CUSTODIA WEALTH MANAGEMENT SA
RESEARCH AND ANALYSIS

MARKETS AND ARTIFICIAL INTELLIGENCE

COMPUTING POWER AS A COMMODITY

GPU rental futures, bank exposure to the AI boom, the Anthropic IPO and the outlook for quantum computing

Situation as at 11 November 2026

Thematic report — Main sources: Financial Times, AFP (10 November 2026), Silicon Data and CME Group documentation

Document intended exclusively for professional and institutional investors

Contents

Introduction 2
1 Background: why renting computing power can be a commodity 4
1.1 What exactly is traded 4
1.2 The three conditions applied to compute 5
1.3 Why now 6
2 How the Silicon Data H100 Rental Index (SDH100RT) is built 8
2.1 Scope: why only neoclouds 8
2.2 Eligible and excluded data 8
2.3 Step one: conversion to a unit price 8
2.4 Step two: econometric basis adjustment 8
2.5 Two-tier aggregation and weights 9
2.6 Filters, missing data and expert judgement 10
2.7 Governance and conflicts of interest 10
2.8 The CME contract built on the index 10
2.9 The issues raised 11
3 The impact of the future on creditors and banks 13
3.1 Who is exposed 13
3.2 Positive effects 13
3.3 Risks 13
3.4 The link with political risk 14
4 Investment-bank financing of AI 16
4.1 The facts: banks have hit their limits 16
4.2 How the instrument works 16
4.3 What the price of protection tells us 16
4.4 Overall assessment 17
5 The link with the Anthropic IPO 18
5.1 Terms of the deal 18
5.2 The link exists, and it is substantial 18
5.3 Grounds for caution 19
6 The possible impact of quantum computing on AI 20
6.1 The facts 20
6.2 Implications 20
6.3 A cautious assessment 20
7 Conclusions 22
Sources 23

Introduction

Computing power for artificial intelligence is taking on the typical features of a raw material: it is an essential input that can be measured in a standard unit (the GPU-hour), its price is highly volatile, and supply and demand are represented by players with opposing interests and concrete reasons to hedge. On this basis CME Group has filed two cash-settled futures contracts on the rental prices of Nvidia H100 and B200 GPUs, using indices calculated by Silicon Data as the underlying. The launch, originally scheduled for 5 October 2026, has been postponed: the CFTC has extended its review by at least 60 days, pushing the start back to early 2027 at the earliest, over concerns that indices built on opaque prices in a concentrated market could be manipulated.

For creditors, and banks in particular, the future would have a twofold effect. On the one hand, it would offer for the first time a market price, observable every day and with a forward curve of up to 36 months, against which to value GPUs pledged as collateral and the cash flows of financed data centres; on the other, it would turn that price into a public signal capable of accelerating any correction, as happened with the ABX index in 2007.

The credit backdrop is one of banks that have reached their internal concentration limits on AI and are turning to credit insurance (Barclays has bought it, HSBC has looked at it, JPMorgan considered it and turned it down) in order to keep lending. The cost of protection, equal to 50–70% of the margin on the insured portion, shows that the market perceives a significant risk. The Anthropic IPO, which could raise more than $100bn at a valuation of $1.4–1.5tn, is the other side of the same coin: when bank balance sheets are saturated, risk capital becomes the outlet for funding compute spending estimated at around $900bn. Finally, quantum computing is today a medium-to-long-term promise (2029–2033) which, if fulfilled, would affect data-centre energy demand and the residual value of GPUs — precisely the variables that the future and lenders need to estimate.

$2.77	≥ 60 days	50–70%	> $100bn
H100 rental per GPU-hour (Silicon Data index, early October 2026)	Extension of the CFTC review: futures will not launch before 2027	Share of the bank's margin ceded to credit insurers	Potential proceeds of the Anthropic IPO, including the private placement

Figure 1. Key indicators of the report. Sources: Financial Times (3 October – 11 November 2026), Silicon Data.

In-Depth Box: the ABX index

The **ABX** was an index of credit derivatives on US subprime mortgages, launched in early 2006 by Markit, which became the “thermometer” of the subprime crisis.

How it worked. Each series of the index consisted of credit default swaps on a basket of 20 subprime mortgage securitisations (RMBS) issued in the previous six months. The series were rolled every six months (06-1, 06-2, 07-1, 07-2), and each was split into sub-indices by rating class: AAA, AA, A, BBB and BBB-. The price was quoted relative to par: 100 meant the market expected full repayment, while 50 implied expected losses of roughly half the principal. Buyers of protection on the index were betting on the mortgages deteriorating; sellers were betting on them holding up.

Why it became crucial. Before the ABX, subprime securities traded rarely and over the counter, so nobody saw a daily market price. The index provided, for the first time, a public price observed every day. From early 2007 the BBB and BBB- sub-indices collapsed, followed in 2007–2008 by the AAA tranches as well. Banks, insurers and funds began to use it to mark their own similar securities to market, even though these were not part of

the basket. Huge write-downs, margin calls and forced sales followed. The series planned for 2008 was never launched, because there were no longer enough new securitisations.

The criticism. Many argued that the ABX exaggerated losses. It was illiquid and dominated by protection buyers, and its prices also embedded a large premium for illiquidity and fear, not just expected mortgage losses. Used as an accounting reference, it would therefore have amplified the crisis beyond the fundamentals.

The link with this report. This is Rubinstein's point. A future on the GPU rental price would make a currently opaque market transparent, as the ABX did with subprime. If compute prices fell, that public number could become the benchmark for valuing the collateral of data-centre loans. It could thus trigger the same chain of write-downs and forced sales, regardless of how liquid trading actually was. This is why the article concludes that the future may matter more as a signal than as a market.

Box 1. The ABX index and the consequences of its adoption.

