CUSTODIA WEALTH MANAGEMENT SA
RICERCA ED ANALISI

MERCATI E INTELLIGENZA ARTIFICIALE

LA POTENZA DI CALCOLO COME COMMODITY

Futures sul noleggio di GPU, esposizione delle banche al boom dell'AI, IPO di Anthropic e prospettive del quantum computing

Situazione aggiornata all'11 novembre 2026

Report tematico — Fonti principali: Financial Times, AFP (10 novembre 2026), documentazione Silicon Data e CME Group

Documento destinato esclusivamente a investitori professionali ed istituzionali

Indice

Introduzione 2
1 Premessa: perché il noleggio di potenza di calcolo può essere una commodity 4
1.1 Che cosa si scambia esattamente 4
1.2 Le tre condizioni applicate al calcolo 5
1.3 Perché adesso 6
2 Come è costruito il Silicon Data H100 Rental Index (SDH100RT) 8
2.1 Perimetro: perché solo le neocloud 8
2.2 Dati ammessi ed esclusi 8
2.3 Primo passaggio: conversione in prezzo unitario 8
2.4 Secondo passaggio: aggiustamento di base econometrico 8
2.5 Aggregazione a due livelli e pesi 9
2.6 Filtri, dati mancanti e giudizio esperto 10
2.7 Governance e conflitti di interesse 10
2.8 Il contratto CME costruito sull'indice 10
2.9 Le criticità sollevate 11
3 L'impatto del future sui creditori e sulle banche 13
3.1 Chi è esposto 13
3.2 Effetti positivi 13
3.3 Rischi 13
3.4 Il nesso con il rischio politico 14
4 I finanziamenti all'AI delle banche d'investimento 16
4.1 Il fatto: le banche hanno raggiunto i limiti 16
4.2 Come funziona lo strumento 16
4.3 Che cosa dice il prezzo della protezione 16
4.4 Valutazione complessiva 17
5 Il collegamento con l'IPO di Anthropic 18
5.1 I termini dell'operazione 18
5.2 Il collegamento esiste, ed è sostanziale 18
5.3 Elementi di prudenza 19
6 Il possibile impatto del quantum computing sull'AI 20
6.1 Il fatto 20
6.2 Implicazioni 20
6.3 Una valutazione prudente 20
7 Conclusioni 22
Fonti 23

Introduzione

La potenza di calcolo per l'intelligenza artificiale sta acquisendo le caratteristiche tipiche di una materia prima: è un input essenziale e misurabile in un'unità standard (l'ora-GPU), il suo prezzo è molto volatile e la domanda e l'offerta sono rappresentate da operatori con interessi opposti che hanno ragioni concrete per coprirsi. Su queste basi CME Group ha depositato due contratti futures regolati per cassa, sui prezzi di noleggio delle GPU Nvidia H100 e B200, che usano come sottostante gli indici calcolati da Silicon Data. L'avvio, inizialmente previsto per il 5 ottobre 2026, è stato rinviato: la CFTC ha esteso la revisione di almeno 60 giorni, il che sposta la partenza almeno all'inizio del 2027, per timori legati alla manipolabilità di indici costruiti su prezzi opachi in un mercato concentrato.

Per i creditori, e in particolare per le banche, il future avrebbe un effetto duplice. Da un lato offrirebbe per la prima volta un prezzo di mercato, osservabile ogni giorno e con una curva a termine fino a 36 mesi, con cui valutare le GPU date in garanzia e i flussi di cassa dei data center finanziati; dall'altro trasformerebbe quel prezzo in un segnale pubblico capace di accelerare un'eventuale correzione, come accadde con l'indice ABX nel 2007.

Il contesto creditizio è quello di banche che hanno raggiunto i propri limiti interni di concentrazione verso l'AI e che ricorrono all'assicurazione del credito (Barclays l'ha acquistata, HSBC l'ha valutata, JPMorgan l'ha considerata e scartata) per continuare a erogare. Il costo della protezione, pari al 50–70% del margine sulla quota assicurata, indica che il mercato percepisce un rischio rilevante. L'IPO di Anthropic, che potrebbe raccogliere oltre 100 miliardi di dollari a una valutazione di 1.400–1.500 miliardi, è l'altra faccia della stessa medaglia: quando i bilanci bancari sono saturi, il capitale di rischio diventa la valvola di sfogo per finanziare una spesa in calcolo stimata in circa 900 miliardi. Infine, il quantum computing è oggi una promessa di medio-lungo periodo (2029–2033) che, se mantenuta, inciderebbe sulla domanda di energia dei data center e sul valore residuo delle GPU, cioè proprio sulle variabili che il future e i finanziatori devono stimare.

2,77 $	≥ 60 giorni	50–70%	> 100 mld $
Noleggio H100 per ora-GPU (indice Silicon Data, inizio ottobre 2026)	Proroga della revisione CFTC: avvio dei futures non prima del 2027	Quota del margine bancario ceduta agli assicuratori del credito	Raccolta potenziale dell'IPO di Anthropic, incluso il collocamento privato

Figura 1. Indicatori chiave del report. Fonti: Financial Times (3 ottobre – 11 novembre 2026), Silicon Data.

Box di Approfondimento: l'indice ABX

L'**ABX** era un indice di derivati creditizi sui mutui subprime statunitensi, lanciato all'inizio del 2006 da Markit e diventato il "termometro" della crisi dei subprime.

Come funzionava. Ogni serie dell'indice era composta da credit default swap su un paniere di 20 cartolarizzazioni di mutui subprime (RMBS) emesse nei sei mesi precedenti. Le serie si rinnovavano ogni semestre (06-1, 06-2, 07-1, 07-2), e ognuna era divisa in sottoindici per classe di rating: AAA, AA, A, BBB e BBB-. Il prezzo era espresso rispetto alla pari: 100 voleva dire che il mercato si aspettava il rimborso integrale, mentre 50 indicava perdite attese di circa metà del capitale. Chi comprava protezione sull'indice scommetteva sul deterioramento dei mutui, chi la vendeva sulla loro tenuta.

Perché divenne cruciale. Prima dell'ABX i titoli subprime erano scambiati poco e fuori borsa, quindi nessuno vedeva un prezzo di mercato quotidiano. L'indice fornì per la prima volta un prezzo pubblico, osservato ogni giorno. A partire da inizio 2007 i sottoindici BBB e BBB- crollarono, seguiti nel 2007-2008 anche dalle tranche AAA. Banche, assicurazioni e fondi iniziarono a usarlo per valutare a mercato i propri titoli simili, anche se questi

non facevano parte del paniere. Ne seguirono svalutazioni enormi, richieste di margine e vendite forzate. La serie prevista per il 2008 non fu mai lanciata, perché non c'erano più abbastanza nuove cartolarizzazioni.

