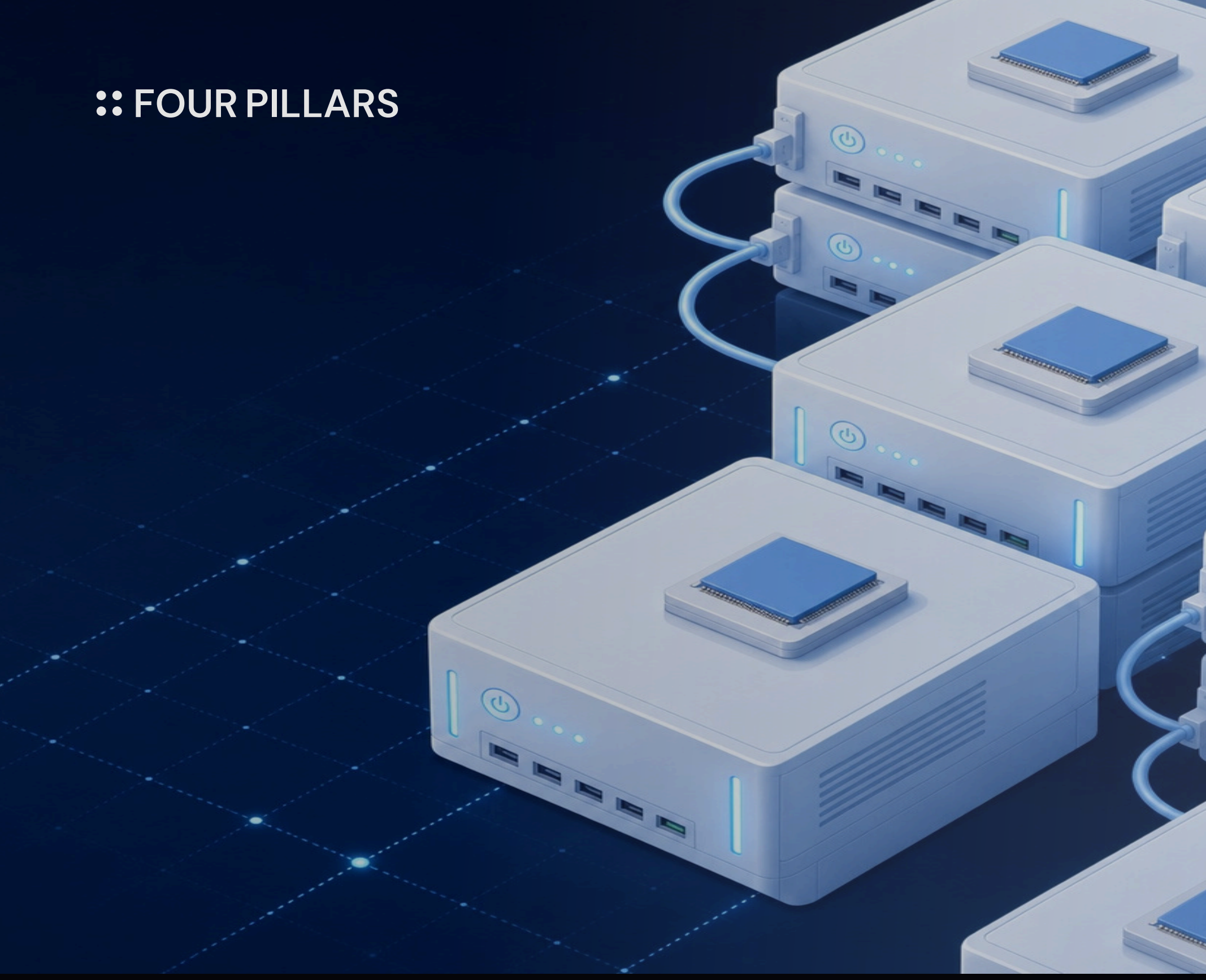
:: FOUR PILLARS



The Making of a Commodity Market for Compute

Research by

Sungjun Eun — Sphere Labs | Head of Research, Korea
moyed@spherepay.co

Reviewed by

Jinsol Bok — Four Pillars | Head of Research
100y@4pillars.io

sphere Ornn

© 2026. Sphere Labs & Four Pillars. All Rights Reserved.

[Disclaimer]

The author of this report may have personal holdings or financial interests in assets or tokens discussed herein. However, the author affirms that no transactions have conducted using material non-public information obtained in the course of research or drafting. This report is intended solely for general information purposes and does not constitute legal, business, investment, or tax advice. It should not be used as a basis for making any investment decisions or as guidance for accounting, legal, or tax matters. Any references to specific assets or securities are made for informational purposes only and should not be construed as an offer, solicitation, or recommendation to invest. The opinions expressed herein are those of the author and may not reflect the views of any affiliated institutions, organizations, or individuals. The opinions and analyses expressed herein are subject to change without prior notice. In addition, beyond the individual disclosures included in each report, Four Pillars, may hold existing or prospective investments in some of the assets or protocols discussed herein. Furthermore, FP Validated, a division of Four Pillars, may already be operating as a node in certain networks or protocols discussed herein or may do so in the future. Please see below links in the footer for FP Validated's participating network disclosures and for broader disclosure details.

© 2026. Sphere Labs & Four Pillars. All Rights Reserved.

Contents

00 Executive summary 03

01 Introduction 04

02 Capital is already moving 05

03 The object of trade 11

04 How compute is priced 15

05 Where price risk can be transferred 18

06 What hedging leaves behind 23

07 Conclusion 25

References 28

© 2026. Sphere Labs & Four Pillars. All Rights Reserved.

Executive summary

Executive summary
1. Introduction
2. Capital is already moving
3. The object of trade
4. How compute is priced
5. Where price risk can be transferred
6. What hedging leaves behind
7. Conclusion
References

This report examines whether compute can trade as a commodity and what would have to exist first. Commodity markets such as crude oil were built in sequence. A standard for what is traded came first, and price discovery, hedging and clearing followed. In compute the sequence is reversed. Capital at the scale of an energy industry and futures have arrived first, while the definition of what is traded and the clearing arrangements are not yet in place. The reversal matters to lenders and investors because the terms of that financing could change if compute acquired a price that lenders could underwrite and hedge.

Capital reaches compute on the strength of counterparties' credit rather than the price of compute.

Compute financing takes several forms, but in each one lenders look to the credit of parties whose ability to pay can be checked, such as customers buying GPU capacity on long-term contracts, data center tenants and chip vendors that provide guarantees. Compute price risk enters only through renewal assumptions after contracts expire and a small segment of loans secured on GPUs. None of the facilities reviewed references a published compute price. Part of the reason lenders rely on credit is that no market price yet exists for them to underwrite against.

The instruments for trading compute price risk are being built before the foundations of that price.

Commodity futures define the unit and grade of what is delivered and then price it. Compute is quoted in GPU-hours, yet the same chip delivers different amounts of work depending on where it is installed and what it runs. No unit or grade that would reconcile those differences has been adopted. Meanwhile four published benchmarks track H100-class capacity, each measuring a different slice of the market, and NYMEX and ICE have specified futures on two of them. A US prediction market and several offshore venues already trade contracts on compute prices, but no futures contract is yet cleared with margin by a US-regulated clearing house. At the request of the Commodity Futures Trading Commission (CFTC), NYMEX agreed on 4 November 2026 to extend the review of its contracts to 8 January 2027, and responses to the CFTC's request for comment centered on whether any index is yet fit to settle them.

Hedging with these contracts converts price risk into basis risk, the risk that the hedge and the hedged exposure move apart. No existing or proposed mechanism would bound the basis that arises between chip generations.

Differences of location, specification and contract terms can be bounded as volume accumulates or as grading and physical delivery take hold. The drift between chip generations has no such path, because no process converts an H100 into a B200 and no settlement rule fixes the relationship between their prices. On Ornn's indices, weekly changes in H100 and B200 rental prices correlated at only 0.25 over the year to August 2026. Each new generation moves a generation-specific contract further from the revenue it is meant to hedge, yet no dedicated contract on the spread between generations has been filed.

© 2026. Sphere Labs & Four Pillars. All Rights Reserved.

1. Introduction

Executive summary
1. Introduction
2. Capital is already moving
3. The object of trade
4. How compute is priced
5. Where price risk can be transferred
6. What hedging leaves behind
7. Conclusion
References

Computing capacity for AI has become one of the largest categories of corporate investment, and a market structure for trading it is being assembled around that investment. The report asks what must exist before compute trades as a commodity.

Compute here means the processing capacity of GPU servers used to train and run AI models, mostly rented by the GPU-hour. Sellers are hyperscalers such as Amazon Web Services, Microsoft Azure and Google Cloud, along with neoclouds that specialize in renting GPUs. Buyers are AI laboratories and companies that build AI features into their services. Contracts range from spot capacity that can be reclaimed at short notice to offtake agreements that fix price and volume for years. The larger the contract, the less likely its price is disclosed. The chips themselves, such as NVIDIA's A100, H100 and B200, turn over quickly from one generation to the next.

Commodity markets are built in layers. A standard defines what is traded, spot transactions form a price, futures transfer price risk, and a clearing house stands between the two sides of each trade. In crude oil these layers accumulated over more than a century, while in compute the capital has arrived before most of them are in place. Section 2 therefore begins with how that capital is financed and what lenders underwrite, because their terms show which layers are still missing. Sections 3 to 5 take standards, price formation, and venues and clearing in turn. Section 6 asks what a hedge leaves behind, and Section 7 sets out the questions the evidence has not yet answered.

© 2026. Sphere Labs & Four Pillars. All Rights Reserved.

2. Capital is already moving

Executive summary
1. Introduction
2. Capital is already moving
3. The object of trade
4. How compute is priced
5. Where price risk can be transferred
6. What hedging leaves behind
7. Conclusion
References

The clearest view of what the compute market lacks comes from how its build-out is financed. **Lenders are financing AI compute through four channels, and in each they rely on a creditworthy party to repay them rather than on the market value of the compute being produced.**

2.1 A capital program on the scale of an energy industry

The natural comparison for the scale of AI investment is the oil and gas industry. Estimates compiled by Ribbit Capital in July 2026 from company data, Morgan Stanley Research and BMI put the combined capital expenditure of Amazon, Microsoft, Alphabet, Meta and Oracle at $449 billion in 2025 and about $805 billion in 2026. Global oil and gas capital expenditure runs at about $640 billion a year over the same period, so on these estimates the five companies spent about 30% less than the industry in 2025 and would exceed it in 2026. Federico Bandi and Yinan Su of Johns Hopkins estimate that the compute installed by the end of 2025 could deliver services worth $430 billion to $1.3 trillion a year, or 1.4% to 4.0% of US GDP, if it were fully used and priced at posted rental rates.