1 Background: why renting computing power can be a commodity

A futures contract is a standardised agreement, traded on an exchange and guaranteed by a clearing house, to exchange a given quantity of an asset at a future date at a price fixed today. For an asset to become the underlying of a successful future, being important is not enough: the literature cited by Marc Rubinstein in the Financial Times shows that between two-thirds and three-quarters of new contracts fail to reach and sustain sufficient volumes. There are essentially three necessary conditions: underlying prices volatile enough to make hedging worthwhile, a standard contract on which the market converges, and a broad and diversified base of buyers and sellers.

1.1 What exactly is traded

The underlying asset is not the chip itself but the **right to use it for one hour**: the GPU-hour. AI model developers, companies and researchers rarely buy GPUs; more often they rent them from hyperscalers (Amazon, Microsoft, Google) or from "neoclouds", specialist operators that buy GPUs, install them in data centres and sell capacity by the hour, by the month or under multi-year contracts. A spot market with observable prices therefore already exists; what is missing is a transparent, regulated forward market for longer-dated contracts.

The GPU-hour has a feature that makes it closer to electricity than to oil: it cannot be stored. An hour of compute not used today is lost forever. As a result, the classic arbitrage relationship between spot and forward prices (buy today, store and sell forward) does not work, and the futures price mainly reflects expectations about future supply and demand. It is the same logic as electricity futures, a market that has, incidentally, worked for decades.

In-Depth Box: DePIN networks

DePIN networks (Decentralized Physical Infrastructure Networks) are the blockchains that make it possible to trade GPU compute. The main ones for computing power are Render, Akash, io.net and Aethir.

How they work. They are marketplaces where those with idle GPUs (smaller data centres, repurposed miners, sometimes individual users) make them available, and those who need compute rent them. The network token has three functions. It is used to pay for the service, to reward GPU providers and, often, as collateral: on io.net, for example, providers deposit IO tokens as collateral to take part in the network. The main attraction is price: Akash offers access to an H100 at $1.20–1.80 an hour, against $4.50–5.50 on AWS.

A recent development concerns the link between tokens and real usage. Initially these networks grew by distributing new tokens to providers, i.e. through subsidies. They now tie the token's value to actual demand. In March 2026 Akash introduced a mechanism whereby, every time a customer pays for compute, AKT is automatically bought back and burned. Similarly, in June 2026 io.net linked token issuance and burning to network revenue, paying providers a stable dollar target.

They are still small. Aethir reports $166mn of annualised revenue and more than 150 enterprise customers. io.net was approaching $20mn of annualised revenue with 139,000 GPUs on the network. By comparison, Anthropic alone reports $81bn of annualised revenue. It is a niche, useful mainly for inference, rendering and less critical workloads, while the large labs rely on hyperscalers and neoclouds under multi-year contracts.

Difference from CME futures. These tokens are not derivatives on the price of compute. They are payment and incentive instruments on spot markets, and their price depends as much on crypto sentiment as on GPU demand. They therefore do not make it possible to lock in today the cost of a GPU-hour two years from now, which is exactly what the future offers.

The risks are twofold. First: the crypto exchange Hyperliquid is among those supporting compute futures before the CFTC — the crypto world wants to take part in this market; but the crypto world is not yet easy to frame and supervise, and the computing-power market requires rigorous traceability (much as happens with green energy and guarantees of origin). Much of the data on these networks comes from the projects themselves or from crypto-exchange blogs, not from independent auditors. Figures on revenue and usage should therefore be treated with caution. Second: the prices on these marketplaces are exactly the kind of data that worries index critics — thin platforms, with few transactions, easy to move. This is why Silicon Data excludes unverified data and caps the weight of each provider.

Box 2. The tokenisation of computational work: processor use per unit of time.

1.2 The three conditions applied to compute

Volatility. This is the best-met condition. The hourly rental of an H100 rose as high as $8 in early 2024, when the chip shortage was at its peak, before falling below $2 at the end of 2025. The Silicon Data index (see Figura 2) has ranged between $1.86 and $3.40 between September 2024 and November 2026: after its February 2025 low it climbed back to $2.96 in early November 2026. The structural cause is the long lead time for new capacity: Oracle's Project Jupiter, due in 2028, has suffered delays linked to a gas pipeline and to litigation over water and air-quality permits. When supply responds to demand with a lag of years, prices swing widely, and those who produce or consume capacity have an incentive to lock them in.