Le critiche. Molti sostennero che l'ABX esagerasse le perdite. Era poco liquido e dominato da chi comprava protezione, e i suoi prezzi incorporavano anche un forte premio per l'illiquidità e la paura, non solo le perdite attese sui mutui. Usato come riferimento contabile, avrebbe quindi amplificato la crisi oltre i fondamentali.

Il collegamento con il report. È questo il punto di Rubinstein. Un future sul prezzo del noleggio delle GPU renderebbe trasparente un mercato oggi opaco, come fece l'ABX con i subprime. Se i prezzi del calcolo scendessero, quel numero pubblico potrebbe diventare il riferimento per valutare le garanzie dei prestiti ai data center. Potrebbe così innescare la stessa catena di svalutazioni e vendite forzate, a prescindere dall'effettiva liquidità degli scambi. Per questo l'articolo conclude che il future potrebbe contare più come segnale che come mercato.

Scheda 1. L'indice ABX e le conseguenze della sua adozione.

1 Premessa: perché il noleggio di potenza di calcolo può essere una commodity

Un contratto futures è un accordo standardizzato, negoziato in borsa e garantito da una cassa di compensazione, per scambiare a una data futura una certa quantità di un bene a un prezzo fissato oggi. Perché un bene possa diventare il sottostante di un future di successo non basta che sia importante: la letteratura citata da Marc Rubinstein sul Financial Times indica che tra due terzi e tre quarti dei nuovi contratti non riesce a raggiungere e mantenere volumi sufficienti. Le condizioni necessarie sono essenzialmente tre: prezzi del sottostante abbastanza volatili da rendere utile la copertura, un contratto standard su cui il mercato converga e una platea ampia e diversificata di compratori e venditori.

1.1 Che cosa si scambia esattamente

Il bene sottostante non è il chip in sé, ma il **diritto di usarlo per un'ora**: l'ora-GPU. Le società che sviluppano modelli di AI, le imprese e i ricercatori raramente acquistano GPU; più spesso le noleggiano da hyperscaler (Amazon, Microsoft, Google) o da "neocloud", operatori specializzati che comprano GPU, le installano in data center e ne vendono capacità a ore, a mesi o con contratti pluriennali. Esiste quindi già un mercato a pronti (spot) con prezzi osservabili; ciò che manca è un mercato a termine trasparente e regolamentato per i contratti di durata più lunga.

L'ora-GPU ha una caratteristica che la avvicina all'energia elettrica più che al petrolio: non è immagazzinabile. Un'ora di calcolo non usata oggi è persa per sempre. Ne consegue che la classica relazione di arbitraggio tra prezzo a pronti e prezzo a termine (comprare oggi, stoccare e vendere a termine) non funziona, ed il prezzo futures riflette soprattutto le aspettative su domanda e offerta futura. È la stessa logica dei futures sull'elettricità, mercato che peraltro funziona da decenni.

Box di Approfondimento: le reti DePIN

Sono chiamate reti **DePIN** (Decentralized Physical Infrastructure Networks), quelle blockchain che consentono di negoziare il calcolo su GPU. Le principali per la potenza di calcolo sono Render, Akash, io.net e Aethir.

Come funzionano. Sono marketplace in cui chi ha GPU inutilizzate (data center minori, miner riconvertiti, a volte singoli utenti) le mette a disposizione, e chi ha bisogno di calcolo le noleggia. Il token della rete svolge tre funzioni. Serve a pagare il servizio, a remunerare i fornitori di GPU e, spesso, come garanzia: per esempio, su io.net i fornitori depositano token IO come collaterale per partecipare alla rete. Il richiamo principale è il prezzo: Akash offre l'accesso a una H100 a 1,20–1,80 dollari l'ora, contro i 4,50–5,50 di AWS.

Un'evoluzione recente riguarda il legame tra token e uso reale. All'inizio queste reti crescevano distribuendo nuovi token ai fornitori, cioè con sussidi. Ora collegano il valore del token alla domanda effettiva. A marzo 2026 Akash ha introdotto un meccanismo per cui, ogni volta che un cliente paga il calcolo, AKT viene acquistato automaticamente e distrutto. Analogamente, a giugno 2026 io.net ha legato emissioni e distruzione dei token ai ricavi della rete, pagando ai fornitori un obiettivo stabile in dollari.

Le dimensioni sono ancora piccole. Aethir dichiara 166 milioni di ricavi annualizzati e oltre 150 clienti enterprise. io.net si avvicinava a 20 milioni di ricavi annualizzati con 139.000 GPU in rete. Per confronto, Anthropic da sola dichiara 81 miliardi di ricavi annualizzati. È una nicchia, utile soprattutto per inferenza, rendering e carichi meno critici, mentre i grandi laboratori si affidano a hyperscaler e neocloud con contratti pluriennali.

Differenza rispetto ai futures CME. Questi token non sono derivati sul prezzo del calcolo. Sono strumenti di pagamento e incentivo su mercati a pronti, e il loro prezzo dipende tanto dal sentiment crypto quanto dalla domanda di GPU. Non permettono quindi di bloccare oggi il costo di un'ora-GPU tra due anni, che è proprio ciò che il future offre.

I rischi sono duplici. Primo: l'exchange crypto Hyperliquid si inscrive tra i sostenitori dei futures sul calcolo presso la CFTC: il mondo crypto vuole partecipare a questo mercato; ma il mondo crypto non è ancora semplice da inquadrare e sorvegliare ed il mercato della potenza di calcolo esige una rigorosa tracciabilità (similmente a come avviene per l'energia verde e le garanzie d'origine). Gran parte dei dati su queste reti proviene dai progetti stessi o da blog di exchange crypto, non da revisori indipendenti. Le cifre su ricavi e utilizzo vanno quindi prese con prudenza. Secondo: i prezzi di questi marketplace sono esattamente il tipo di dato che preoccupa i critici degli indici: piattaforme sottili, con poche transazioni, facili da muovere. È per questo che Silicon Data esclude i dati non verificati e limita il peso di ogni fornitore.

Scheda 2. La tokenizzazione del lavoro computazionale: uso processore per unità di tempo.