The output of this investment is already priced like a commodity. SemiAnalysis reports that the one-year rental price for H100 capacity rose from $1.70 to $2.35 per GPU-hour between October 2025 and March 2026, about 38% in five months. Oil, gas and power producers face similar capital needs and price swings, and they hedge their output in futures markets. Compute providers hedge their interest rates, and in some facilities their power costs, because their lenders require it. Yet the price of the capacity they sell is largely unhedged, because no listed contract yet exists for it.

2.2 What each financing channel underwrites

The four financing channels differ in who repays the lender (Exhibit 1). Moving down the exhibit, each channel relies less on a contracted customer, and in the last one repayment can come from the producer's output itself.

Financing channels and what each underwrites, November 2026

Channel	Who repays	Representative structure	What the lender underwrites
Asset finance	The offtake customer, through contract payments	CoreWeave's delayed draw term loans	The customer's credit for the life of the contract
Developer finance	The tenant, through rent	Meta and Blue Owl's Hyperion joint venture	The tenant's credit, plus a guarantee on the building's value
Demand finance	A guarantor, if the tenant fails	NVIDIA's residual value guarantees and capacity commitments	The tenant's credit, backstopped by the guarantor on building, power and transmission value
Producer finance	The producer, in cash or in output	OpenAI's loan to Cerebras	The producer's delivery and credit, with in-kind repayment valued at contract prices

Source: CoreWeave, Meta, NVIDIA and Cerebras filings; Moody's; Fitch

© 2026. Sphere Labs & Four Pillars. All Rights Reserved.

Asset finance

In asset finance, a lender funds the GPUs themselves and is repaid from a customer's contract for the capacity they produce. CoreWeave, which owns GPUs and rents out their capacity, has set the template. Since July 2023 it has raised seven Delayed Draw Term Loans (DDTLs), each of which lends against the cash flows of a customer contract. DDTL 4.0, closed in March 2026, is the cleanest example. It provides up to $8.5 billion, and its floating-rate half is priced at the Secured Overnight Financing Rate (SOFR) plus 225 basis points (bp). The loan is repaid in full within a six-year contract with a single customer, whom Moody's identifies as Meta, and Moody's rates it A3. The lenders are in effect lending against Meta's payments, and their repayment does not depend on what the servers would fetch on resale.

DDTL 5.5 shows what happens when the loan outlasts the contracts. The $2.6 billion facility, signed in August 2026, lends for about five years against customer contracts that average about three. Its price was SOFR plus 550bp, and CoreWeave itself had to guarantee it in full. Its lenders therefore rely partly on renewals, which Section 2.3 examines.

Developer finance

In developer finance, a lender funds the building and power for a data center and is repaid from the rent a tenant pays for it. Hyperion, the Louisiana campus that Meta developed with Blue Owl, shows how a tenant's credit can finance a project kept off the tenant's balance sheet. The campus sits in a joint venture owned 80% by Blue Owl and 20% by Meta, funded with $27.3 billion of bonds rated one notch below Meta. Meta's initial lease runs only four years from 2029, so Meta also guarantees a minimum value for the campus if it does not renew. Bondholders therefore rely on Meta twice, as tenant and as guarantor.

In June 2026 a subsidiary of the data center developer Cipher Digital sold $810 million of 6% notes due 2031 to build a Texas facility leased to AWS for fifteen years, with Amazon.com guaranteeing the rent. The lease runs about a decade beyond the notes, and the financing carries no guarantee on the building's value.

Demand finance

In demand finance, a party with a stake in the build-out backs the demand side, by guaranteeing a facility's value if its tenant defaults or by committing to buy capacity that would otherwise go unsold.

- NVIDIA has begun to support the credit of the facilities that buy its chips. In August 2026 it guaranteed a minimum value, capped at $105 billion, on data center leases between the developer SB Energy and an OpenAI affiliate. The guarantee covers buildings, power and transmission, and excludes the chips. NVIDIA pays only if the tenant defaults, and OpenAI must then repay NVIDIA. NVIDIA has also committed $36 billion to buy cloud capacity that its partners fail to sell to others, which gives those partners a buyer of last resort.
- Meta took a similar role for the GPU cloud provider Nebius in March 2026. Alongside $12 billion of dedicated capacity, it agreed to buy up to $15 billion of capacity on certain Nebius clusters that Nebius does not sell to other customers.

Executive summary
1. Introduction
2. Capital is already moving
3. The object of trade
4. How compute is priced
5. Where price risk can be transferred
6. What hedging leaves behind
7. Conclusion
References

© 2026. Sphere Labs & Four Pillars. All Rights Reserved.

Producer finance

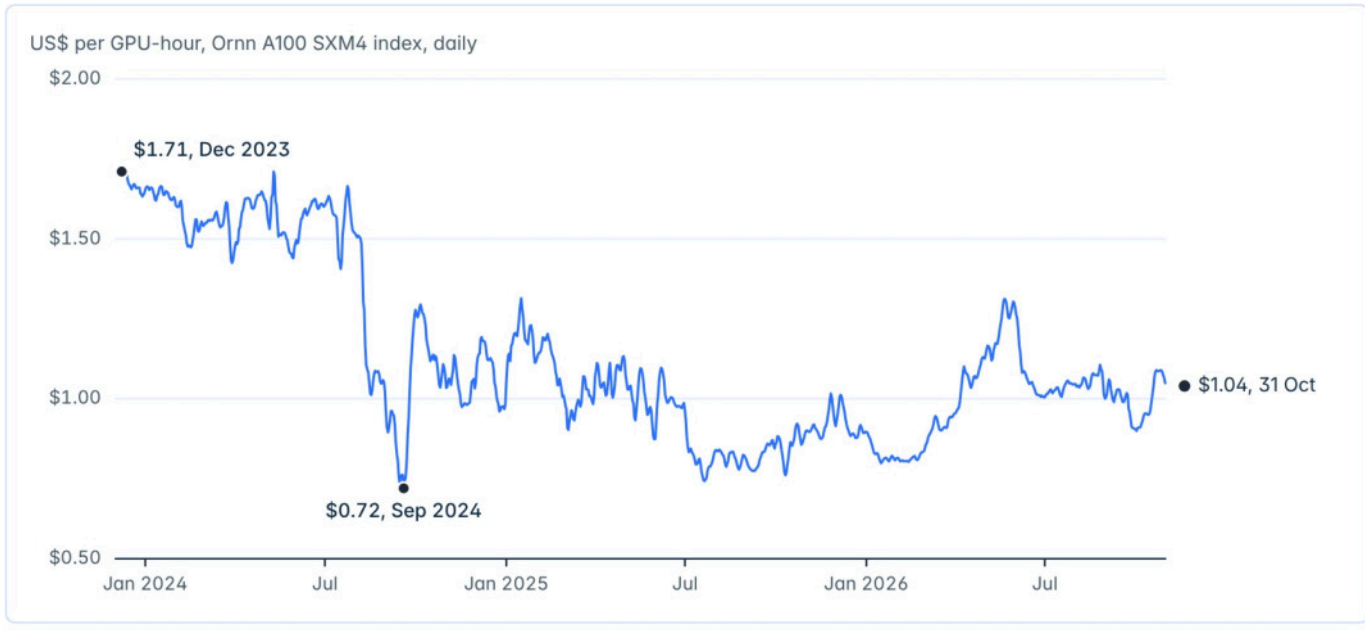
In producer finance, a buyer of compute lends to the company that produces it, and the loan can be repaid in capacity rather than cash. Cerebras, a chipmaker that also sells computing capacity, disclosed such a loan in its May 2026 prospectus. Oil and gas producers have long used the same structure. OpenAI lent Cerebras about $1.0 billion in January 2026, repayable in cash or in compute capacity, with interest waived on any part repaid in capacity. Repayment runs in installments over three years once the first 250 megawatts of capacity are delivered.

2.3 Where the price of compute sits in these structures

Each channel has a creditworthy party behind the cash flows, so the price of compute is a secondary concern for lenders. It matters most in DDTL 5.5. That loan runs beyond its current customer contracts, so repayment in the later years depends on CoreWeave signing new contracts at rents high enough to service the debt. Its credit agreement shows how the lenders handle those years. It lends only against GPUs that are already under contract, up to 70% of their cost, and never values the chips themselves. When a contract runs out, CoreWeave must show that expected income still covers its debt payments by at least 1.35 times, or repay part of the loan. For months with no customer yet, that expected income comes from CoreWeave's own forecast, which the rating agency Fitch also accepted. This is where the price of compute enters the loan, and the lenders accept the borrower's estimate of it because no market price exists to check it against.

The longest record of what a chip earns after its contracts end belongs to the A100. Ornn's index for the A100, a chip released in 2020, fell from $1.71 per GPU-hour in December 2023 to $0.72 in September 2024 as H100 supply grew, and has since moved between about $0.73 and $1.36 (Figure 1). Over the same period the share of A100s on rent rose from 61% in 2024 to about 86% in October 2026. Ornn, which runs a marketplace for GPU capacity, attributes the end of the decline to scarce power and demand for open-weight models. CoreWeave has also signed a new A100 contract running into 2029. Older chips stay on rent, but their rents can fall below half their early level, so for lenders the rent a renewal brings matters more than whether it happens. This is still only one generation's record, and it does not show what H100s and B200s will fetch when their contracts end.

A100 rental price, December 2023 to October 2026



Source: Ornn OCPI daily index, archived 26 August 2026; Ornn public index data to 31 October 2026

Beyond DDTL 5.5 the exposure is narrower still. The guarantees from Meta and NVIDIA cover buildings and power rather than GPUs, and lending directly against GPUs is confined to a small segment of the

Executive summary
1. Introduction
2. Capital is already moving
3. The object of trade
4. How compute is priced
5. Where price risk can be transferred
6. What hedging leaves behind
7. Conclusion
References

© 2026. Sphere Labs & Four Pillars. All Rights Reserved.