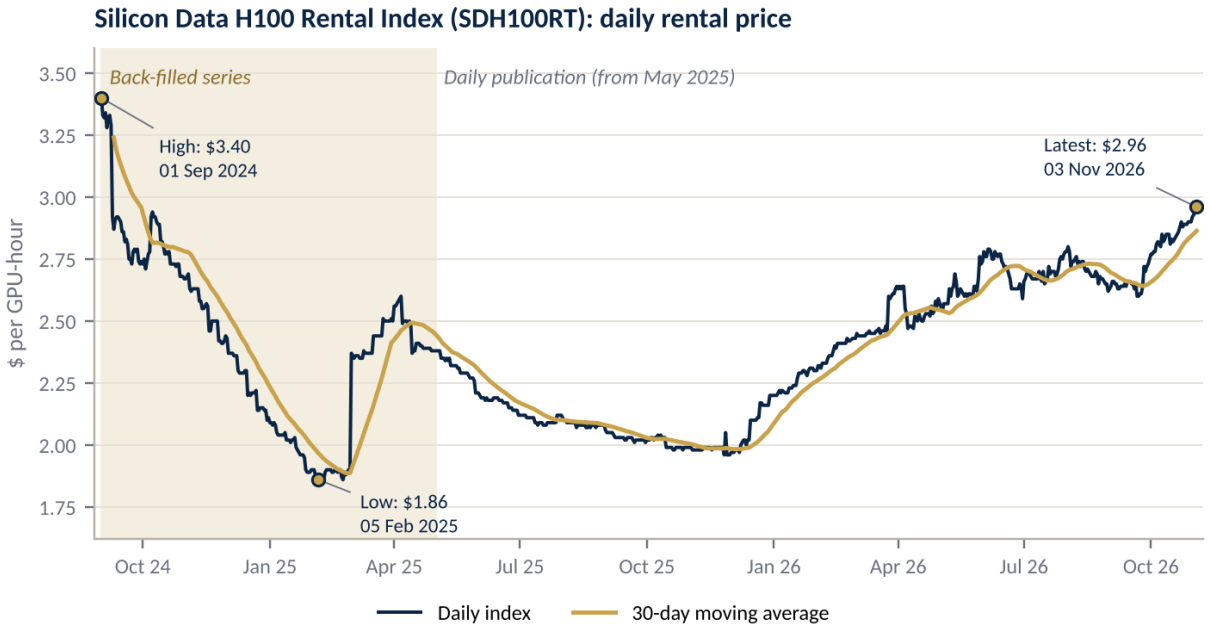


Figure 2. Silicon Data H100 Rental Index (SDH100RT), daily values in dollars per GPU-hour, 1 September 2024 – 3 November 2026, with 30-day moving average. The shaded area shows the back-filled series before daily publication began (May 2025). Source: Silicon Data; author's calculations.

Standardisation. This is the most delicate condition. The success story is WTI crude, listed since 1983 with delivery at Cushing, Oklahoma: although it represents a small share of world output, it has become the benchmark for the whole market. The failure is bandwidth in 1999, where service quality and reliability varied too much and the market collapsed together with its sponsor, Enron. A GPU-hour is not perfectly homogeneous: performance depends on cluster configuration, networking, software and location, and rapid obsolescence means an H100-hour cannot be substituted for a B200-hour. This is why CME lists separate contracts for each model, and why the heart of the project is an index that "normalises" the observations, as discussed in chapter 2.

Diversified participation. Compute is highly concentrated: Nvidia dominates chip supply; hyperscalers control much of the capacity and a handful of AI labs generate much of the demand. Rubinstein recalls that onion futures were banned in the United States in 1958 after two traders cornered the entire market. A "corner" in compute is unlikely in the same form, but concentration is exactly the point raised by the CFTC.

Condition	Oil (WTI)	AI compute (GPU-hour)	Assessment
Price volatility	High, driven by geopolitics and cycles	Very high: H100 from $8 to below $2 in about two years	Met
Standard asset	WTI delivered at Cushing	Indices by model (H100, B200) with econometric normalisation	Partial
Natural users	Producers and refiners, airlines	Neoclouds (sellers), AI labs and companies (buyers), lenders	Met
Diversified participation	Very broad	Concentrated: Nvidia, hyperscalers, a few labs	Weak
Storability	Yes	No (like electricity)	Price driven by expectations

Table 1. The conditions for a functioning futures market: oil compared with compute.

1.3 Why now

The size of the market explains the exchanges' interest: Boston Consulting Group estimates that spending on AI compute will rise from $360bn in 2025 to around $2.3tn in 2030. For CME, which already lists soybeans, metals, hogs, oil and interest rates, adding a product with this potential is a natural choice; the executive in charge explicitly compared the path to that of oil, which moved from spot trading to global derivatives, and BlackRock's Larry Fink spoke in May of the birth of a new asset class. The race is crowded: besides CME, Intercontinental Exchange (ICE, with Ornn indices — see Section 2.9 and Scheda 4) and Kalshi have announced similar products, and firms such as DRW, Jump Trading, Hyperliquid and Andreessen Horowitz have sent supportive comments to the CFTC. According to Andreessen, without a domestic futures market price discovery would migrate to fragmented or offshore venues.

In-Depth Box: the Kalshi exchange

Kalshi is a US exchange for "event contracts", so-called prediction markets. A Federal Reserve study describes it as the largest federally regulated prediction market, supervised by the CFTC — the same authority that is reviewing the CME and ICE futures. It was founded by Tarek Mansour and Luana Lopes Lara and is based in New York.

How it works. Each contract is a yes/no question about a future event. The contract pays $1 if the event occurs and nothing if it does not, so the price at which it trades equals the probability the market assigns to the event. An example on compute: a contract may ask whether a B200 will exceed $7 an hour by the end of 2026; if it costs $0.40, it implies a 40% probability. One can bet on elections, Fed decisions, inflation, sporting events, even the words spoken in a speech.

Size and controversy. In 2025 Kalshi recorded growth of more than 1,100% and total volume of $23.8bn, driven mainly by sports betting. According to DeFi Rate, contracts worth $137bn were traded on US prediction markets in the last 30 days, 76% of them on Kalshi. Sport is precisely the legal sticking point. Traditional bookmakers

and many states argue that this is betting dressed up as derivatives, and by mid-2026 Bloomberg counted lawsuits against Kalshi in at least 19 states.

Kalshi and AI compute. This is why the FT mentions it, and on this front Kalshi got there before CME and ICE:

- The first compute markets opened on 2 March 2026. They cover the hourly prices of the H100, H200, B200 and RTX 5090 and settle on Ornn index data — the same provider that ICE uses for its futures.
- In July Kalshi launched forward curves for the B200, H200 and A100. The curves are not tradable but, according to the company, are used to price bilateral compute deals. The curve is derived by aggregating the probabilities of the various yes/no contracts.
- The stated plan is to move later to perpetual futures on compute, i.e. contracts with no expiry traded around the clock, since data centres do not close at weekends.

Volumes remain small: according to Allium, by 27 July Kalshi's GPU markets had totalled $4.4mn of notional value. The current status is unclear. Kalshi has asked the CFTC for approval to relist compute contracts in the United States, which suggests they were suspended pending the close of the CFTC consultation, but this should be checked.

