htm

15. IREN Secures $9.7bn AI Cloud Contract with Microsoft (Exhibit 99.1 to Form 8-K). IREN, 3 November 2025, viewed 15 November 2026. https://www.sec.gov/Archives/edgar/data/1878848/000114036125040072/ef20058139_ex99-1.htm

16. IREN Reports FY26 Results (Form 8-K). IREN, 27 August 2026, viewed 15 November 2026. https://www.sec.gov/Archives/edgar/data/0001878848/000187884826000051/irenreportsfy26results.htm

17. Announcing up to 45% price reduction for Amazon EC2 NVIDIA GPU-accelerated instances. AWS News Blog, 5 June 2025, viewed 15 November 2026. https://aws.amazon.com/blogs/aws/announcing-up-to-45-price-reduction-for-amazon-ec2-nvidia-gpu-accelerated-instances/

18. H100 Rental Price Over Time (2023 to 2025): A Complete Market Analysis. Silicon Data, 21 December 2025, viewed 15 November 2026. https://www.silicondata.com/blog/h100-rental-price-over-time

19. H100 Rental Price Index. Silicon Data, viewed 9 October 2026. https://www.silicondata.com/products/silicon-index/h100

20. The Great GPU Shortage: Rental Capacity. SemiAnalysis, 2 April 2026, viewed 15 November 2026. https://newsletter.semianalysis.com/p/the-great-gpu-shortage-rental-capacity

21. KBRA Releases Research: AI Compute Financings, Evaluating Re-Leasing Risk for GPUs and TPUs. KBRA, 22 June 2026, viewed 15 November 2026. https://www.kbra.com/publications/KNzgvHxx/kbra-releases-research-ai-compute-financings-evaluating-re-leasing-risk-for-gpus-and-tpus

22. CoreWeave Continues to Contract New Compute Capacity at Higher Prices. CoreWeave, 17 September 2026, viewed 15 November 2026. https://investors.coreweave.com/news/news-details/2026/CoreWeave-Continues-to-Contract-New-Compute-Capacity-at-Higher-Prices/default.aspx

23. GPU cloud pricing. Lambda, viewed 15 November 2026. https://lambda.ai/pricing

24. H100 Rental Price Index. Silicon Data, reading as of 14 November 2026, viewed 15 November 2026. https://www.silicondata.com/products/silicon-index/h100

About this report

Published by Rillor on 16 November 2026. Superintelligence (SI) tools assist research and drafting, and every figure cites its source. Rillor builds and secures SI systems, supplies the data they learn from and delivers the infrastructure they run on. We apply SI to financial and compute markets.

Cite as: Rillor. Rent-to-value: a simple way to read GPU payback. 16 November 2026. https://rillor.com/insights/rent-to-value-gpu-payback

Notices

Rillor is not a registered investment adviser or commodity trading advisor and does not provide investment or trading advice. Research, forecasts, data and software described on this site are for research and engineering use. Nothing here is an offer or recommendation to buy or sell any security, commodity interest or digital asset. Past or simulated results do not indicate future results.

Rillor Compute Index series are for research and planning use. They are not offered for settlement of any contract.

NVIDIA, AMD, Ceph, Kubernetes and related product names are trademarks of their owners and are used only to identify products. No affiliation or endorsement is implied.

Work with Rillor

Send the series, GPU classes and history you need, and how you plan to use them. Contact us at rillor.com/contact?type=index or contact@rillor.com. More on the Rillor Compute Index: rillor.com/markets/compute-index