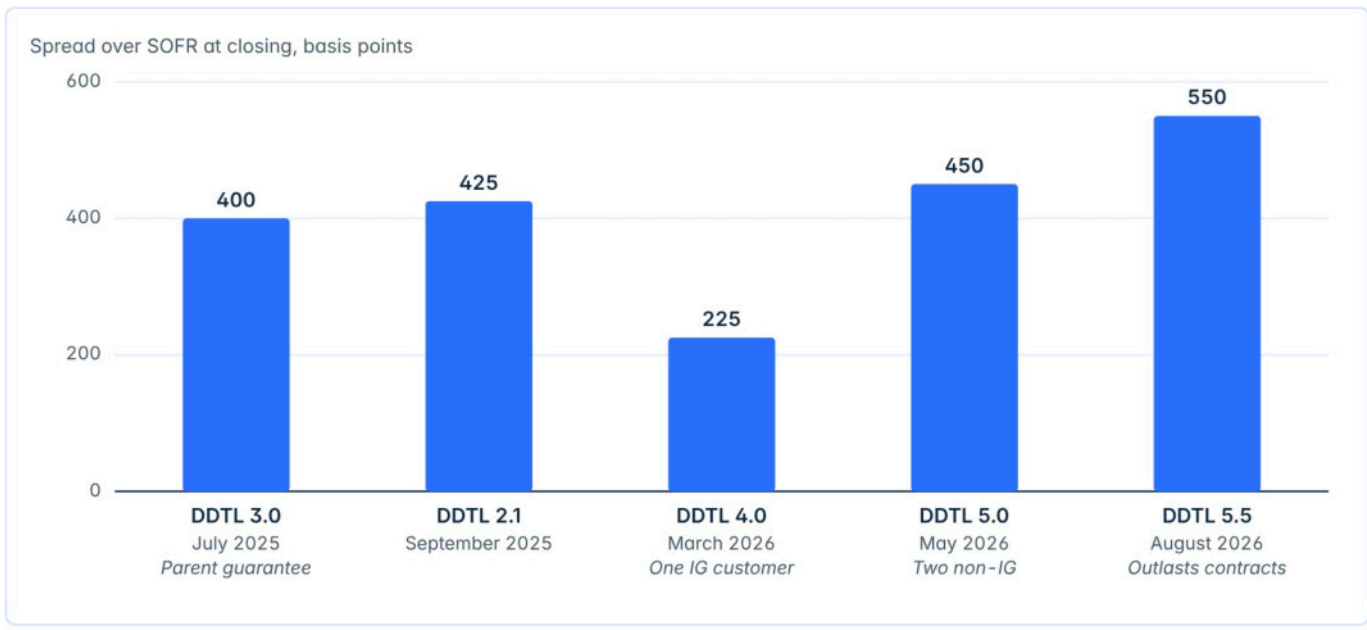
Executive summary
1. Introduction
2. Capital is already moving
3. The object of trade
4. How compute is priced
5. Where price risk can be transferred
6. What hedging leaves behind
7. Conclusion
References

market led by lenders such as USD.AI. No facility reviewed refers to a published compute price or forward curve. A lender asking where it is exposed to the price of compute itself would need such a curve, and the curves available in October 2026 are vendor estimates, prediction-market prices and thinly traded offshore contracts, none of them liquid.

The reliance on creditworthy parties shapes who can build. Providers that can sign long contracts with strong customers borrow most cheaply. Lambda's $926 million loan of August 2026, which is repaid in full within the contract of a single investment-grade customer, was priced at SOFR plus 300bp and rated Baa2 by Moody's, according to Lambda.

Others borrow on a third party's credit, as the GPU cloud provider Fluidstack did with a Google guarantee on its leases. CoreWeave's own two loans show how pricing changes with the contract support behind a loan (Figure 2, Exhibit 2). The two loans also differ in guarantees, covenants and the credit market at the time they priced, so the gap cannot be traced to one cause.

CoreWeave delayed draw term loan spreads, 2025 to 2026



Source: CoreWeave 10-K, 10-Q and 8-K filings; Moody's; Fitch

Two CoreWeave facilities compared, 2026

	DDTL 4.0	DDTL 5.5
Size and closing	Up to $8.5 billion, March 2026	$2.6 billion, August 2026
Spread over SOFR	225bp on the floating half	550bp
Customers	One, Meta according to Moody's	Several, reported by Bloomberg as including Anthropic, Jane Street and Hudson River Trading
Guarantee from CoreWeave	Only against misconduct	Full guarantee
Loan against contract	Repaid within a six-year contract	About five years against contracts averaging about three
Rating	A3 Moody's	Ba2 Moody's, BB+ Fitch

Source: CoreWeave filings; Moody's; Fitch; Bloomberg

© 2026. Sphere Labs & Four Pillars. All Rights Reserved.

2.4 What credit markets are signaling

Credit markets began to price the risk of AI infrastructure debt from late 2025, and the debt of CoreWeave, the largest compute-specific borrower, was repriced sharply. The five largest US data center builders have added about $350 billion of debt over five years, according to Bloomberg data reported in July 2026, and nearly 80% of data center bonds issued since early 2025 were trading at wider spreads than when they were issued.

CoreWeave's own debt shows the move (Figure 3). DDTL 5.5 was offered in mid-July at SOFR plus 425bp to 450bp and closed at 550bp two weeks later, as investor appetite for AI-related debt weakened. In FINRA's trade reports, the yield on CoreWeave's 9.25% notes due 2030 rose from 8.1% in early October 2025 to 12.1% in mid-December. It eased through the spring and then climbed to 14.1% on 14 October 2026, the day Bloomberg reported that CoreWeave's credit default swaps, the cost of insuring its debt against default, had returned to a record high. The yield peaked at 15.1% on 23 October and stood at 14.3% on 30 October, when the company's newer notes due 2032 yielded 14.5%.

Yields on CoreWeave senior notes, May 2025 to October 2026



Source: FINRA TRACE trade reports, daily average of trades of $1 million or more; report calculation of yield to maturity

Higher borrowing costs have now spread even to loans that the customer's payments would repay in full before the contract ends. In October the AI infrastructure developer Volta marketed a $5 billion loan to finance GPUs for a six-year compute contract with Anthropic, PitchBook and Bloomberg reported. The loan is repaid within the contract's term but carries no rating and no parent guarantee. Price talk of SOFR plus 625bp to 650bp at 97 to 98 cents on the dollar did not draw enough demand, and the arranging banks sold it at 650bp and 95 cents.

The comparison with CoreWeave's DDTL 5.5 shows what lenders charged for. On contract terms DDTL 5.5 is the riskier loan, because it runs about two years beyond its customer contracts, yet it priced lower, at 550bp in August. Its customers reportedly also include Anthropic, so the difference does not lie in the customer. It lies in the support behind the borrower. DDTL 5.5 is rated, and CoreWeave, a listed company with several years of operating history, guaranteed it in full. Volta's loan has no rating, and the company behind it, founded in January 2026, gave no guarantee. Volta's wider pricing is consistent with lenders charging for that missing support, even though its customer contract covered the whole loan.

Taken together, these cases suggest that the cost of a loan turns on who stands behind repayment.

Executive summary
1. Introduction
2. Capital is already moving
3. The object of trade
4. How compute is priced
5. Where price risk can be transferred
6. What hedging leaves behind
7. Conclusion
References

© 2026. Sphere Labs & Four Pillars. All Rights Reserved.

Rates are low where a strong customer's contract repays the loan and high where repayment depends on what compute will fetch later. Recently, rates have also risen further the less proven the borrower is.

Executive summary
1. Introduction
2. Capital is already moving
3. The object of trade
4. How compute is priced
5. Where price risk can be transferred
6. What hedging leaves behind
7. Conclusion
References

© 2026. Sphere Labs & Four Pillars. All Rights Reserved.

3. The object of trade

Executive summary
1. Introduction
2. Capital is already moving
3. The object of trade
4. How compute is priced
5. Where price risk can be transferred
6. What hedging leaves behind
7. Conclusion
References

As Section 2.3 showed, lenders accept the borrower's estimate because no market price exists to check it against. Such a price requires agreement on what is being priced, which is standardization, the first layer of a commodity market. A standardized contract needs something a buyer will accept without inspecting it. That means settling what is delivered, how it is counted and how it is graded. **For compute, none of the three has yet been agreed.**

3.1 The delivery layer

Compute is delivered in stacked layers. At the bottom sits the bare-metal server, a whole physical machine rented without virtualization. It comes with the network fabric, power and cooling, and with the control software that provisions and monitors it remotely. Virtual machines, containers and aggregators that resell many suppliers' capacity all sit above it, and none can deliver more than the hardware beneath. The case for placing the standard at the bare-metal server rests on this, since every layer above inherits its properties.

Ariel Deschapell, a co-founder of Hydra Host, argues that this layer plays the role crude plays in oil. In his account, cloud providers compete on orchestration, the software that schedules work across servers, much as refiners compete on products. They resist standardizing orchestration, which leaves the bare-metal layer beneath it as the place where a standard can take hold. NVIDIA's DGX Cloud Lepton, a software layer meant to give developers one interface across GPU clouds, met partner resistance.

Supply at the bare-metal layer is fragmented, which is a precondition for price discovery and also the reason delivery needs a standard. The suppliers are mostly neoclouds operating outside the three hyperscalers. Operators such as CoreWeave and Lambda own GPUs and rent them out, while integrated operators such as Crusoe, IREN and Nebius also control land, buildings or power. Although NVIDIA certifies cloud partners and invests in many of them, supply remains spread across dozens of firms. A buyer that spreads work across several of these providers integrates with each one separately, because no common delivery standard exists.

Capacity at this layer is sold on four contract forms, which differ in commitment and in the supplier's right to reclaim it (Exhibit 3).

Contract forms for GPU capacity, November 2026

Form	Commitment	Can the supplier reclaim it?	Price	Typical use
Spot	None	Yes, on short notice, two minutes on Amazon Web Services (AWS)	Lowest, at discounts of up to about 90%	Interruptible training and batch jobs
On-demand	None	No, once provisioned	Posted rate	Experiments and short bursts
Reserved	One month to several years	No	Discount to on-demand	Steady inference workloads
Multi-year offtake	Several years, fixed price and volume	No	Negotiated, largest contracts by value	Hyperscalers and frontier AI laboratories