Difference from CME futures. A CME future pays in proportion to the index's deviation from the agreed price, so it covers the entire price exposure. A Kalshi contract instead pays a fixed amount if the price exceeds a given threshold, and nothing otherwise. For a neocloud or a bank that wants to fix a cost or a revenue, the future is the natural instrument. The binary contract lends itself more to speculation or to hedging extreme scenarios. Kalshi's advantages are accessibility, including for retail investors via an app, and the speed with which it produced a first forward price signal.

Kalshi reinforces two points: compute is becoming a financial underlying on several fronts at once; and Ornn is becoming a reference used by several exchanges, which matters in the competition with Silicon Data.

Box 3. Prediction markets: the Kalshi exchange and how it works.

2 How the Silicon Data H100 Rental Index (SDH100RT) is built

The SDH100RT index was launched by Silicon Data in May 2025 on Bloomberg terminals as the first daily benchmark for GPU rental prices[1]. It is part of a family that also includes the B200 (SDB200RT, the underlying of the second CME contract), A100, H200, B300, MI300X and an index of token spending on language models. The description that follows is based on the governance and methodology document published by Silicon Data, which states that it aligns with the IOSCO Principles for Financial Benchmarks.

2.1 Scope: why only neoclouds

The index that will serve as the underlying measures rental prices in the neocloud segment and deliberately excludes hyperscaler list prices. The reason is technical: hyperscaler list prices include ancillary services and enterprise service levels, are slow to react and do not represent the marginal price at which capacity actually changes hands. Neoclouds, by contrast, are where supply and demand meet first. Silicon Data publishes separate hyperscaler indices, which show a persistent price premium (around three times over the period measured), so that those exposed to that segment can measure and hedge the basis risk separately.

2.2 Eligible and excluded data

Only two categories of data enter the calculation: transactions actually executed, collected directly from providers and marketplaces under binding licence agreements and covering spot, on-demand and reserved rentals; and publicly available executable offers whose availability and price are verified. Excluded are static marketing price lists, data relayed by unverified aggregators, promotional or trial rates and any record lacking verified hardware metadata. At launch the company said it had used 3.5 million global price observations.

2.3 Step one: conversion to a unit price

Rental contracts take very different forms: a bare-metal server with eight GPUs rented for a month cannot be compared directly with a single-GPU instance paid by the hour. Each record is therefore converted into a standard price in dollars per GPU per hour:

$$Unit\ price = Total\ rental\ price \div (Number\ of\ GPUs \times Duration\ in\ hours)$$

Equation 1. Conversion of the rental price into a price per GPU per hour.

2.4 Step two: econometric basis adjustment

The unit price is not yet comparable, because it embeds premiums or discounts linked to the specific features of the contract. Each record is therefore brought back to a reference configuration (on-demand, in the United States, with standard hardware specifications for the GPU model) using a deterministic econometric model calibrated on historical data:

[1] The historical series, however, starts on 1 September 2024 and was reconstructed using data collected by Silicon Data before real-time prices were launched. The methodology used is, of course, the same as that used to calculate the index in real time.

$$Adjusted\ price = Raw\ unit\ price + Basis\ adjustment$$

Equation 2. Basis adjustment towards the reference configuration.

The adjustments cover three dimensions.

a) Contract duration: a long-term reserved commitment costs less than an on-demand rental and is converted to its on-demand equivalent.

b) Hardware configuration: memory, interconnect and cluster scale are brought back to the baseline specifications.

c) Geography: prices outside the United States or regional prices are brought back to the US reference.

The aim is that a shift in the composition of a given day's data (for example more long contracts or more European offers) should not produce a change in the index that does not correspond to a genuine price change.

2.5 Two-tier aggregation and weights

To prevent a provider with many quotes from dominating the result, the calculation takes place in two stages. In the first, the arithmetic mean of all of each provider's adjusted prices for the day is calculated, giving a single consolidated price per provider. In the second, the index is the weighted average of the consolidated prices of the active providers:

$$Index(t) = \sum_{j}^{N} w_j(t) \times P_j(t)$$

Equation 3. Daily index value as the weighted average of providers' consolidated prices.

where:

- $P_j(t)$ is the consolidated price of provider j on day t;
- $w_j(t)$ is the share with which the price of provider j enters the index on day t;
- N is the total number of providers.

Rule	Specification
Minimum number of contributors	At least 5 independent providers every business day
Weights	Static between reviews, distributed almost equally with limited adjustments for size and market share
Concentration cap	No provider may exceed 20% of the total weight
Panel entry	Quality review, at least 90 consecutive days of history, oversight committee approval
Rebalancing	Panel and weights reviewed at least every 3 months; model parameters locked under version control
Publication	Once per business day by 21:00 UTC via portal, API, Bloomberg, Refinitiv, Kaiko

Table 2. Index aggregation and governance parameters. Source: Silicon Data, Governance, Methodology & Control Framework.

2.6 Filters, missing data and expert judgement

Before entering the calculation, each provider–GPU pair is checked against historical ranges on four metrics: price level, daily price change, number of records and daily change in the number of records. Anomalies are suspended and verified; duplicates, records without hardware metadata, promotional rates, data older than 30 days and data from providers experiencing outages are permanently discarded. If a provider cannot be reached, its last valid price is used for up to 30 days, after which it is excluded and its weight redistributed. As a last resort, a price derived from a model based on historical data is used.

The calculation is automated; human intervention is allowed only in predefined cases, including the assessment of a daily change greater than 5%, with dual sign-off and approval by the head of compliance, and is publicly flagged when applied. Once published, the daily value is never retroactively restated — an essential rule for an index on which contracts are settled.

2.7 Governance and conflicts of interest

Oversight is entrusted to a benchmark oversight committee chaired by the head of compliance, with at least one independent member; a third-party calculation agent (Syntax Data) computes the final value from the normalised prices, and an independent audit firm issues an attestation on the controls. Silicon Data states that it does not trade compute derivatives and owns no GPU capacity. Two points nonetheless deserve attention. First, Silicon Data's chief executive also runs Compute Exchange, a spot GPU marketplace: the company cites information barriers and the executive's exclusion from index production, but the potential conflict of interest exists. Second, the list of providers contributing to the index is confidential (available only to the authorities): this protects against those who might try to front-run or manipulate the index, but it reduces external verifiability.