1.2 Le tre condizioni applicate al calcolo

Volatilità. È la condizione meglio soddisfatta. Il noleggio orario di una H100 è salito fino a 8 dollari all'inizio del 2024, quando la scarsità di chip era massima, per poi scendere sotto i 2 dollari alla fine del 2025. L'indice Silicon Data (vedi Figura 2) è oscillato tra 1,86 e 3,40 dollari tra settembre 2024 e novembre 2026: dopo il minimo di febbraio 2025 è risalito fino a 2,96 dollari a inizio novembre 2026. La causa strutturale sono i lunghi tempi di realizzazione della nuova capacità: il progetto Jupiter di Oracle, atteso per il 2028, ha subito ritardi legati a un gasdotto e a contenziosi sui permessi per acqua e qualità dell'aria. Quando l'offerta reagisce con anni di ritardo alla domanda, i prezzi oscillano molto, e chi produce o consuma capacità ha un incentivo a bloccarli.

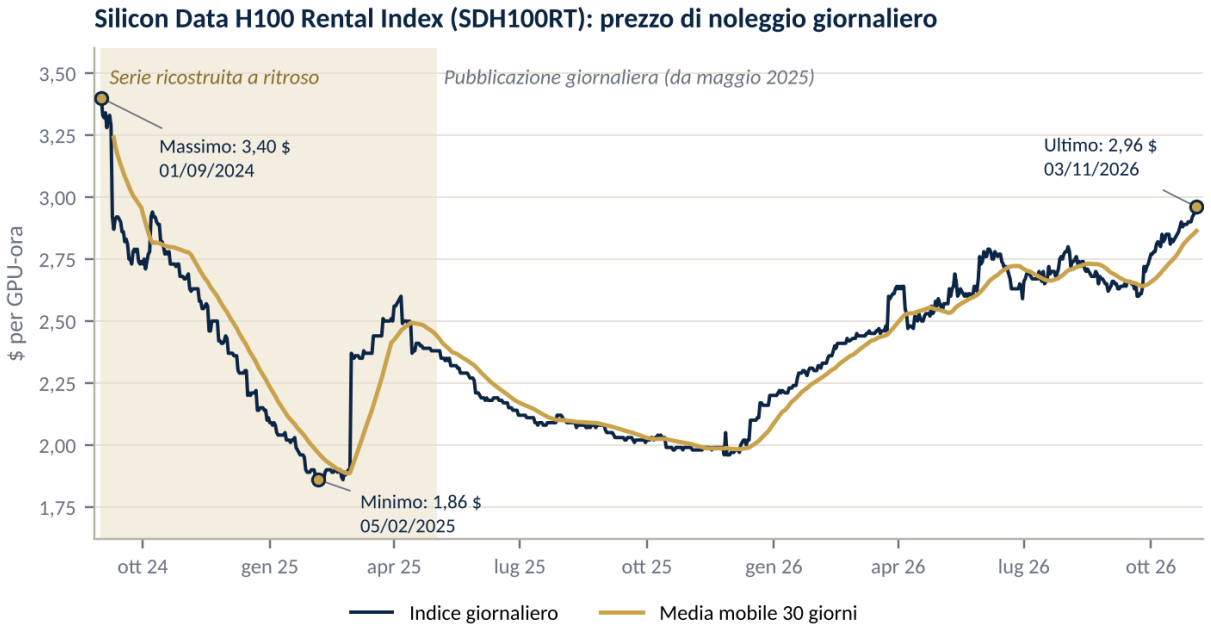


Figura 2. Silicon Data H100 Rental Index (SDH100RT), valori giornalieri in dollari per GPU-ora, 1° settembre 2024 – 3 novembre 2026, con media mobile a 30 giorni. L'area ombreggiata indica la serie ricostruita a ritroso prima dell'avvio della pubblicazione giornaliera (maggio 2025). Fonte: Silicon Data; elaborazione dell'autore.

Standardizzazione. È la condizione più delicata. Il caso di successo è il petrolio WTI, quotato dal 1983 con consegna a Cushing, Oklahoma: pur rappresentando una piccola quota della produzione mondiale, è diventato il riferimento per tutto il mercato. Il caso di insuccesso è la banda larga del 1999, dove qualità e affidabilità del servizio variavano

troppo e il mercato crollò insieme al suo promotore, Enron. Un'ora-GPU non è perfettamente omogenea: le prestazioni dipendono dalla configurazione del cluster, dalla rete, dal software e dalla località, e l'obsolescenza rapida rende un'ora H100 non sostituibile con un'ora B200. Per questo CME quota contratti distinti per ciascun modello e per questo il cuore del progetto è un indice che "normalizzi" le osservazioni, come si vedrà nel capitolo 2.

Platea diversificata. Il calcolo è molto concentrato: Nvidia domina l'offerta di chip; gli hyperscaler controllano gran parte della capacità e pochi laboratori di AI generano gran parte della domanda. Rubinstein ricorda che nel 1958 i futures sulle cipolle furono vietati negli Stati Uniti dopo che due operatori si erano accaparrati l'intero mercato. Un "corner" sul calcolo è improbabile nella stessa forma, ma la concentrazione è esattamente il punto sollevato dalla CFTC.

Condizione	Petrolio (WTI)	Calcolo AI (ora-GPU)	Giudizio
Volatilità del prezzo	Alta, guidata da geopolitica e cicli	Molto alta: H100 da 8 $ a meno di 2 $ in circa due anni	Soddisfatta
Bene standard	WTI consegnato a Cushing	Indici per modello (H100, B200) con normalizzazione econometrica	Parziale
Utilizzatori naturali	Produttori e raffinatori, compagnie aeree	Neocloud (venditori), laboratori AI e imprese (compratori), finanziatori	Soddisfatta
Platea diversificata	Molto ampia	Concentrata: Nvidia, hyperscaler, pochi laboratori	Debole
Immagazzinabilità	Sì	No (come l'elettricità)	Prezzo guidato dalle aspettative

Tabella 1. Le condizioni per un mercato futures funzionante, petrolio a confronto con il calcolo.

1.3 Perché adesso

Le dimensioni del mercato giustificano l'interesse delle borse: Boston Consulting Group stima che la spesa per calcolo AI passi da 360 miliardi di dollari nel 2025 a circa 2.300 miliardi nel 2030. Per CME, che già quota soia, metalli, suini, petrolio e tassi, aggiungere un prodotto con questo potenziale è una scelta naturale; il manager responsabile ha esplicitamente paragonato il percorso a quello del petrolio, passato dal mercato spot ai derivati globali, e Larry Fink di BlackRock ha parlato a maggio della nascita di una nuova classe di attività. La corsa è affollata: oltre a CME, Intercontinental Exchange (ICE, con indici Ornn, vedi la Sezione 2.9 e la Scheda 4) e Kalshi hanno annunciato prodotti simili, e operatori come DRW, Jump Trading, Hyperliquid e Andreessen Horowitz hanno inviato commenti favorevoli alla CFTC. Secondo Andreessen, senza un mercato futures domestico la formazione del prezzo migrerebbe verso sedi frammentate o estere.