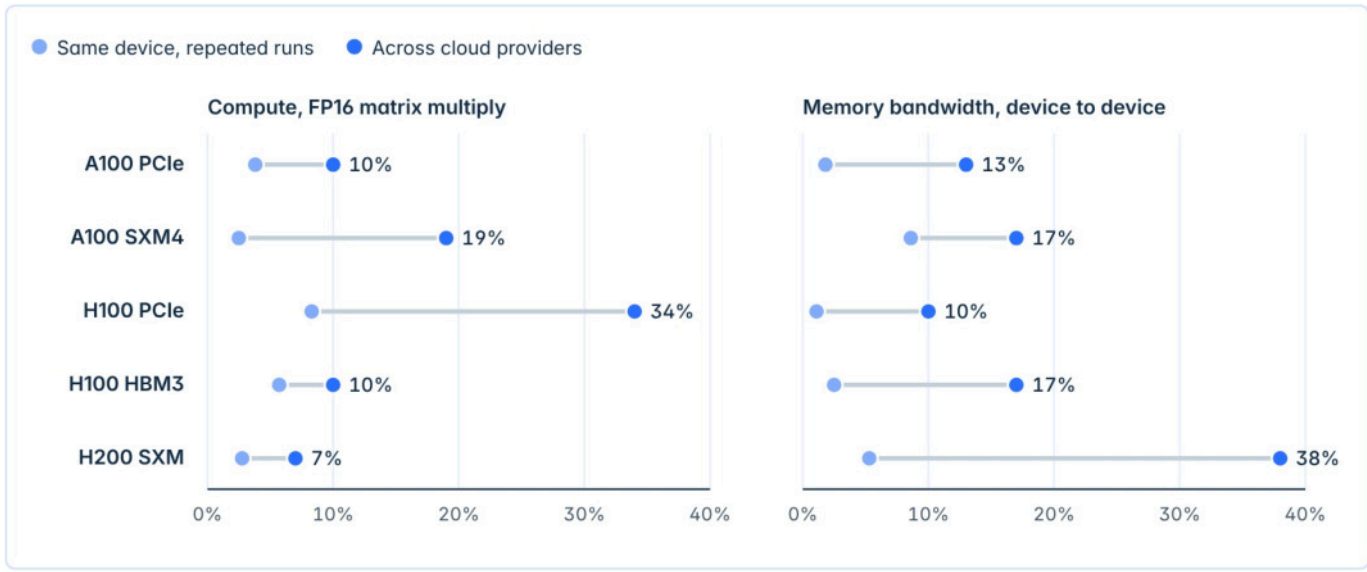
Source: AWS; provider documentation; Moonrunner, Moonletter

© 2026. Sphere Labs & Four Pillars. All Rights Reserved.

3.2 Why the GPU-hour and the token fail as units

The market quotes compute in dollars per GPU-hour, but the same nominal GPU-hour delivers materially different amounts of work. A study by Silicon Data and the College of William & Mary, presented at GPGPU 2026, measured memory bandwidth varying by up to 38% across H200 units and peak 16-bit compute by up to 34.5% across H100 units. Repeated runs on the same device varied by less than 9%, so the spread comes from the hardware and its host rather than from measurement noise (Figure 4). Differences between individual chips are the largest single source, accounting for 32% to 73% of the variance depending on the GPU model.

Performance range within one GPU model, repeated runs compared with cloud providers



Source: Silicon Data and College of William & Mary, Did You Win the GPU Cloud Lottery?, GPGPU 2026, Table 3

The underlying reason the GPU-hour falls short as a unit is that many variables decide the quality of delivered compute. A crude oil futures contract specifies a gravity band of 37 to 42 degrees API and sulfur of 0.42% or less for delivery at Cushing, and a power contract specifies a quantity at a node and an hour. Compute, by contrast, depends on chip generation, memory, interconnect, network, location, reliability, tenor and cooling, and several of these vary continuously. Like power, compute cannot be stored. Unlike either oil or power, the hardware that produces it also loses value to each new generation whether or not it wears out.

The token, widely used to bill AI usage, does not resolve the problem either. Many GPU workloads from image generation to rendering do not use tokens. Where tokens are used, an input token and an output token are different work for the GPU. Input tokens are the prompt the model reads, processed in parallel and limited by arithmetic throughput. Output tokens are the reply the model writes one at a time, limited by memory bandwidth. Model providers price the two separately. Anthropic's published API prices, for example, set output tokens at five times the input rate. A price per token therefore depends on how much of a workload is reading and how much is writing. Even one model has no single token price.

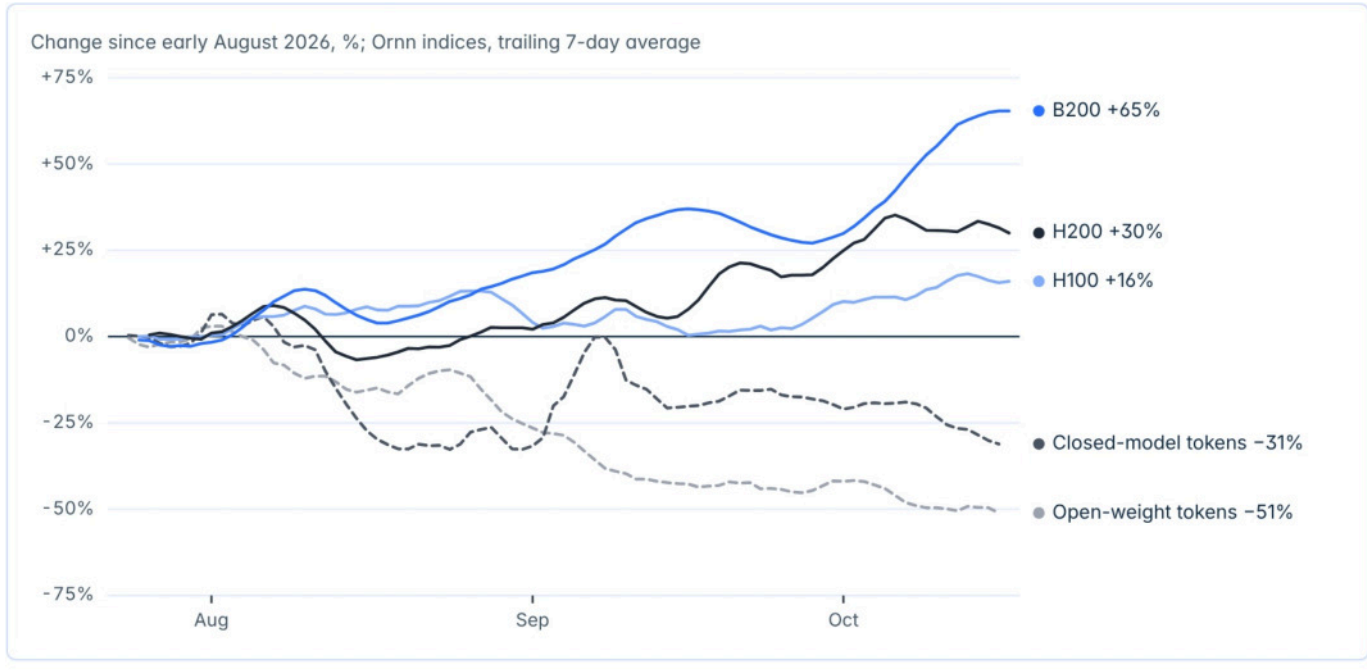
Neither the GPU-hour nor the token works as a unit on its own, and the two prices do not move together either. From early August to the end of October 2026, Ornn's rental indices for H100, H200 and B200 rose by about 16% to 65%, while its token price indices fell by about 31% for closed models and 51% for open-weight models (Figure 5). A provider that buys GPU-hours and sells tokens would therefore have found that a GPU index did not track its margin. The laboratories whose prices make up these token indices limit that exposure by buying most of their capacity on long-term contracts. Some market participants now chart the gap as a spread, borrowing from the crack spread between crude oil and its refined products. The comparison is loose, however. A barrel of crude yields a roughly fixed

Executive summary
1. Introduction
2. Capital is already moving
3. The object of trade
4. How compute is priced
5. Where price risk can be transferred
6. What hedging leaves behind
7. Conclusion
References

© 2026. Sphere Labs & Four Pillars. All Rights Reserved.

volume of products, but no fixed ratio converts a GPU-hour into tokens, so spending on tokens can rise while token prices fall.

GPU rental prices and token prices, August to October 2026



Source: Ornn GPU rental indices and token price indices by model class, public data read 1 November 2026; trailing 7-day averages; report calculation

3.3 Two routes to a standard

There are two ways to make capacity from different suppliers comparable. One quotes capacity in a composite unit, a single number blending a server's compute, memory, network and storage. The other sorts the differences into grades, as crude oil does.

The most developed attempt at a composite unit was 6fusion's Workload Allocation Cube, a patented metric for compute, network and storage consumption.

In 2014, 6fusion and CME Group signed a definitive agreement to build a spot exchange for physically delivered contracts in that unit, running on trading technology licensed from CME. The exchange, the Universal Compute Xchange, opened a beta in November 2014 and planned a full launch for the first half of 2015, but no record shows it moving beyond the beta. A composite unit struggles for three reasons:

1. Chip generations are different products with no universal conversion ratio. Bandi and Su convert rental prices into units of H100 computing power, measured by peak 8-bit throughput. Measured that way, a unit of computing power cost about $4.54 an hour on an A100, against $2.51 on an H100 and $2.02 on a B200, so the market does not price the generations by computing power alone.
2. A single number hides the network and facility attributes that decide what a buyer can run. Two clusters can receive the same value even if only one has the high-speed links between GPUs that training a large model requires.
3. The weights would need an industry agreement that different workloads make impossible, because each is limited by a different resource. Large-scale training leans on interconnect while generating replies leans on memory bandwidth, so a weighting that suits one prices the other wrongly.

Executive summary
1. Introduction
2. Capital is already moving
3. The object of trade
4. How compute is priced
5. Where price risk can be transferred
6. What hedging leaves behind
7. Conclusion
References

© 2026. Sphere Labs & Four Pillars. All Rights Reserved.

Today a composite unit exists only inside individual benchmarks, not as a standard across the market. Compute Desk's H100 index converts each H100 variant to the performance of one reference model, the SXM5 with 80GB of memory, before averaging prices. The weights are set by the index provider rather than agreed across the industry, and they apply only within one chip family.

Grading takes the opposite route and keeps the differences visible. Crude oil standardized a unit, a few quality measures and a set of delivery points, and many grades trade at differentials to a few reference crudes. The grade defines what a contract refers to, and the assay measures where a given cargo sits within it.

This route was proposed by Ariel Deschapell, whose proposal maps compute onto oil's three axes, with components in place of gravity, network in place of sulfur and data center in place of delivery point (Exhibit 4). Each grade would describe a specific configuration that any counterparty could test.

Crude oil grading set against compute, November 2026

Dimension	Crude oil	Compute proposal	Status in compute
Unit	Barrel of 42 gallons	GPU-hour	Quoting convention; work delivered per hour varies by host and workload
Inherent quality	API gravity	Components: GPU model, memory, thermal behavior, chassis	Proposed; no agreed thresholds
Capability boundary	Sulfur content	Network: switched fabric, bridged pairs or single GPU	Proposed; used informally in pricing
Location and reliability	Delivery point	Data center: power redundancy, cooling, security, uptime	Proposed; no agreed thresholds
Measurement standard	ASTM test methods named in the contract	Programmatic benchmarks under the inspector's control	Offered by market participants; no neutral standard
Reference contract	NYMEX light sweet crude	NYMEX H100 and B200 rental futures	Under CFTC review

Source: NYMEX Rule 200101; Deschapell, Compute Doesn't Need a Unit; CFTC

Grading would still face two problems that oil does not:

1. The reference grade itself becomes obsolete. NVIDIA releases a new generation roughly every twelve to eighteen months, while the core gravity and sulfur band of the West Texas Intermediate (WTI) contract has stayed in place.
2. Some differences are performance boundaries that price cannot bridge. A buyer that needs a network linking many GPUs for distributed training cannot use a single-GPU configuration at any discount. Such differences cannot be adjusted through basis, the price differential between grades, so in effect they should be treated as separate products.