2.8 The CME contract built on the index

Feature	Silicon Data H100 Rental Index Futures
Code	GPU1 (B200: GPU2), listed under NYMEX rules
Size	730 GPU-hours, i.e. about one month of continuous use of one GPU
Price quotation	Dollars and cents per GPU-hour
Minimum tick	$0.01, equal to $7.30 per contract
Settlement	Financial (cash): no physical delivery of capacity
Maturities	Up to 36 months (according to the FT)
Status	Launch planned for 5 October 2026, postponed by the CFTC to early 2027 at the earliest; since 15 October CME has published an indicative, non-tradable forward curve

Table 3. Contract specifications. Sources: CME Group, Special Executive Report SER-9785 (11 August 2026).

To give an order of magnitude: at the price of $2.77 per hour reported by the FT in early October, an H100 contract is worth about $2,022 (730 × 2.77); a B200 contract at $5.86 is worth about $4,278. A 50-cent move in the index shifts the value of a contract by $365. A neocloud with 1,000 H100 GPUs produces about 730,000 GPU-hours a month, equal to 1,000 contracts: by selling them forward it can lock in today the revenue of the coming months or years. An AI lab worried about price increases would do the opposite.

2.9 The issues raised

The sticking point behind the postponement is how the index is built. The CFTC fears manipulation because of the opacity of rental prices and the dominance of a few providers, and is seeking assurances that exchanges can monitor prices derived from private deals. Some market participants argue that indices based on thinly traded short-term rental platforms are easy to move with a few transactions; a Columbia Business School professor has flagged the risk of affiliated entities trading with each other to create a false perception of demand. Moreover, there is still no consensus on how to measure prices: Ornn's competing index (used by ICE) does not always match Silicon Data's, and Silicon Data already revised its methodology in December 2025, restating historical values. The methodology described above (five-provider threshold, 20% cap, filters, verified transactions) is largely a response to these objections; the CFTC will have to decide whether it is sufficient for the index to qualify as a reliable underlying for launching its futures.

In-Depth Box: the OCPI index

The **OCPI (Ornn Compute Price Index)** is the competitor to Silicon Data's index. It is published by the New York start-up Ornn and is the underlying of the compute futures that ICE, the owner of the New York Stock Exchange, plans to list. It is the one Rubinstein refers to in the FT article when he notes that the two benchmarks do not always align.

What it is and how it is built. Ornn publishes the OCPI, which tracks actually traded spot prices of GPU compute for several models, including the Nvidia H100, H200 and B200. The feature Ornn stresses most is the origin of the data: according to the company, it is the first compute index built solely from transactions actually executed. This is the main stated difference from Silicon Data, which also uses executable public offers (verified listings) and normalises them with an econometric model. The OCPI series cover the H100, H200, B200, RTX 5090 and other models to be added as the market develops. The index is distributed on Bloomberg, and Ornn also runs its own platform for transferring compute risk; it says it executed the first cleared compute swap in December 2025.

The ICE contracts. Announced on 19 May 2026, a week after CME's. The first two planned are the OCPI H100 Future (symbol HPR) and the OCPI B200 Future (symbol BKL), to be listed on ICE Futures U.S. Like CME's, they are cash-settled in dollars, but with an important technical difference. They are monthly contracts whose settlement is based on the arithmetic mean of the index's daily prices over the month — so-called "Asian" settlement. The advantage is that moving the settlement price would require moving the index not on a single day but for the whole month. It is a direct response to the manipulation concerns raised by the CFTC.

	CME + Silicon Data	ICE + Ornn
Input data	Verified transactions and executable offers	Executed transactions only (stated)
Treatment	Econometric normalisation to a standard configuration	Traded spot prices
Initial models	H100, B200	H100, B200 (then H200, RTX 5090 and others)
Settlement	Cash, on the index	Cash, on the monthly average of daily values
Status	Awaiting CFTC approval	Awaiting approval

Why the two indices diverge. Measuring "the price of a GPU-hour" requires choices: which providers to include, whether to use offers or only transactions, how to treat duration, location and configuration. Different choices produce different numbers. One example is the B200 volatility recorded by Ornn: according to its index, the

spot rental of Blackwell rose 48% between mid-February and mid-April 2026, from $2.75 to $4.08 per GPU-hour. Over the same period Silicon Data reported the B200 at around $5–6. Until a single benchmark emerges, as happened with WTI in oil, anyone hedging with one index while having costs linked to the other carries basis risk.

Two critical observations:

- **The same kind of conflict of interest as Silicon Data.** Ornn publishes the index and also runs a compute trading venue. Silicon Data has a similar issue, as its CEO also runs the Compute Exchange marketplace.
- **"Transactions only" is more robust in theory.** It eliminates offers that nobody takes up. However, if transactions are few, the index becomes thin — precisely the risk some commenters flagged to the CFTC.

The historical parallel. ICE has been through a similar contest before. It owns the Brent contract, which established itself as a global benchmark alongside the WTI contract listed by CME. It is plausible that in compute, too, the two indices will coexist for a long time before the market chooses.

Box 4. Ornn's OCPI index and the ICE futures compared with CME and Silicon Data.

3 The impact of the future on creditors and banks

According to its supporters, a liquid futures market would help those who finance AI chips to value collateral using real-time market prices. From here, both positive effects and risks can be identified. It should be stressed that the link between futures and credit insurance developed in this chapter is an analytical interpretation reflecting the views of the author.