Box di Approfondimento: la borsa Kalshi

Kalshi è una borsa statunitense di "contratti su eventi", i cosiddetti mercati di previsione (prediction market). Una ricerca della Federal Reserve la descrive come il più grande mercato di previsione regolamentato a livello federale e vigilato dalla CFTC, la stessa autorità che sta esaminando i futures CME e ICE. È stata fondata da Tarek Mansour e Luana Lopes Lara, e ha sede a New York.

Come funziona. Ogni contratto è una domanda con risposta sì/no su un evento futuro. Il contratto paga 1 dollaro se l'evento si verifica e zero se non si verifica, quindi il prezzo a cui si scambia equivale alla probabilità che il mercato attribuisce all'evento. Un esempio sul calcolo: un contratto può chiedere se una B200 supererà

i 7 dollari l'ora entro fine 2026; se costa 0,40 dollari, implica una probabilità del 40%. Si possono scommettere elezioni, decisioni della Fed, inflazione, eventi sportivi, persino le parole pronunciate in un discorso.

Dimensioni e controversie. Nel 2025 Kalshi ha registrato una crescita di oltre il 1.100% e un volume totale di 23,8 miliardi di dollari, trainato soprattutto dalle scommesse sportive. Secondo DeFi Rate, negli ultimi 30 giorni nei mercati di previsione statunitensi sono stati scambiati contratti per 137 miliardi di dollari, e Kalshi ne ha gestito il 76%. Proprio lo sport è il nodo legale. I bookmaker tradizionali e molti Stati sostengono che si tratti di scommesse mascherate da derivati, e a metà 2026 Bloomberg contava contenziosi contro Kalshi in almeno 19 Stati.

Kalshi e il calcolo AI. Questo è il motivo per cui l'FT la cita, e su questo terreno Kalshi è arrivata prima di CME e ICE:

- I primi mercati sul calcolo sono stati aperti il 2 marzo 2026. Coprono i prezzi orari di H100, H200, B200 e RTX 5090 e si risolvono sui dati dell'indice Ornn. È lo stesso fornitore che ICE usa per i suoi futures.
- A luglio Kalshi ha lanciato curve forward per B200, H200 e A100. Le curve non sono negoziabili ma, secondo la società, servono a prezzare accordi bilaterali sul calcolo. La curva si ricava aggregando le probabilità dei vari contratti sì/no.
- Il piano dichiarato prevede di passare in seguito a futures perpetui sul calcolo, cioè contratti senza scadenza negoziati senza interruzione, dato che i data center non chiudono nel fine settimana.

I volumi restano piccoli: secondo Allium, fino al 27 luglio i mercati GPU di Kalshi avevano totalizzato 4,4 milioni di dollari di valore nominale. Lo stato attuale non è chiaro. Kalshi ha chiesto alla CFTC l'approvazione per riquotare i contratti sul calcolo negli Stati Uniti. Questo fa pensare che siano stati sospesi in attesa che si chiuda la consultazione della CFTC, ma conviene verificare.

Differenza rispetto ai futures CME. Un future CME paga in proporzione allo scostamento dell'indice dal prezzo concordato, quindi copre l'intera esposizione al prezzo. Un contratto Kalshi paga invece una cifra fissa se il prezzo supera una certa soglia, e zero altrimenti. Per una neocloud o una banca che vuole fissare un costo o un ricavo, il future è lo strumento naturale. Il contratto binario si presta di più a speculare o a coprire scenari estremi. I vantaggi di Kalshi sono l'accessibilità, anche agli investitori al dettaglio tramite app, e la velocità con cui ha prodotto un primo segnale di prezzo a termine.

Kalshi rafforza due punti: il calcolo sta diventando un sottostante finanziario su più fronti contemporaneamente; e Ornn sta diventando un riferimento usato da più borse, il che conta nella competizione con Silicon Data.

Scheda 3. Il mercato delle previsioni: la borsa Kalshi ed il suo funzionamento.

2 Come è costruito il Silicon Data H100 Rental Index (SDH100RT)

L'indice SDH100RT è stato lanciato da Silicon Data a maggio 2025 sui terminali Bloomberg come primo benchmark giornaliero dei prezzi di noleggio delle GPU[1]. Fa parte di una famiglia che comprende anche B200 (SDB200RT, sottostante del secondo contratto CME), A100, H200, B300, MI300X e un indice sulla spesa in token dei modelli linguistici. La descrizione che segue si basa sul documento di governance e metodologia pubblicato da Silicon Data, che dichiara di allinearsi ai Principi IOSCO per i benchmark finanziari.

2.1 Perimetro: perché solo le neocloud

L'indice che fungerà da sottostante misura il prezzo di noleggio nel segmento delle neocloud ed esclude volutamente i listini degli hyperscaler. La ragione è tecnica: i listini degli hyperscaler includono servizi accessori e livelli di servizio enterprise, sono poco reattivi e non rappresentano il prezzo marginale a cui la capacità effettivamente si scambia. Le neocloud sono invece il punto in cui domanda e offerta si incontrano per prime. Silicon Data pubblica a parte indici sugli hyperscaler, che mostrano un premio di prezzo persistente (circa tre volte nel periodo misurato), così che chi ha esposizione verso quel segmento possa misurare e coprire separatamente il rischio di base.

2.2 Dati ammessi ed esclusi

Entrano nel calcolo due sole categorie di dati: le transazioni effettivamente eseguite, raccolte direttamente da fornitori e marketplace in base a contratti di licenza vincolanti, che coprono noleggi spot, on-demand e riservati; e le offerte pubbliche effettivamente eseguibili, la cui disponibilità e il cui prezzo sono verificati. Sono esclusi i listini di marketing statici, i dati rilanciati da aggregatori non verificati, le tariffe promozionali o di prova e qualunque record privo di metadati hardware verificati. Al lancio la società dichiarava di aver utilizzato 3,5 milioni di osservazioni di prezzo globali.