Executive summary
1. Introduction
2. Capital is already moving
3. The object of trade
4. How compute is priced
5. Where price risk can be transferred
6. What hedging leaves behind
7. Conclusion
References

© 2026. Sphere Labs & Four Pillars. All Rights Reserved.

4. How compute is priced

Executive summary
1. Introduction
2. Capital is already moving
3. The object of trade
4. How compute is priced
5. Where price risk can be transferred
6. What hedging leaves behind
7. Conclusion
References

As Section 3 showed, compute still has no agreed unit or grade. Price discovery, the second layer of a commodity market, rests on that agreement. Compute therefore has yet to form a single price the whole market can refer to, although several benchmarks have already emerged that seek to provide one. **Compute prices are set separately across several sales channels, however, and the larger the channel, the less it discloses. Each benchmark therefore draws on a different part of the market.**

4.1 Six pricing channels from posted to private

A buyer looking for the price of an hour on an H100 finds several, depending on where it looks and how long it wants the capacity. Moonrunner describes six overlapping channels through which GPU capacity is priced (Exhibit 5). They range from hourly rates that anyone can read to multi-billion-dollar contracts whose terms are disclosed only in headlines. The largest volume trades in reserved and committed capacity, whose prices are negotiated privately. The channels that are easiest to observe therefore account for the least of what buyers spend. Multi-year offtakes, the largest contracts of all, bundle price with credit terms, delivery terms and extension or termination options.

Six channels through which GPU capacity is priced, November 2026

Channel	How the price is set	What is visible
Hyperscaler list prices	AWS, Google Cloud and Microsoft Azure post hourly rates set by internal methods	Public rates, updated irregularly and varying by region and commitment
Spot capacity	AWS and Google run internal spot markets and manage their prices	The cleanest public time series
Marketplaces	Aggregators such as Vast.ai, RunPod and Prime Intellect pool many operators' capacity	Prices close to an auction, for buyers who do not need guaranteed uptime
Neocloud prices	Specialized providers negotiate enterprise rates through sales teams	Some list prices; negotiated rates are private
Reserved and committed capacity	One-year and three-year commitments at a discount	Private; the largest volume
Bilateral long-term contracts	Multi-year agreements for frontier laboratories and national clusters	Headlines only

Source: Moonrunner, Inside the Fragmented Market for AI Compute; provider documentation

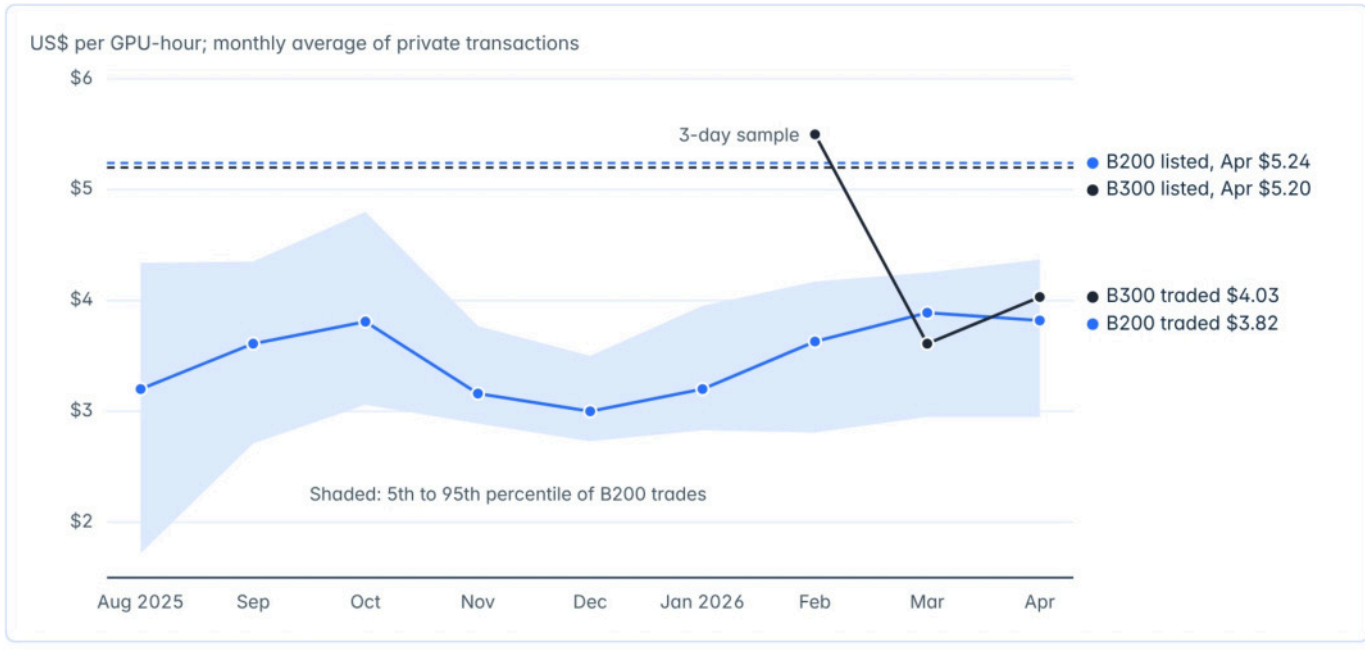
4.2 What buyers pay against posted prices

Even where prices are posted, they are a poor guide to what buyers pay. When David Lopez Mateos of Compute Desk compared private transaction data with posted rates for NVIDIA's B200 and the newer, higher-memory B300 in April 2026, both traded below their posted rates. The B200 transacted at $3.82 per GPU-hour, 27% below its posted $5.24, and the B300 at $4.03, 22% below its posted $5.20. On posted rates the B300 was the cheaper chip, but in transactions it was the more expensive one, and the ranking of the two reversed four times in two months. Individual trade prices were also widely scattered, with April's B200 trades running from about $2.95 to $4.37 between the 5th and 95th percentiles (Figure 6).

© 2026. Sphere Labs & Four Pillars. All Rights Reserved.

Executive summary
1. Introduction
2. Capital is already moving
3. The object of trade
4. How compute is priced
5. Where price risk can be transferred
6. What hedging leaves behind
7. Conclusion
References

B200 and B300 transacted prices against listed rates, August 2025 to April 2026



Source: Compute Desk, What does a B300 cost?, 12 May 2026; private transaction data with undisclosed sample sizes; values read from its chart and approximate

4.3 Benchmarks built from different data

With prices this scattered, benchmarks have emerged to condense them into reference numbers. **But the four that track H100-class capacity are built from different data, and their readings differ.**

Ornn and Silicon Data publish daily indices of H100 rental prices, Compute Desk aggregates the Hopper family including the H200, and SemiAnalysis surveys one-year contract prices each month (Exhibit 6). When Compute Desk compared the four from January 2025 to April 2026, they crossed several times. The gap between the highest and lowest ran from about $0.40 to $0.90 per GPU-hour, roughly 20% to 40% of the average level (Figure 7).

H100 and Hopper price benchmarks, January 2025 to April 2026



Source: Compute Desk, There is no GPU price, Figure 2, values digitized and approximate; Bloomberg; SemiAnalysis

What matters to a hedger is whether an index tracks the exposure it has. A provider renting out H100s by the hour and a laboratory buying a year of reserved capacity face different prices, and each would need a contract settled on its own product. Which providers a benchmark draws on also matters.

© 2026. Sphere Labs & Four Pillars. All Rights Reserved.

On 18 April 2026 Silicon Data's own hyperscaler H100 series stood at $7.43 against $2.50 for its neocloud series, about three times apart. Crude oil also trades on several benchmarks, but each is tied to a defined grade and delivery location. Compute has yet to define the products its benchmarks stand for, which is the gap in standards described in Section 3.

H100 benchmark methodologies compared, November 2026

	Silicon Data	Ornn	Compute Desk	SemiAnalysis
Inputs	Over 10,000 daily rental records from 40+ providers, standardized by a weekly-recalibrated model	Realized on-demand rentals only	Completed transactions blended with neocloud list prices	Monthly survey of 100+ participants, checked against deals
Contract types	On-demand, reserved and spot, adjusted to a benchmark contract	On-demand	One-year reserved; Hopper US series compared	One-year contracts
Segmentation	Per chip; neocloud and hyperscaler readings for H100 and A100	Per GPU type	Per chip and chip family; US only	H100 public; other chips and tenors paid
Distribution	Bloomberg	Bloomberg	Bloomberg, Refinitiv and General Index	Own dashboard, free
Contracts referencing it	NYMEX futures under CFTC review to 8 Jan 2027; offshore perpetuals on Bitget, unlicensed according to Silicon Data	ICE futures specified; live on Kalshi; perpetuals on Lighter	Futures live on AX, offshore; Nodal futures announced; American Innovation Exchange futures under review	None

Source: index methodologies; CFTC filings; exchange notices; Compute Desk; General Index; Silicon Data comment letter; venue documentation

Executive summary
1. Introduction
2. Capital is already moving
3. The object of trade
4. How compute is priced
5. Where price risk can be transferred
6. What hedging leaves behind
7. Conclusion
References

© 2026. Sphere Labs & Four Pillars. All Rights Reserved.

5. Where price risk can be transferred

Executive summary
1. Introduction
2. Capital is already moving
3. The object of trade
4. How compute is priced
5. Where price risk can be transferred
6. What hedging leaves behind
7. Conclusion
References

The lack of a single price has not stopped venues from specifying, filing and in some cases trading contracts on the benchmarks described in Section 4. **Venues for compute risk have multiplied in 2026, and the record of earlier attempts shows which conditions they still have to meet.**

5.1 Four failure modes from six decades of attempts

Attempts to trade compute are older than cloud computing. The first dates to 1968, when Ivan Sutherland described an auction for allocating time on Harvard's PDP-1. Ariel Deschapell's review of attempts to commoditize compute centers on four later cases:

- **Memory chip futures:** At least six attempts to list futures on dynamic random-access memory (DRAM) between 1989 and 2003, from the Pacific Stock Exchange to the Singapore Exchange, produced no trades.
- **Enron Broadband:** Enron standardized bandwidth units and built a trading desk for them, and the business collapsed with the company in 2001.
- **6fusion:** The composite-unit exchange described in Section 3.3, developed with CME Group from 2014, left no record of trading after its beta.
- **AWS Spot:** Amazon priced spare capacity through bidding from 2009 and replaced bidding with managed pricing in 2017, leaving a single seller to set the price.