3.1 Who is exposed

There are three types of creditor exposed to the price of compute. Banks financing data-centre construction (project finance), whose repayment depends on the rents paid by hyperscalers and labs. Lenders providing loans secured on the GPUs themselves, typically to neoclouds, where the value of the collateral depends on how much the GPU can earn in rental over its remaining life. Finally, credit insurers, who are taking on a growing share of this risk from the banks.

3.2 Positive effects

Collateral valuation. Today the value of a GPU held as collateral is estimated using internal models and opaque price lists. A futures curve of up to 36 months would provide an independent reference price for the revenue expected from that GPU, which could be used to calculate loan-to-value ratios and to set more objective covenants.

Stabilising the borrower's cash flows. A neocloud that sells futures locks in the price of its future revenue. For the bank, this reduces the variability of debt-service capacity, just as in lending to oil companies the bank often requires the borrower to hedge part of its production. It is plausible that, once the market is liquid, banks will start requiring hedges as a condition of financing.

An additional hedging tool. Banks currently manage the risk with credit default swaps (increasingly expensive) and credit insurance. These instruments cover the borrower's default; the future would instead hedge the economic factor that often causes it, namely a collapse in the price of compute. It does not replace existing instruments but could reduce their cost, because a hedged borrower is less risky.

Pricing risk for insurers. Credit insurers, who today can afford to be very selective, would have a market data point for pricing protection on data-centre loans.

3.3 Risks

The price becomes a signal. As with the ABX index on subprime mortgages, which shed light on an opaque corner of finance and became the reference point for the ensuing sell-off, if the compute futures curve fell the decline would be immediately visible to all, could trigger collateral-value covenants, margin calls and forced sales, and thus amplify the fall. The conclusion is that the future could act more as a signal than as a market.

Basis risk. The index measures on-demand neoclouds with a standard US configuration. A data centre leasing to a hyperscaler under a multi-year contract, or a European neocloud with particular clusters, has different prices. The hedge might move in the right direction but not by the right amount.

Manipulation and index risk. If the index proved manipulable, a bank using it to value collateral would import that error into its own balance sheet.

Liquidity still to be built. With the launch postponed and most new futures failing, in the short term the benefit for banks remains potential.

3.4 The link with political risk

A relevant piece of context: in the midterm elections 38 ballot measures opposing data centres were passed. Political and permitting risk can delay or block projects, as in the Jupiter case. The future does not hedge this project-specific risk, but a general slowdown in new capacity would tend to support rental prices, partly compensating those already in operation.

In-Depth Box: Project Jupiter

Project Jupiter is one of the largest AI data-centre campuses under construction in the United States, and Rubinstein cites it in the FT as an example of why compute prices are volatile: new capacity arrives years late and with uncertain outcomes.

What it is. It is a 1,400-acre campus of more than 2 gigawatts in New Mexico, part of the Stargate programme of OpenAI and its partners, with an estimated cost of $165bn; Oracle is the tenant. It is located in Santa Teresa, Doña Ana County, near El Paso and the Mexican border; Oracle has an 18-year lease and the main customer is OpenAI. To give a sense of scale, at 2.5 gigawatts it could power more than half of New Mexico. According to the FT, it is due to come online in 2028.

The crux: energy. The campus will not connect to the public grid but will have its own microgrid: Oracle and its partner BorderPlex Digital Assets have announced that the main source will be Bloom Energy fuel cells. But the cells run on gas, and this is where the problems begin. The microgrid would still depend on a gas pipeline, which the state public lands commissioner has twice rejected for the section crossing state land, and whose construction appears to have been delayed until at least early next year.

The legal challenges. The microgrid needs an air-quality permit, and that is where the litigation began:

- In two orders, the New Mexico Supreme Court suspended the permit hearing and ordered the relevant state office to halt the use of water from a specific well for the construction site. The appeal was brought by an environmental group, which argued that the hearing was premature because the pipeline does not yet have its authorisations.
- The Court subsequently unanimously rejected the developers' request to resume the proceedings, even in part.
- As recently as last week, the authorities and the developer asked for more time to examine changes to the permit, with a further postponement of the hearing. According to a local press headline, the developer also waived the 23 November deadline for the permit decision; the full article could not be consulted.

In the background there is also the question of water in a desert area. Oracle maintains that cooling and fuel cells together will consume each year, averaged over 15 years, about as much as nine American households, but critics dispute the estimates.

Meanwhile construction continues. According to Oracle, nearly 700 local residents are already working on the site, and the company says it has awarded more than $994mn of contracts to New Mexico businesses.

The case touches on three themes:

- **Compute volatility.** If a campus of more than 2 gigawatts slips, a large amount of planned capacity does not arrive when expected. This is one of the factors that make the price of a GPU-hour hard to predict, and therefore worth hedging with a future.

- **Risk for lenders.** Buildings are going up while essential permits (energy, water) are still under discussion. The money spent is hard to recover, and a delay in permits pushes back the revenue that is supposed to repay the debt.
- **Political risk.** The Jupiter case is a concrete example of the local opposition to data centres that the FT documents through the 38 ballot measures passed in the midterm elections. It is also the kind of risk for which banks value credit insurance that pays out regardless of the cause of default.

No public information was found on which banks specifically finance Project Jupiter, so it cannot be linked directly to Barclays, HSBC or JPMorgan.

Box 5. The Jupiter case as an example of the aversion to large data centres, a likely factor in the outcome of the US midterm elections.

4 Investment-bank financing of AI

4.1 The facts: banks have hit their limits

Banks are increasingly turning to credit insurance to transfer part of the risk of their data-centre loans, so as to get around internal exposure limits on single companies, projects or sectors and keep lending. Barclays recently bought a policy to hedge its exposure to AI infrastructure; HSBC has looked at the instrument; JPMorgan, the largest US bank, considered it in recent weeks but decided not to buy it.

The starting point, then, is not a lack of capital but concentration. Banks have already lent heavily to AI and risk committees prevent them from adding exposure to the same names; insurance serves to “free up room” under those limits.