2.3 Primo passaggio: conversione in prezzo unitario

I contratti di noleggio hanno forme molto diverse: un server bare-metal affittato per un mese con otto GPU non è confrontabile direttamente con un'istanza a una GPU pagata a ore. Ogni record viene quindi convertito in un prezzo standard in dollari per GPU per ora:

$$\textit{Prezzo unitario} = \textit{Prezzo totale del noleggio} \div (\textit{Numero di GPU} \times \textit{Durata in ore})$$

Equazione 1. Conversione del prezzo di noleggio in prezzo per GPU per ora.

2.4 Secondo passaggio: aggiustamento di base econometrico

Il prezzo unitario non è ancora confrontabile, perché incorpora premi o sconti legati alle caratteristiche specifiche del contratto. Ogni record viene allora ricondotto a una configurazione di riferimento (on-demand, negli Stati Uniti, con specifiche hardware standard per il modello di GPU) tramite un modello econometrico deterministico calibrato sui dati storici:

[1] La serie storica parte invece dal 1° settembre 2024 ed è stata ricostruita con i dati raccolti da Silicon Data prima del lancio dei prezzi in tempo reale. La metodologia utilizzata è ovviamente la stessa usata per calcolare l'indice in tempo reale.

$$\textit{Prezzo aggiustato} = \textit{Prezzo unitario grezzo} + \textit{Aggiustamento di base}$$

Equazione 2. Aggiustamento di base verso la configurazione di riferimento.

Gli aggiustamenti riguardano tre dimensioni.

a) La durata contrattuale: un impegno riservato di lungo periodo costa meno di un noleggio on-demand e viene riportato al suo equivalente on-demand.

b) La configurazione hardware: memoria, interconnessione e scala del cluster sono ricondotte alle specifiche di base.

c) La geografia: i prezzi fuori dagli Stati Uniti o regionali sono riportati al riferimento statunitense.

Lo scopo è che uno spostamento nella composizione dei dati di un dato giorno (per esempio più contratti lunghi o più offerte europee) non produca una variazione dell'indice che non corrisponde a una vera variazione di prezzo.

2.5 Aggregazione a due livelli e pesi

Per evitare che un fornitore con molte quotazioni domini il risultato, il calcolo avviene in due stadi. Nel primo, per ogni fornitore si calcola la media aritmetica di tutti i suoi prezzi aggiustati del giorno, ottenendo un unico prezzo consolidato per fornitore. Nel secondo, l'indice è la media ponderata dei prezzi consolidati dei fornitori attivi:

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

Equazione 3. Valore giornaliero dell'indice come media ponderata dei prezzi consolidati dei fornitori.

dove:

- $P_j(t)$ rappresenta il prezzo consolidato del fornitore j nel giorno t;
- $w_j(t)$ rappresenta la quota con cui il prezzo del fornitore j entra nell'indice nel giorno t;
- N rappresenta il numero totale dei fornitori.

Regola	Specifica
Numero minimo di contributori	Almeno 5 fornitori indipendenti ogni giorno lavorativo
Pesi	Statici tra una revisione e l'altra, distribuiti in modo quasi uguale con correzioni limitate per dimensione e quota di mercato
Tetto di concentrazione	Nessun fornitore può superare il 20% del peso totale
Ingresso nel panel	Revisione di qualità, almeno 90 giorni consecutivi di storico, approvazione del comitato di sorveglianza
Ribilanciamento	Panel e pesi rivisti almeno ogni 3 mesi; parametri del modello bloccati sotto controllo di versione
Pubblicazione	Una volta al giorno lavorativo entro le 21:00 UTC via portale, API, Bloomberg, Refinitiv, Kaiko

Tabella 2. Parametri di aggregazione e governance dell'indice. Fonte: Silicon Data, Governance, Methodology & Control Framework.

2.6 Filtri, dati mancanti e giudizio esperto

Prima di entrare nel calcolo, ogni coppia fornitore-GPU viene confrontata con gli intervalli storici su quattro metriche: livello di prezzo, variazione giornaliera del prezzo, numero di record e variazione giornaliera del numero di record. Le anomalie vengono sospese e verificate; sono scartati definitivamente i duplicati, i record senza metadati hardware, le tariffe promozionali, i dati più vecchi di 30 giorni e quelli di fornitori in disservizio. Se un fornitore non è raggiungibile, si usa il suo ultimo prezzo valido per un massimo di 30 giorni, poi viene escluso e il suo peso redistribuito. Come ultima risorsa è previsto un prezzo ricavato da un modello sui dati storici.

Il calcolo è automatico; l'intervento umano è ammesso solo in casi predefiniti, tra cui la valutazione di una variazione giornaliera superiore al 5%, con doppia firma e approvazione del responsabile compliance, ed è segnalato pubblicamente quando applicato. Una volta pubblicato, il valore giornaliero non viene più rettificato retroattivamente, regola indispensabile per un indice su cui si regolano contratti.

2.7 Governance e conflitti di interesse

Il presidio è affidato a un comitato di sorveglianza del benchmark presieduto dal responsabile compliance, con almeno un membro indipendente; un agente di calcolo terzo (Syntax Data) calcola il valore finale a partire dai prezzi normalizzati, e una società di revisione indipendente rilascia un'attestazione sui controlli. Silicon Data dichiara di non negoziare derivati sul calcolo e di non possedere capacità GPU. Due elementi meritano però attenzione. Primo, l'amministratrice delegata di Silicon Data guida anche Compute Exchange, un marketplace spot di GPU: la società dichiara barriere informative e l'esclusione della manager dalla produzione dell'indice, ma il potenziale conflitto d'interessi esiste. Secondo, l'elenco dei fornitori che contribuiscono all'indice è riservato (accessibile solo alle autorità): è una protezione contro chi vorrebbe anticipare o manipolare l'indice, ma riduce la verificabilità esterna.

2.8 Il contratto CME costruito sull'indice

Caratteristica	Silicon Data H100 Rental Index Futures
Codice	GPU1 (B200: GPU2), quotato secondo le regole NYMEX
Dimensione	730 ore-GPU, cioè circa un mese di uso continuo di una GPU
Quotazione	Dollari e centesimi per ora-GPU
Tick minimo	0,01 $, pari a 7,30 $ per contratto
Regolamento	Finanziario (per cassa): nessuna consegna fisica di capacità
Scadenze	Fino a 36 mesi (secondo il FT)
Stato	Avvio previsto il 5 ottobre 2026, rinviato dalla CFTC almeno all'inizio del 2027; dal 15 ottobre CME diffonde una curva forward indicativa non negoziabile

Tabella 3. Specifiche del contratto. Fonti: CME Group, Special Executive Report SER-9785 (11 agosto 2026).