Deschapell derives four failure modes from these cases, namely fungibility, delivery infrastructure, supply concentration and sequencing. Delivery infrastructure has since changed and supply concentration has changed in part, while fungibility and sequencing still remain unresolved (Exhibit 7).

Failure modes of earlier attempts to commoditize compute and their current status, November 2026

Failure mode	Definition	Status	Evidence
Fungibility	How far different deliveries can be treated as the same product	Unresolved	A grading proposal exists and has not been adopted
Delivery infrastructure	Systems that provision, meter and deliver traded capacity	Changed	Bare-metal control planes and on-demand rental operate commercially
Supply concentration	How far supply is concentrated enough for a few sellers to set the price	Changed in part	Supply is spread across dozens of providers, and hyperscalers have disclosed no participation as traders or data contributors, although some indices read their posted rates
Sequencing	The order in which delivery standards, indices and derivatives are put in place	Unresolved	Indices and futures are ahead of delivery standards, and ComputeConnect and Liquid Compute are attempting both together

Source: Deschapell, Every Attempt to Commoditize Compute Has Failed; CFTC filings; Architect and Liquid Compute announcements

© 2026. Sphere Labs & Four Pillars. All Rights Reserved.

5.2 Listed and pending contracts as of November 2026

Four kinds of venue now offer or plan contracts on compute prices, though the US-regulated futures that would anchor those prices have not begun trading (Exhibit 8). The four are established futures exchanges, newer Designated Contract Markets (DCMs), prediction markets and offshore venues that trade dated and perpetual futures.

NYMEX filed submission 26-370 on 11 August 2026 for two cash-settled contracts of 730 GPU-hours each. GPU1 references Silicon Data's H100 index and GPU2 its B200 index. Both settle on the monthly average of daily values, with 36 consecutive months listed. CME had targeted trading from 5 October. But on 21 September the CFTC's Division of Market Oversight extended its review to the end of 9 November 2026, and on 4 November NYMEX consented at the Commission's request to a further 60-day extension to 8 January 2027.

Other exchanges, existing and prospective, are preparing rival contracts.

- **ICE Futures U.S.:** Published specifications on 29 September 2026 for futures on Ornn's H100 and B200 indices. ICE sizes each contract to the hours in its calendar month and lists up to 24 months. It has not yet set a listing date, and it has published Ornn reference data on its platform since 5 October.
- **Nodal Exchange:** Announced futures on Compute Desk's indices on 3 September, to be cleared by Nodal Clear.
- **American Innovation Exchange:** Architect acquired IMX Health, an existing DCM, in May 2026 and is relaunching it as the American Innovation Exchange for compute futures, which remain under regulatory review.
- **Liquid Compute:** Has applied to become both a contract market and a clearing organization in February 2026, and now publishes its own price indices.

Prediction markets and offshore venues have started trading.

- **Kalshi:** Has offered contracts on Ornn's GPU price indices since 14 July 2026 and publishes market-implied forward curves from them. Monthly contracts are listed to August 2027, and the weekly series has not been relisted since 9 October.
- **Polymarket US:** Self-certified Ornn-index contracts in August but has not listed them in the US pending regulatory clarity, according to its comment letter.
- **AX:** Architect's exchange in Bermuda lists quarterly cash-settled futures on Compute Desk's H100, H200, B200 and B300 indices, outside the US regulatory perimeter. It earlier listed a perpetual on Ornn's H100 index and had removed it by late August 2026.
- **Lighter:** Lists a perpetual future on Ornn's H100 rental index, outside the US regulatory perimeter, with minimal trading.
- **Bitget:** Since 8 September 2026 this offshore exchange has listed perpetual futures, with leverage of up to 10 times, on H100 and B200 rental indices that it attributes to Silicon Data. Silicon Data told the CFTC that it has not licensed its indices for perpetual contracts and is aware of unlicensed use.
- **Injective:** Launched a perpetual on 26 October 2026 on an index of five GPU models at fixed weights, built by LO:TECH.

Executive summary
1. Introduction
2. Capital is already moving
3. The object of trade
4. How compute is priced
5. Where price risk can be transferred
6. What hedging leaves behind
7. Conclusion
References

© 2026. Sphere Labs & Four Pillars. All Rights Reserved.

Executive summary
1. Introduction
2. Capital is already moving
3. The object of trade
4. How compute is priced
5. Where price risk can be transferred
6. What hedging leaves behind
7. Conclusion
References

Venues and contracts for compute prices, 7 November 2026

Venue	Contract	Reference	Settlement	Status
NYMEX	H100 and B200 monthly futures	Silicon Data	Cash	Under review
ICE Futures U.S.	H100 and B200 monthly futures	Ornn	Cash	Specified
ICE Futures U.S.	Energy-normalized compute futures	NATIVX	Cash	Announced
Nodal Exchange	Compute futures	Compute Desk	Cash	Announced
American Innovation Exchange	Compute futures with exchange for physical	Compute Desk	Cash	Under review
Liquid Compute	Exchange and clearing	Own index	Physical	License pending
Kalshi	Monthly price contracts	Ornn	Fixed payout	Live
Polymarket US	Price contracts	Ornn	Fixed payout	Certified, not listed
AX, Bermuda	H100, H200, B200 and B300 quarterly futures	Compute Desk	Cash	Live
Lighter	H100 perpetual futures	Ornn	Cash	Live, minimal trading
Bitget	H100 and B200 perpetual futures	Silicon Data, unlicensed	Cash	Live
Injective	Perpetual on a five-GPU index	LO:TECH	Cash	Live

Source: CFTC filings and product pages; ICE and Nodal notices; Architect, Kalshi and Liquid Compute releases; Polymarket US and Silicon Data comment letters; venue documentation

5.3 How contracts settle and clear

A compute contract can settle in one of two ways, and each places a different burden on the market. A cash-settled contract pays the difference between the agreed price and an index value at expiry, and no capacity changes hands. Under a physically delivered contract, the seller provides capacity of a stated grade at a stated location over a stated period, and the buyer pays for it.

Every compute contract trading or filed so far settles in cash. NYMEX's GPU1 and GPU2 and ICE's Ornn contracts settle on the monthly average of an index. Kalshi's contracts pay a fixed amount according to where an index settles, and Lighter's perpetual tracks Ornn's H100 index with no expiry.

Physical delivery exists today only in bilateral contracts, the reserved and multi-year offtake agreements in Exhibit 3, which fix price and volume now for later delivery. Their terms are negotiated case by case, so they are not interchangeable and do not trade on an exchange. Liquid Compute is building a physical order book intended to make capacity transferable, with derivatives listed on top of it, and its exchange and clearing applications are pending.

The two methods present a trade-off between how easily a contract can be listed and how closely its hedge fits the exposure (Exhibit 9). A cash-settled contract needs only a defensible index, and the CFTC expects most compute derivatives to settle in cash at first because delivering the underlying capacity poses infrastructure challenges. The hedge it provides is only as good as the match between the index and the capacity a party buys or sells.

A physically delivered contract removes the gap between index and spot for capacity of the contract

© 2026. Sphere Labs & Four Pillars. All Rights Reserved.

Executive summary
1. Introduction
2. Capital is already moving
3. The object of trade
4. How compute is priced
5. Where price risk can be transferred
6. What hedging leaves behind
7. Conclusion
References

grade at the contract location, because delivery at expiry forces the futures price to converge on spot. The price of this is fragmentation. Capacity cannot be stored or shipped to another site, so each grade, location and delivery window would form a separate contract and divide the market's liquidity among them.

The two methods also cover different risks. A cash-settled hedge pays when prices rise but leaves the buyer to find capacity when it needs it, whereas delivery secures the capacity itself, which for many buyers is the larger risk.

Cash settlement and physical delivery compared for compute, November 2026

	Cash settlement	Physical delivery
Requirements	An index that can be computed and defended against manipulation	A deliverable grade, verification of delivered hardware, and a delivery location and window
Strength	Concentrates liquidity in one contract without defining a deliverable grade	Converges to spot for the contract grade at expiry
Weakness	Leaves basis between the index and the hedged capacity, and the settlement price rests on the integrity of index inputs	Divides liquidity across grades, locations and windows, because capacity cannot be stored or shipped
Risk covered	Price only; the hedger must still obtain capacity	Price and access to the capacity itself
Compute examples	NYMEX and ICE futures, Kalshi contracts, offshore futures and perpetuals	Reserved and offtake contracts, bilaterally; Liquid Compute's planned order book

Source: provider and venue documentation; CFTC request for comment on compute derivatives; Liquid Compute

Exchange for Physical (EFP) transactions link the two methods by letting the holders of opposite futures positions swap them for actual capacity. A laboratory hedging its future compute costs holds a long H100 futures position, and a neocloud hedging its future rental revenue holds a short one. The two can agree privately on a month of capacity and report both trades to the exchange, which closes the positions against each other.

The two parties price the capacity at the index plus or minus a negotiated basis that reflects the particular chip, location and period. A single cash-settled contract can then hold the market's liquidity and still end in delivery, without the exchange defining a deliverable grade. ComputeConnect, which Architect and Compute Desk announced on 8 July 2026, would pair cash-settled futures on the American Innovation Exchange with EFP delivery into H100, H200, B200 and B300 capacity, subject to regulatory review.

No compute futures contract is cleared with margin under US regulation today, although AX margins its offshore futures itself. ICE has named ICE Clear Europe and Nodal has named Nodal Clear for the contracts they have specified. Liquid Compute has applied to become a clearing organization, and Kalshi's contracts clear on a fully collateralized basis. Clearing would also matter for the lending described in Section 2. A cleared hedge carries no bilateral counterparty risk, so a lender could in principle accept a hedged borrower with less offtake support.

Margin is harder to set for compute than for mature energy contracts. Initial margin is meant to cover the loss a clearing house could face while it closes out a defaulting member's position, and it is usually sized from the contract's past price moves. Two features of compute make that harder:

© 2026. Sphere Labs & Four Pillars. All Rights Reserved.

- **Generation jumps:** The launch of a new chip generation can cut the price of the previous one in a single move larger than anything in the contract's history. Margin sized on that history would not cover such a move, and the clearing house would absorb the shortfall if a member defaulted.
- **Collateral that falls with the position:** A clearing member that is long compute and posts GPUs as collateral would see both lose value when compute prices fall, so its collateral would shrink just as its losses grew. Clearing houses already refuse a member's own securities as collateral to avoid this kind of wrong-way risk, and the same logic would argue against accepting GPUs.