4.2 How the instrument works

In a typical deal, a bank that wants to lend a few hundred million to a single project transfers between 20% and 50% of the exposure to insurers. European and Canadian banks have long used credit insurance also to reduce their regulatory capital requirements; US banks generally do not get this benefit and have used credit default swaps. The rising cost of CDS and the volume of AI debt issued are now prompting several US banks to look at policies which, even without capital relief, allow them to comply with internal limits. The policy pays out on default regardless of the cause: the bank does not have to prove, for example, that the failure was due to political opposition to the data centre.

The market has broadened: there are now more than 70 credit insurers for project finance, including AIG, Chubb and Axa, compared with a handful a decade ago.

4.3 What the price of protection tells us

The most telling figure is the cost. Insurers charge annual premiums equal to 50–70% of what the bank earns on the insured portion of the loan, and these prices have withstood the general fall in credit-insurance rates. In illustrative terms: on a $500mn loan of which 40% ($200mn) is insured, with a hypothetical margin of 200 basis points the bank would earn $4mn a year on that portion and pay between $2mn and $2.8mn in premium.

If a bank agrees to give up much of its margin, the reason is not the return on the loan itself: banks pay to be able to make larger loans to a single data centre, to obtain lead-lender status or to show support for a client. In other words, lending is increasingly a relationship tool that opens the door to more profitable business: advisory, bond issuance, equity placements. On the other hand, the fact that insurers demand such a high premium and can afford to be selective shows that they perceive the risk as high.

Bank	Position on credit insurance for AI loans
Barclays	Has bought cover on its exposure to AI infrastructure loans
HSBC	Has explored using the instrument
JPMorgan Chase	Considered it in recent weeks, then decided not to buy it
US banks in general	Considering policies because of the rising cost of CDS, despite no capital benefit
European and Canadian banks	Have long used them, also to reduce capital requirements

Table 4. Use of credit insurance. Source: FT, 11 November 2026.

4.4 Overall assessment

The picture is that of a mature credit cycle: banks have concentrated exposures, are trying to distribute them to others (insurers, CDS investors) and accept thin margins to stay in the game. The risk does not disappear from the system; it shifts towards the insurance sector. Two factors make this a delicate moment: the tech sell-off after the midterm elections, put at $3.2tn and since recovered, showed how quickly the market can turn, with leverage amplifying the moves; and the growing doubts about the sustainability of the boom mentioned in connection with the Anthropic IPO.

5 The link with the Anthropic IPO

5.1 Terms of the deal

Item	Figure
Price range	$70–78 per share (filing of Tuesday 10 November)
Implied valuation	$1.26–1.41tn (fully diluted); target $1.4tn, possible increase to $1.5tn or more
IPO proceeds	Up to $64bn, plus about $10bn greenshoe if oversubscribed
Private placement	$26bn from Google, Nvidia and Microsoft at the IPO price
Potential total proceeds	Over $100bn with the higher price
Lead banks	Goldman Sachs and Morgan Stanley
Revenue	Annualised revenue of $81bn in October; $17bn in the third quarter
Planned spending	About $900bn on compute and infrastructure in the coming years
Morningstar valuation	$1.7tn fair value, within a $500bn–3tn range
Expected listing	Next week

Table 5. The IPO in numbers. Source: FT, 11 November 2026.

A few ratios help to read the numbers. At $1.4tn the valuation is about 17 times October's annualised revenue (1,400 / 81); at $1.5tn about 18.5 times. The FT's calculator reaches $1.4tn assuming revenue of $270bn in 2028 and a multiple of 5.2 times, lower than Microsoft's current one. Backers' expectations had aimed at $2tn before being scaled back. Meanwhile, some investors with an allocation are already reselling it above the IPO price through dedicated vehicles, a practice the company has tried to curb.

5.2 The link exists, and it is substantial

There are five concrete links between the Anthropic IPO and the financing of data centres dedicated to AI compute.

a) **Equity where bank balance sheets can no longer reach.** Three of Anthropic's backers say the company aims to raise as much as possible to support its compute spending, rather than to maximise its valuation. This is consistent with the picture in Section 4: if banks are close to their concentration limits and must pay dearly for protection, equity becomes the main alternative source. The proportions also make clear the limits of this source: over $100bn represents about 11% of the $900bn of planned spending. The rest will have to come from cash flow, partners and, inevitably, more debt, whether the company's own or that of the data centres working for it.

b) **The ultimate debtor of the data centres.** Anthropic is one of the few labs that generate much of the demand for compute. The loans banks make to data centres are repaid from the rents paid by customers such as Anthropic. An IPO of more than $100bn strengthens the balance sheet of one of the main end customers and therefore, indirectly, the credit quality of the whole chain. Conversely, a disappointing share performance after the listing would be read as a signal about the sustainability of those rents.

c) **The dual role of investment banks.** Goldman Sachs and Morgan Stanley are leading the offering. This illustrates why banks accept thin margins on data-centre loans: the AI ecosystem generates far richer underwriting and advisory fees, and being a lender is part of the commercial relationship.

d) **Suppliers financing their customers.** Nvidia, Google and Microsoft — the dominant chipmaker and two of the main capacity providers — are investing in the private placement. This is circularity: those who sell compute invest in those who buy it. It is the same concentration that worries the CFTC regarding the futures and is the market's weak point.

e) **The price of compute as a valuation variable.** Compute is an AI lab's main cost item. The price of the GPU-hour therefore affects Anthropic's margins, the value of the infrastructure and the solvency of lenders. A lab is a potential buyer of hedging. For equity investors, a public futures curve would become a real-time indicator of expected costs, with the "signal" effect already described.

5.3 Grounds for caution

The context is fragile: the listing comes a few days after a $3.2tn fall in tech stocks, and one major investor observed that in this market a weekend is an eternity. Morningstar's range, from $500bn to $3tn, reflects the huge uncertainty about the pace of adoption of AI tools. The comparison with OpenAI ($10bn of revenue in the third quarter, up 50%) shows rapid growth across the sector, but the sustainability of the ratio between revenue and investment remains the crux the market will have to assess.