Per dare un ordine di grandezza: al prezzo di 2,77 $ per ora riportato dal FT a inizio ottobre, un contratto H100 vale circa 2.022 $ (730 × 2,77); un contratto B200 a 5,86 $ vale circa 4.278 $. Una variazione dell'indice di 50 centesimi sposta il valore di un contratto di 365 dollari. Una neocloud con 1.000 GPU H100 produce circa 730.000 ore-GPU al mese, pari a 1.000 contratti: vendendoli a termine può bloccare oggi il ricavo dei prossimi mesi o anni. Un laboratorio di AI che teme rincari farebbe l'operazione opposta.

2.9 Le criticità sollevate

Il nodo che ha portato al rinvio è la costruzione dell'indice. La CFTC teme la manipolazione a causa dell'opacità dei prezzi di noleggio e del dominio di pochi fornitori, e chiede garanzie che le borse possano monitorare prezzi derivati da accordi privati. Alcuni operatori sostengono che gli indici basati su piattaforme di noleggio a breve, poco liquide, siano facili da muovere con poche transazioni; un professore della Columbia Business School ha segnalato il rischio che entità collegate scambino tra loro per creare una falsa percezione di domanda. Inoltre non c'è ancora consenso su come misurare i prezzi: l'indice concorrente di Ornn (usato da ICE) non sempre coincide con quello di Silicon Data, e Silicon Data ha già rivisto la metodologia a dicembre 2025, ricalcolando i valori storici. La metodologia descritta sopra (soglia di cinque fornitori, tetto del 20%, filtri, transazioni verificate) è in buona parte una risposta a queste obiezioni; la CFTC dovrà stabilire se è sufficiente per qualificare l'indice come sottostante affidabile per lanciare i suoi future.

Box di Approfondimento: l'indice OCPI

L'**OCPI (Ornn Compute Price Index)** è l'indice concorrente di quello di Silicon Data. Lo pubblica la start-up newyorkese Ornn ed è il sottostante dei futures sul calcolo che ICE, la società proprietaria del New York Stock Exchange, vuole quotare. È quello che Rubinstein cita nell'articolo dell'FT quando osserva che i due benchmark non sempre coincidono.

Cos'è e come è costruito. Ornn pubblica l'OCPI, che segue i prezzi spot effettivamente scambiati del calcolo su GPU per diversi modelli, tra cui Nvidia H100, H200 e B200. La caratteristica su cui Ornn insiste di più è l'origine dei dati: secondo la società è il primo indice sul calcolo costruito solo da transazioni effettivamente eseguite. È la principale differenza dichiarata rispetto a Silicon Data, che usa anche le offerte pubbliche eseguibili (listing verificati) e le normalizza con un modello econometrico. Le serie OCPI coprono H100, H200, B200, RTX 5090 e altri modelli da aggiungere man mano che il mercato si sviluppa. L'indice è distribuito su Bloomberg, e Ornn gestisce anche una propria piattaforma per il trasferimento del rischio sul calcolo; dichiara di aver eseguito il primo swap sul calcolo a prezzi compensati nel dicembre 2025.

I contratti ICE. Annunciati il 19 maggio 2026, una settimana dopo quelli CME. I primi due previsti sono l'OCPI H100 Future (simbolo HPR) e l'OCPI B200 Future (simbolo BKL), da quotare su ICE Futures U.S. Come quelli CME sono regolati per cassa e in dollari, ma con una differenza tecnica importante. Sono contratti mensili il cui regolamento si basa sulla media aritmetica dei prezzi giornalieri dell'indice nel mese. Si tratta del cosiddetto regolamento "asiatico". Il vantaggio è che, per spostare il prezzo di liquidazione, non basta muovere l'indice in un singolo giorno: bisognerebbe farlo per tutto il mese. È una risposta diretta al timore di manipolazione sollevato dalla CFTC.

	CME + Silicon Data	ICE + Ornn
Dati di input	Transazioni verificate e offerte eseguibili	Solo transazioni eseguite (dichiarato)
Trattamento	Normalizzazione econometrica su configurazione standard	Prezzi spot scambiati
Modelli iniziali	H100, B200	H100, B200 (poi H200, RTX 5090 e altri)
Regolamento	Per cassa sull'indice	Per cassa sulla media mensile dei valori giornalieri
Stato	In attesa di approvazione CFTC	In attesa di approvazione

Perché i due indici divergono. Misurare "il prezzo di un'ora-GPU" richiede scelte. Bisogna decidere quali fornitori includere, se usare offerte o solo transazioni, come trattare durata, località e configurazione. Scelte diverse producono numeri diversi. Ne è un esempio la volatilità del B200 registrata da Ornn: secondo il suo indice, il noleggio spot di Blackwell è salito del 48% tra metà febbraio e metà aprile 2026, da 2,75 a 4,08 dollari per ora-GPU. Nello stesso periodo Silicon Data riportava un B200 intorno ai 5–6 dollari. Finché non emerge un unico riferimento, come accadde per il WTI nel petrolio, chi si copre con un indice e ha costi legati all'altro si porta dietro un rischio di base.

Due osservazioni critiche:

- **Lo stesso tipo di conflitto di interessi di Silicon Data.** Ornn pubblica l'indice e gestisce anche una sede di negoziazione sul calcolo. Silicon Data ha un problema analogo, perché la sua CEO guida anche il marketplace Compute Exchange.
- **"Solo transazioni" è più robusto in teoria.** Elimina le offerte che nessuno accetta. Però, se le transazioni sono poche, l'indice diventa sottile ed è proprio il rischio segnalato da alcuni commentatori alla CFTC.

Il parallelo storico. ICE ha già vissuto una competizione simile. È proprietaria del contratto sul Brent, che si è affermato come riferimento mondiale accanto al WTI quotato da CME. È plausibile che anche sul calcolo i due indici convivano a lungo prima che il mercato scelga.

Scheda 4. L’indice OCPI di Ornn e i futures di ICE a confronto con CME e Silicon Data.

3 L'impatto del future sui creditori e sulle banche

Secondo i sostenitori, un mercato futures liquido aiuterebbe chi finanzia i chip AI a valutare le garanzie con prezzi di mercato in tempo reale. Partendo da qui si possono individuare effetti positivi e rischi. È bene precisare che il collegamento tra futures e assicurazione del credito, sviluppato in questo capitolo, è un'elaborazione analitica che riflette le opinioni di chi scrive.