5.4 Where hedging demand falls short

A futures market needs parties with opposite exposures who trade in it repeatedly. In compute, providers appear willing to sell forward but rarely do so in a standard contract that can be bought and sold on a listed market.

On the sell side, the parties with a real hedging need are neoclouds hedging their future rental revenue, and lenders secured on GPUs carry the same risk of falling prices. Providers already sell forward by renting capacity on long terms. Bandi and Su infer forward prices from the gap between short and long rental terms and find that buying them forward would have earned annualized returns of 7.5% for A100, 11.2% for B200 and 26.2% for H100. Forward prices sat below the spot prices that followed, which the authors, describing the evidence as preliminary, read as providers accepting a discount to lock in revenue.

The financing structure in Section 2 points the same way. Lenders lend against contracted revenue, so a multi-year contract makes a provider's GPUs easier to finance, and it also fixes utilization and often brings earlier payment. Providers therefore have reasons beyond price hedging to accept less for a long commitment.

These sales take the form of private long-term lease contracts rather than a standard contract traded on a listed market, and disclosure helps keep them there. A neocloud that sold forward at a discount in a listed market would reveal its own view of how fast its equipment loses value. If that pace is faster than the useful lives used for its accounting depreciation, the equity valuation and lenders' credit analysis that assume those lives could come under review. A private long-term lease keeps such a discount out of view.

When capacity is scarce, the premium on short contracts can lead providers to hold back forward sales altogether. Nebius said in August 2026 that it could sell its entire 2027 capacity on its current contract terms but was deliberately keeping some for shorter contracts, which it priced at roughly twice the rate per megawatt. It is funding its build-out partly with a $3.5 billion facility from Citi that matures in 364 days.

Buyers have clear reasons to buy forward. Laboratories commit to compute budgets before they know their revenue, and companies that sell AI features at fixed prices carry the cost of the compute behind them. Fewer of them, however, need a listed price hedge than these reasons suggest. A laboratory's larger exposure is often whether it can obtain capacity at all, and a cash-settled contract pays money rather than delivering GPUs. A company selling at fixed prices needs a hedge only while its customer prices do not follow compute costs, and one that buys through fixed-rate API or cloud agreements has little exposure to a GPU benchmark. Buyers can also move to newer chips as the economics shift, while a provider's revenue stays tied to the generation it owns.

Executive summary
1. Introduction
2. Capital is already moving
3. The object of trade
4. How compute is priced
5. Where price risk can be transferred
6. What hedging leaves behind
7. Conclusion
References

© 2026. Sphere Labs & Four Pillars. All Rights Reserved.

6. What hedging leaves behind

Executive summary
1. Introduction
2. Capital is already moving
3. The object of trade
4. How compute is priced
5. Where price risk can be transferred
6. What hedging leaves behind
7. Conclusion
References

A hedge converts price risk into basis risk, the risk that the hedging contract and the exposure it covers move apart. In compute, the drift between chip generations creates basis that no existing or proposed mechanism would bound. A listed contract is useful to a hedger only if that residual risk has a limit that can be measured.

6.1 Six causes of basis risk

Basis is manageable when something bounds how far it can widen. Exhibit 10 sets the six causes of basis that Friedman identifies against that test. In traded commodities the bound comes from conversion, transport, storage or a settlement rule. A bounded basis can be large and volatile, and a risk management function can still set a haircut from its history.

Six causes of basis in compute and the mechanism that bounds each, November 2026

Cause	Gap between	Bounding mechanism in traded commodities	Compute
Liquidity	Paper and physical in an immature market	Accumulated volume	Narrows with volume
Location	Contract delivery point and the hedger's location	Transport and arbitrage limits	Partly bounded by moving workloads between regions
Quality	Contract specification and the hedger's asset	Blending, refining and delivery specifications	Not yet bounded
Tenor	Contract month and the hedger's exposure period	Storage and carry	Not yet bounded; physical delivery would tie each month to spot
Settlement	Settlement index and the hedger's realized price	Physical delivery or the settlement rule	Bounded to its own index only
Obsolescence	Reference generation and the hedger's generation	None	None, and no known route to one

Source: Friedman, Compute Derivatives Basis Risk Primer; exchange filings

Settlement basis is the easiest cause to measure today, because the indices behind the filed contracts can be compared directly. Each contract converges to its own index at expiry, which makes settlement basis look bounded, but the bound protects only a hedger whose costs track that index.

A hedger carries the gap whenever its revenue or costs follow something other than what its settlement index measures. For example, a provider whose revenue follows on-demand rates and that hedged with a contract settling on an index built from a different mix of rentals and providers would have faced a gap of about $0.40 per GPU-hour in one direction in early 2025 and about $0.86 in the other by April 2026. Bandi and Su find the same separation between the two H100 indices behind the filed contracts, with some pairs negatively correlated, and attribute it to differences in index construction or in the market segments each provider covers. Weekly price changes point the same way. Fernández-Fuertes and Grégoire find that weekly changes in the two H100 indices correlated at 0.17 from late 2024 to July 2026, so either index would have removed about 3% of the other's weekly variance as a hedge. The authors note that their data cannot tell whether the gap comes from how the indices are built or from real differences in price. Either way, a hedger's settlement basis depends on how closely its chosen index matches the capacity it buys or sells.

© 2026. Sphere Labs & Four Pillars. All Rights Reserved.

6.2 What a generation shift leaves unhedged

Generational obsolescence is the one cause of basis in compute that no existing or proposed mechanism would bound. No conversion process turns an H100 into a B200, no transport route carries one into the other, and no settlement rule fixes the relationship between their indices. The inability to store compute does not explain the absence of a bound, because power cannot be stored either and its basis is bounded by fuel conversion and transmission.

A provider whose fleet is migrating from H100 to B200 cannot fix in advance how much of its revenue to hedge in each contract. Its hedge and its revenue diverge as the mix shifts, and the divergence starts again with each generation. Hedge effectiveness would therefore improve within a generation as volume accumulates and fall back at each generational boundary.

On Ornn's transaction-based daily indices, the price relationship between chip generations has been unstable. From September 2025 to August 2026, B200 rented at between 1.5 and 2.7 times the hourly rate of H100, and weekly changes in the two indices correlated at 0.25. A provider hedging B200 revenue with an H100 contract over that period would have offset little of its weekly variation, even though Ornn's indices for A100, H100, H200 and B200 all rose over the year. Nor does an older generation fade on a predictable path, as the A100's record in Section 2.3 shows.

The NYMEX and ICE designs specify one contract per generation, which turns the drift between generations into a spread that a desk can price. The NYMEX rules allow the two contracts to trade as a spread in one transaction, but no dedicated contract on the spread between generations has been filed. Nor does any venue reviewed offer standardized generation swaps or residual value puts, which would pay the difference if a GPU's resale value fell below a set floor.

Protection is emerging outside exchanges first. Architect structures bespoke over-the-counter compute swaps, and Liquid Compute arranges over-the-counter put options on its own rental index between hedgers and third parties. NVIDIA has also held early talks with insurers about covering lenders' losses when a borrower defaults and its pledged GPUs sell for less than the debt, the Financial Times reported in September 2026. Such residual value cover, however, carries wrong-way risk. Claims arise when borrowers default, which is most likely in a downturn for GPU cloud providers. The likeliest buyers of the seized chips are other providers that depend on the same lenders, so resale values would be lowest just when the insurer has to pay.

Executive summary
1. Introduction
2. Capital is already moving
3. The object of trade
4. How compute is priced
5. Where price risk can be transferred
6. What hedging leaves behind
7. Conclusion
References

© 2026. Sphere Labs & Four Pillars. All Rights Reserved.

7. Conclusion

Executive summary
1. Introduction
2. Capital is already moving
3. The object of trade
4. How compute is priced
5. Where price risk can be transferred
6. What hedging leaves behind
7. Conclusion
References

7.1 Where the market stands

What this report finds is a market built in reverse order. In crude oil a standard for what is traded came first, and price discovery, hedging and clearing were layered on top of it in turn. In compute, capital arrived first and indices and futures followed, while the unit and grade of what is traded and margined clearing are not yet in place. Benchmarks and venues are multiplying quickly, but the foundation they rest on is still missing.

The reversal shows directly in how capital handles compute price risk. With no market price to underwrite, lenders look to credit instead. Loans repaid by a strong customer's contract are funded at low rates, and rates rise the further a loan runs beyond its contracts or the less proven its borrower. For lenders, compute price risk is a secondary factor that surfaces only in renewal assumptions and loans secured on GPUs. Its size is set by the borrower's estimate rather than by a market price.

Once listed futures begin trading, part of this gap could be filled. Hedging, however, only converts price risk into basis risk. Basis from location, specification and settlement index can narrow as volume builds and as grading and physical delivery take hold, but the drift between chip generations has no such mechanism yet. On the evidence, what remains before compute trades as a commodity has less to do with opening more venues than with three foundations. They are trust in settlement indices, real hedging demand on both the buy and sell sides, and a mechanism to bound the risk of generational change.

7.2 Open questions

7.2.1 Whether regulators will accept compute indices as settlement references

The CFTC has asked two questions. The first is whether an index calculated mainly from private bilateral trades can meet its core principle that a contract must not be readily susceptible to manipulation, and the second is how to stop a provider from moving an index built partly from prices that providers themselves post. Every compute future filed so far settles on an index that relies on private trades or posted prices. The answer will therefore shape whether and on what terms they list.

The more than 100 comments posted by the end of October agree that the settlement index is the central question. Most exchanges and trading firms would accept an index that follows benchmark standards and favors executed trades, while others conclude that none yet qualifies. One letter found that weekly changes in four published H100 indices correlated at only 0.02 to 0.31 over fifteen months. It attributes the dispersion to each index sampling a different slice of a heterogeneous market that no arbitrage connects, so a contract on one index would hedge only the slice that index measures. Several index providers are also tied to trading businesses, a conflict that natural gas indices had to address after false trade reports drew more than $130 million in CFTC penalties in 2002 and 2003.

The CFTC's request for a further 60 days is consistent with those questions remaining open. The point to watch is what the Commission does with the NYMEX contracts by 8 January 2027 and what conditions, if any, it attaches, in particular on index governance and on the independence of administrators.

© 2026. Sphere Labs & Four Pillars. All Rights Reserved.