6 The possible impact of quantum computing on AI

6.1 The facts

According to Agence France-Presse (AFP), the Australian start-up Diraq is conducting the world's first field test of a small silicon-chip quantum computer inside a commercial data centre, Equinix's facility in Sydney. Founder Andrew Dzurak argues that quantum computing could reduce the enormous energy needs of AI data centres and will be ready to do so within five years, because some of the problems tackled by AI would be better suited to quantum computation.

Milestone	Diraq target
Today (November 2026)	8 qubits on silicon chips, consumption below 20 kW
2029	150,000 qubits
2031	2 million qubits, DARPA threshold reached two years early; consumption of a few hundred kW
2033	DARPA deadline to demonstrate a useful quantum computer at sustainable cost

Table 6. Diraq's roadmap. Source: AFP, 10 November 2026.

The Australian landscape is lively: Diraq and PsiQuantum (backed by A$1bn of public investment) are among the six companies in the third stage of the DARPA programme, Silicon Quantum Computing was recognised by Time for a quantum-enhanced AI chip, and the country has more than 50 companies in the sector. The acknowledged limit is commercialisation: Dzurak himself stresses that everything depends on the ability to build machines at competitive cost.

6.2 Implications

Energy. The comparison is stark: a few hundred kilowatts for an advanced quantum system versus gigawatt-scale data centres. If even part of the workloads migrated, it would ease the energy constraint that is currently driving new power plants, pressure on grids and the local opposition documented by the 38 ballot measures.

Demand for and price of the GPU-hour. A credible technological substitute would reduce future GPU demand and hence the expected rental price. This directly affects the long end of the futures curve, which runs to 36 months, i.e. into 2029 — the very year in which Diraq is targeting 150,000 qubits.

Residual value of bank collateral. For creditors, quantum computing is an obsolescence risk that adds to the one within the GPU world itself (an H100 is overtaken by a B200). The maturities of data-centre loans and Anthropic's spending plan extend beyond 2030, so the 2029–2033 horizon falls within the life of today's loans.

New applications and a possible new commodity. Diraq points to uses in chemistry, pharmaceuticals and batteries, estimating that in drug trials quantum computing could raise the success rate by 10% and cut development costs by a third. In a success scenario, quantum computing hours could one day also become an underlying with an index of their own.

6.3 A cautious assessment

The forecasts come from the company's founder, who has an obvious interest in presenting them optimistically, and do not yet demonstrate a quantum advantage on typical AI workloads. Training and running large language models rely mainly on matrix multiplications, at which GPUs are extremely efficient; the prevailing consensus sees

quantum computing as a complementary accelerator for specific problems (molecular simulation, optimisation), installed alongside classical systems — as in the Sydney test — rather than as a general substitute. The most likely scenario in the coming years is therefore a hybrid one: a limited impact on GPU demand until the end of the decade, but a risk factor that long-term GPU lenders and those valuing the futures curve should start to consider in their adverse scenarios.

7 Conclusions

Renting computing power meets two of the three conditions for becoming a commodity with a futures market well (volatility and the presence of natural buyers of hedging) and only partly the third, standardisation, which the Silicon Data index addresses with a detailed methodology: conversion to a price per GPU-hour, econometric adjustment to a reference configuration, two-tier aggregation with concentration caps and governance based on the IOSCO principles. The CFTC's postponement shows that the weak point remains the concentration and opacity of the underlying market.

For banks, the future would be a useful valuation and hedging tool, but also a potentially pro-cyclical public signal. It arrives at a time when the banking system has already reached its credit exposure limits on AI and is shifting risk to insurers at a price that absorbs most of the margin. The Anthropic IPO fits into this same dynamic: it is the way the sector taps risk capital when credit is saturated, but even so it covers only a fraction of planned spending, and its outcome will be an indicator of market confidence in the whole financing chain. Finally, quantum computing is today a medium-to-long-term factor with many uncertainties, but it touches precisely the variables that determine the value of the chain (energy, GPU demand, residual value of collateral) and falls within the time horizon of the financing decided today.

Indicators to watch in the coming months: the CFTC decision and the initial liquidity of the CME and ICE contracts; convergence or divergence between the Silicon Data and Ornn indices; Anthropic's final price and post-listing performance; the cost of CDS and insurance premiums on data-centre loans; interim results of the DARPA quantum programme.

Sources

1. Financial Times, L. Harris and O. Aliaj, “Banks turn to credit insurance as they hit limits on lending to AI boom”, 11 November 2026.
2. Financial Times, M. Rubinstein, “The coming futures market in AI compute”, 3 October 2026.
3. Financial Times, M. Chan and J. R. Shah, “US regulator postpones Wall Street push to trade AI computing futures”, 10 November 2026.
4. Financial Times, G. Hammond, A. Pollard and G. Steer, “Anthropic bankers prepare to increase IPO valuation to $1.5tn or more”, 11 November 2026.
5. AFP, “Quantum could solve AI energy drain, says tech founder”, 10 November 2026.
6. Silicon Data, “GPU Rental Index Series: Governance, Methodology & Control Framework”, silicondata.com.
7. Silicon Data, SDH100RT index launch announcement, May 2025; Silicon Index product pages.
8. CME Group, Special Executive Report SER-9785, “Initial Listing of Two Compute Futures Contracts”, 11 August 2026.
9. Silicon Data, daily historical series of the Silicon Data H100 Rental Index (SDH100RT), 1 September 2024 – 3 November 2026.

This document is provided for information and research purposes only; it does not constitute financial advice, an investment recommendation or a solicitation to buy or sell financial instruments. The opinions expressed are those of the author as at November 2026 and are subject to change without notice.