3.1 Chi è esposto

I creditori esposti al prezzo del calcolo sono di tre tipi. Le banche che finanziano la costruzione di data center (project finance), il cui rimborso dipende dai canoni pagati da hyperscaler e laboratori. I prestatori che concedono finanziamenti garantiti dalle GPU stesse, tipicamente a neocloud, dove il valore della garanzia dipende da quanto la GPU può rendere in noleggio nella sua vita residua. Infine gli assicuratori del credito, che stanno assumendo una quota crescente di questo rischio dalle banche.

3.2 Effetti positivi

Valutazione delle garanzie. Oggi il valore di una GPU in garanzia si stima con modelli interni e listini poco trasparenti. Una curva futures fino a 36 mesi darebbe un prezzo di riferimento indipendente per i ricavi attesi da quella GPU, utilizzabile per calcolare il rapporto tra prestito e valore della garanzia e per fissare covenant più oggettivi.

Stabilizzazione dei flussi del debitore. Una neocloud che vende futures blocca il prezzo dei suoi ricavi futuri. Per la banca questo riduce la variabilità della capacità di servizio del debito, così come nei finanziamenti alle compagnie petrolifere la banca spesso chiede al debitore di coprire una parte della produzione. È plausibile che, una volta liquido il mercato, le banche inizino a chiedere coperture come condizione del finanziamento.

Uno strumento di copertura in più. Le banche oggi gestiscono il rischio con credit default swap (sempre più costosi) e assicurazione del credito. Questi strumenti coprono l'insolvenza del debitore; il future coprirebbe invece il fattore economico che spesso la causa, cioè il crollo del prezzo del calcolo. Non sostituisce gli strumenti esistenti ma potrebbe ridurne il costo, perché un debitore coperto è meno rischioso.

Prezzo del rischio per gli assicuratori. Gli assicuratori del credito, che oggi possono permettersi di essere molto selettivi, avrebbero un dato di mercato per tariffare la protezione sui prestiti ai data center.

3.3 Rischi

Il prezzo diventa un segnale. Come nel caso dell'indice ABX sui mutui subprime, che rese visibile un angolo opaco della finanza e divenne il punto di riferimento della liquidazione successiva, se la curva futures del calcolo scendesse, il ribasso sarebbe immediatamente visibile a tutti, potrebbe far scattare covenant sul valore delle garanzie, richieste di margine e vendite forzate, e quindi amplificare la discesa. La conclusione è che il future potrebbe operare più come segnale che come mercato.

Rischio di base. L'indice misura le neocloud on-demand con configurazione standard negli Stati Uniti. Un data center che affitta a un hyperscaler con un contratto pluriennale, o una neocloud europea con cluster particolari, ha prezzi diversi. La copertura potrebbe muoversi nella direzione giusta ma non nella misura giusta.

Rischio di manipolazione e di indice. Se l'indice si rivelasse manipolabile, una banca che lo usa per valutare garanzie importerebbe quell'errore nei propri bilanci.

Liquidità ancora da costruire. Con l'avvio rinviato e la maggior parte dei nuovi futures che fallisce, nel breve periodo il beneficio per le banche resta potenziale.

3.4 Il nesso con il rischio politico

Un elemento di contesto rilevante: alle elezioni di metà mandato sono passate 38 misure referendarie contrarie ai data center. Il rischio politico e autorizzativo può ritardare o bloccare progetti, come nel caso Jupiter. Il future non copre questo rischio specifico del singolo progetto, ma un rallentamento generalizzato della nuova capacità tenderebbe a sostenere i prezzi di noleggio, compensando in parte chi è già operativo.

Box di Approfondimento: il progetto Jupiter

Il progetto **Jupiter** è uno dei più grandi campus di data center per l'AI in costruzione negli Stati Uniti, e Rubinstein lo cita sull'FT come esempio di perché i prezzi del calcolo sono volatili: la nuova capacità arriva con anni di ritardo e con esiti incerti.

Cos'è. È un campus da 1.400 acri e oltre 2 gigawatt nel New Mexico, parte del programma Stargate di OpenAI e dei suoi partner, con un costo stimato di 165 miliardi di dollari; Oracle ne è il locatario. Sorge a Santa Teresa, nella contea di Doña Ana, vicino a El Paso e al confine messicano; Oracle ha un contratto di affitto di 18 anni e il cliente principale è OpenAI. Per dare un'idea della scala, con 2,5 gigawatt potrebbe alimentare più di metà del New Mexico. Secondo l'FT l'entrata in funzione è prevista per il 2028.

Il nodo: l'energia. Il campus non si allaccerà alla rete pubblica ma avrà una propria micro-rete: Oracle e il partner BorderPlex Digital Assets hanno annunciato che la fonte principale saranno le celle a combustibile di Bloom Energy. Le celle però vanno a gas, e qui nascono i problemi. La micro-rete dipenderebbe comunque da un gasdotto, che la commissaria statale per i terreni pubblici ha bocciato due volte per il tratto su terreni dello Stato, e i cui lavori risultano rallentati almeno all'inizio del prossimo anno.

Le cause legali. La micro-rete ha bisogno di un permesso sulla qualità dell'aria, e lì si è aperto il contenzioso:

- Con due ordinanze, la Corte suprema del New Mexico ha sospeso l'udienza sul permesso e ha ordinato all'ufficio statale competente di bloccare l'uso dell'acqua di un pozzo specifico per il cantiere. Il ricorso era di un gruppo ambientalista, secondo cui l'udienza era prematura perché il gasdotto non ha ancora le autorizzazioni.
- In seguito la Corte ha respinto all'unanimità la richiesta degli sviluppatori di riprendere almeno in parte il procedimento.
- Ancora la settimana scorsa, le autorità e lo sviluppatore hanno chiesto altro tempo per esaminare modifiche al permesso, con un nuovo rinvio dell'udienza. Secondo un titolo della stampa locale, lo sviluppatore ha anche rinunciato alla scadenza del 23 novembre per la decisione sul permesso; non ho potuto leggere l'articolo completo.

Sullo sfondo c'è anche la questione dell'acqua in una zona desertica. Oracle sostiene che raffreddamento e celle a combustibile insieme consumeranno ogni anno, in media su 15 anni, circa quanto nove famiglie americane, ma i critici contestano le stime.

Intanto il cantiere va avanti. Secondo Oracle, quasi 700 residenti lavorano già sul sito, e la società dichiara di aver assegnato oltre 994 milioni di dollari di contratti a imprese del New Mexico.