Executive summary
1. Introduction
2. Capital is already moving
3. The object of trade
4. How compute is priced
5. Where price risk can be transferred
6. What hedging leaves behind
7. Conclusion
References

7.2.2 What must be built first

Practitioners disagree on whether physical delivery has to work before financial contracts can succeed. One view holds that delivery infrastructure must come first, as pipelines and metering came before natural gas futures. The other holds that cash-settled contracts on a credible benchmark can launch before a deep physical market exists. The same author points out that oil, gas and metals each began with bilateral prepaid forwards on output from particular wells and mines and standardized only later, which would make compute's current bilateral contracts a normal first stage. The same split runs through the venues being built. Liquid Compute and two other prospective venues, Hyperbolic and OneChronos, favor starting from physical delivery, while the NYMEX, ICE and Nodal designs start from cash settlement.

The coming months will offer evidence on both sides. If open interest builds in the NYMEX contracts once they trade, cash-settled contracts will have drawn users before a deep physical market exists. If buyers and sellers use the physical delivery routes that ComputeConnect and Liquid Compute are building once those routes open, delivery will have begun to catch up.

7.2.3 Whether obsolescence can be managed

As Section 6.2 showed, a provider moving from H100 to B200 has to shift its hedge from one contract to the other as its fleet changes. The NYMEX rules allow the two contracts to trade as a spread, and the shift becomes easier still if the price gap between generations draws enough trading to form a market of its own.

A generational shift also shortens the tenor over which a contract is useful. In the next few months an H100 contract matches the chips that providers rent out. Two or three years ahead the H100 will have moved down to cheaper work as the A100 has, and buyers with a real hedging need will have moved on to newer generations. Trading would therefore likely concentrate in the near months among the 36 listed, and a provider that finds its hedge out of line with its revenue may not return to the contract. The point to watch is whether the spread between the generation contracts trades once they list.

7.2.4 Whether lenders will require compute price hedges

Lenders already require compute borrowers to hedge interest rates, and in one facility reviewed power costs as well, but no credit agreement found in this review requires them to hedge the price of the capacity they sell. The CoreWeave agreements count payments received under hedges as cash available for debt service, but those hedges cover interest rates and power, and the projected revenue that sizes the debt comes only from customer contracts. There is also a proposal to make a price hedge a condition of lending. Providers that have expanded with borrowed money would then bring steady selling to listed contracts, the form of forward selling that Section 5.4 found missing. The Federal Energy Regulatory Commission's orders of the late 1990s, starting with Order No. 888 in 1996, did something similar when they brought generators into central power markets.

There are two obstacles, however.

- First, a loan covenant is a private term that each lender sets, whereas the regulator's orders applied to a whole industry. A lender that insists on a hedge therefore risks losing the borrower to one that does not.
- Second, a provider that had sold forward would have to post margin when compute prices rose, before the higher revenue reached it. Metallgesellschaft's US oil subsidiary met the mirror image of

© 2026. Sphere Labs & Four Pillars. All Rights Reserved.

this problem in 1993. It had hedged ten-year supply contracts with short-dated futures and swaps, and when oil prices fell the margin calls came due long before the offsetting gains. The episode ended with losses of about $1.3 billion and a $1.9 billion rescue by its banks.

Lenders are unlikely to act until reference contracts trade. The first credit agreement that gives hedged revenue more borrowing capacity than unhedged revenue, or that requires a compute price hedge, would mark the shift.

Executive summary
1. Introduction
2. Capital is already moving
3. The object of trade
4. How compute is priced
5. Where price risk can be transferred
6. What hedging leaves behind
7. Conclusion
References

© 2026. Sphere Labs & Four Pillars. All Rights Reserved.

References

Executive summary

1. Introduction

2. Capital is already moving

3. The object of trade

4. How compute is priced

5. Where price risk can be transferred

6. What hedging leaves behind

7. Conclusion

References

Company filings and credit documents

- CoreWeave, 8-K for DDTL 4.0, closing release, credit agreement and limited guarantee, March 2026
- CoreWeave, 8-K for DDTL 5.5, closing release and credit agreement, August 2026
- CoreWeave, press release on DDTL 5.5, 10 August 2026, and Q2 2026 earnings call transcript, 11 August 2026
- CoreWeave, 10-K for 2025; 10-Q for the quarter to 30 June 2026, August 2026
- Meta Platforms, 10-Q for the quarter to 30 September 2025
- NVIDIA, 8-K on the SB Energy residual value guaranties, 17 August 2026, and 10-Q for the quarter to 26 July 2026
- Cerebras Systems, prospectus, May 2026
- Cipher Digital, 8-K and investor presentation on the Stingray notes, 8 June 2026, and 8-K on closing, 15 June 2026
- TeraWulf, 8-K on the Fluidstack joint venture, October 2025
- USD.AI, protocol documentation
- Nebius Group, 6-K on the Meta infrastructure agreement, March 2026; letter to shareholders, Q2 2026, 12 August 2026; Citi facility release, 22 October 2026
- Lambda, closing of $926 million term loan, 27 August 2026

Regulators and exchanges

- CFTC, NYMEX submission 26-370, 11 August 2026; review extension letter, 21 September 2026; product pages for the H100 and B200 contracts
- CFTC, Request for Comment on the Listing of Compute Derivatives Contracts, Federal Register, 21 August 2026
- CFTC, Polymarket US product certification, August 2026
- Comment letters on the request for comment, docket CFTC-2026-1850, October 2026, including OneChronos, Silicon Data, Polymarket US, Hyperbolic and Liquid Compute
- CFTC, enforcement releases 4728-02, 19 December 2002, 4846-03, 1 October 2003, and 4869-03, 25 November 2003
- ICE Futures U.S., exchange notice on Ornn compute futures, 29 September 2026
- NYMEX, Rulebook chapter 200, light sweet crude oil futures
- CME Group, 6fusion definitive agreement, 14 April 2014
- Universal Compute Xchange, beta launch release, 4 November 2014
- Nodal Exchange and Compute Desk, compute futures announcement, 3 September 2026

© 2026. Sphere Labs & Four Pillars. All Rights Reserved.

Executive summary
1. Introduction
2. Capital is already moving
3. The object of trade
4. How compute is priced
5. Where price risk can be transferred
6. What hedging leaves behind
7. Conclusion
References

- Architect, ComputeConnect announcement, 8 July 2026; American Innovation Exchange compute futures, May 2026; Liquid Inference launch, 7 October 2026
- Compute Exchange, website
- Liquid Compute, launch release, 15 September 2026, website and Hedging When Credit Tightens, 20 October 2026
- Kalshi, Compute has a Market-Implied Forward Curve Now, 14 July 2026, and market listings, accessed 1 November 2026
- Lighter, market specifications
- Architect, AX dated futures documentation and instrument list, accessed 1 November 2026, with archived lists of 28 July and 28 August 2026
- Bitget, compute perpetuals explained, September 2026
- Injective, Super Intelligence Compute Index launch, 26 October 2026
- Amazon Web Services, EC2 Spot and new Spot pricing model
- Anthropic, API pricing, accessed 16 October 2026

Ratings, indices and market data

- Moody's, rating of CoreWeave DDTL 4.0, March 2026
- Fitch Ratings, rating of CoreWeave DDTL 5.5, as syndicated, 17 July 2026
- Ornn, website and index methodology; Ornn OCPI daily index, archived 26 August 2026
- Silicon Data, index methodology, August 2026, and Silicon Index
- General Index, Compute Desk H100 US GPU Rental methodology factsheet, version 2, October 2026
- Ornn, The A100 Refuses to Die, 15 October 2026
- FINRA, TRACE trade reports for CoreWeave 9.25% senior notes due 2030, CUSIP 21873SAB4, and 9.625% senior notes due 2032, CUSIP 21873SAK4, May 2025 to October 2026

News reporting

- Bloomberg Law, DDTL 5.5 pricing, July 2026; data center debt, July 2026; CoreWeave bonds, 14 October 2026
- Los Angeles Times and Bloomberg, Big Tech debt, 10 July 2026
- IFR, Hyperion financing package, 17 December 2025
- DatacenterDynamics, NVIDIA DGX Cloud Lepton
- Financial Times, Nvidia turns to insurers to spread the risk of AI build-out, 29 September 2026
- PitchBook LCD, Volta launches $5B leveraged loan package, 9 October 2026; Bloomberg, JPMorgan, Goldman boost yield on Volta AI loan, 15 October 2026; Octus, Volta, JPM sound out investors for $5B GPU financing, 9 September 2026

© 2026. Sphere Labs & Four Pillars. All Rights Reserved.

Executive summary
1. Introduction
2. Capital is already moving
3. The object of trade
4. How compute is priced
5. Where price risk can be transferred
6. What hedging leaves behind
7. Conclusion
References

Research and articles

- Silicon Data and College of William & Mary, Did You Win the GPU Cloud Lottery? Benchmarking from TFLOPS to Tokens/$, GPGPU 2026, March 2026
- Federico Bandi and Yinan Su, (Early) AI Compute Asset Pricing, arXiv:2607.12156, version of 21 October 2026
- Rubén Fernández-Fuertes and Vincent Grégoire, HEC Montréal, Compute Futures: Who Is Left Exposed?, working paper, 3 August 2026
- Antonio Mello and John Parsons, Maturity Structure of a Hedge Matters: Lessons from the Metallgesellschaft Debacle, Journal of Applied Corporate Finance, 1995
- Shayon Sengupta, Multicoin Capital, Compute Capital Markets, 8 October 2026
- Ivan Sutherland, A futures market in computer time, Communications of the ACM, 1968
- Ribbit Capital, power letter, July 2026, compiling company data, Morgan Stanley Research estimates and BMI
- SemiAnalysis, The Great GPU Shortage: Rental Capacity, 2 April 2026
- Ariel Deschapell, co-founder and chief technology officer of Hydra Host, Bare Metal Is Crude Oil, Compute Doesn't Need a Unit. It Needs a Grading System. and Every Attempt to Commoditize Compute Has Failed
- Dave Friedman, The Market is Pricing Payment Certainty, Not GPUs, Compute Derivatives Basis Risk Primer and What Makes GPU Revenue Bankable
- David Lopez Mateos, Compute Desk, What does a B300 cost?, There is no GPU price, There is no GPU price hedge and Cerebras Is a Commodity Finance Story; Compute Desk, Compute Clear
- a16z, chart on the token and GPU price spread, using Ornn data
- Moonrunner, Moonletter, Inside the Fragmented Market for AI Compute
- Felix Stocker, Chip Futures, April 2024

This report makes no investment recommendation on any company or contract.

© 2026. Sphere Labs & Four Pillars. All Rights Reserved.

Contact

Sungjun Eun

Sphere Labs | Head of Research, Korea

moyed@spherepay.co

Jinsol Bok

Four Pillars | Head of Research

100y@4pillars.io

FOUR PILLARS

4pillars.io/en

x.com/FourPillarsFP

@eth100y

@100y_eth

© 2026. Sphere Labs & Four Pillars. All Rights Reserved.