Il caso tocca tre temi:

- **Volatilità del calcolo.** Se un campus da oltre 2 gigawatt slitta, una grande quantità di capacità prevista non arriva quando atteso. Questo è uno dei fattori che rendono il prezzo dell'ora-GPU difficile da prevedere, e quindi utile da coprire con un future.
- **Rischio per i finanziatori.** Gli edifici si costruiscono mentre permessi essenziali (energia, acqua) sono ancora in discussione. Il denaro speso è difficile da recuperare, e un ritardo dei permessi rinvia i ricavi che dovrebbero rimborsare il debito.
- **Rischio politico.** Il caso Jupiter è un esempio concreto dell'opposizione locale ai data center che l'FT documenta con le 38 misure referendarie approvate alle elezioni di metà mandato. È anche il tipo di rischio per cui le banche apprezzano un'assicurazione del credito che paga indipendentemente dalla causa dell'insolvenza.

Non ho trovato informazioni pubbliche su quali banche finanzino specificamente Project Jupiter, quindi non è possibile collegarlo direttamente a Barclays, HSBC o JPMorgan.

Scheda 5. Il caso Jupiter come esempio di avversione ai grandi datacenter causa probabile dell'esito delle elezioni di metà mandato americane.

4 I finanziamenti all'AI delle banche d'investimento

4.1 Il fatto: le banche hanno raggiunto i limiti

Le banche stanno ricorrendo sempre più spesso all'assicurazione del credito per trasferire parte del rischio dei prestiti ai data center, così da aggirare i limiti interni di esposizione verso singole società, progetti o settori e continuare a prestare. Barclays ha recentemente acquistato una polizza per coprire la propria esposizione alle infrastrutture AI; HSBC ha valutato lo strumento; JPMorgan, la maggiore banca statunitense, lo ha considerato nelle ultime settimane ma ha deciso di non acquistarlo.

Il punto di partenza è dunque non la mancanza di capitale, ma la concentrazione. Le banche hanno già prestato molto all'AI e i comitati rischi impediscono di aggiungere esposizione verso gli stessi nomi; l'assicurazione serve a "liberare spazio" sotto quei limiti.

4.2 Come funziona lo strumento

In un'operazione tipica, una banca che vuole prestare alcune centinaia di milioni a un singolo progetto trasferisce agli assicuratori tra il 20% e il 50% dell'esposizione. Le banche europee e canadesi usano da tempo l'assicurazione del credito anche per ridurre il capitale regolamentare richiesto; le banche statunitensi in genere non ottengono questo beneficio e hanno usato i credit default swap. L'aumento del costo dei CDS e la mole di debito AI emesso spingono ora anche diverse banche statunitensi a valutare le polizze, che pur senza sconto di capitale consentono di rispettare i limiti interni. La polizza paga in caso di insolvenza indipendentemente dalla causa: la banca non deve dimostrare, per esempio, che il fallimento è dovuto all'opposizione politica al data center.

Il mercato si è allargato: gli assicuratori del credito per il project finance, tra cui AIG, Chubb e Axa, sono oltre 70, contro pochi un decennio fa.

4.3 Che cosa dice il prezzo della protezione

Il dato più eloquente è il costo. Gli assicuratori chiedono premi annuali pari al 50–70% di quanto la banca guadagna sulla quota assicurata del prestito, e questi prezzi hanno resistito al calo generale delle tariffe dell'assicurazione del credito. In termini illustrativi: su un prestito di 500 milioni di dollari di cui il 40% (200 milioni) è assicurato, con un margine ipotetico di 200 punti base la banca incasserebbe 4 milioni l'anno su quella quota e ne pagherebbe tra 2 e 2,8 di premio.

Se una banca accetta di cedere gran parte del margine, il motivo non è il rendimento del prestito in sé: le banche pagano per poter erogare prestiti più grandi a un singolo data center, ottenere lo status di prestatore principale o dimostrare sostegno a un cliente. In altre parole, il prestito è sempre più uno strumento di relazione che apre la porta ad attività più redditizie: consulenza, emissioni obbligazionarie, collocamenti azionari. Dall'altro lato, il fatto che gli assicuratori pretendano un premio così alto e possano essere selettivi indica che percepiscono il rischio come elevato.

Banca	Posizione sull'assicurazione del credito per prestiti AI
Barclays	Ha acquistato una copertura sull'esposizione ai prestiti per infrastrutture AI
HSBC	Ha esplorato l'uso dello strumento
JPMorgan Chase	Lo ha considerato nelle ultime settimane, poi ha deciso di non acquistarlo
Banche USA in generale	Valutano le polizze per il costo crescente dei CDS, pur senza beneficio di capitale
Banche europee e canadesi	Le usano da tempo anche per ridurre il capitale richiesto

Tabella 4. Il ricorso all'assicurazione del credito. Fonte: FT, 11 novembre 2026.

4.4 Valutazione complessiva

Il quadro è quello di un ciclo di credito maturo: le banche hanno esposizioni concentrate, cercano di distribuirle ad altri soggetti (assicuratori, investitori in CDS) e accettano margini bassi per restare nel gioco. Il rischio non scompare dal sistema, si sposta verso il settore assicurativo. Due fattori rendono il momento delicato: il crollo dei titoli tecnologici dopo le elezioni di metà mandato, quantificato in 3.200 miliardi di dollari e poi rientrato, ha mostrato quanto rapidamente il mercato possa girare, con la leva che amplifica i movimenti; e i crescenti dubbi sulla sostenibilità del boom citati a proposito dell'IPO di Anthropic.

5 Il collegamento con l'IPO di Anthropic

5.1 I termini dell'operazione

Elemento	Dato
Forchetta di prezzo	70–78 $ per azione (deposito di martedì 10 novembre)
Valutazione implicita	1.260–1.410 miliardi di $ (pienamente diluita); obiettivo 1.400 mld, possibile aumento a 1.500 mld o oltre
Raccolta in IPO	Fino a 64 mld $, più circa 10 mld di greenshoe in caso di eccesso di domanda
Collocamento privato	26 mld $ da Google, Nvidia e Microsoft al prezzo di IPO
Raccolta totale potenziale	Oltre 100 mld $ con il rialzo del prezzo
Banche capofila	Goldman Sachs e Morgan Stanley
Ricavi	Ricavi annualizzati 81 mld $ a ottobre; 17 mld nel terzo trimestre
Spesa pianificata	Circa 900 mld $ in calcolo e infrastrutture nei prossimi anni
Valutazione Morningstar	1.700 mld $ come valore equo, in un intervallo 500–3.000 mld
Quotazione attesa	La prossima settimana

Tabella 5. I numeri dell'IPO. Fonte: FT, 11 novembre 2026.

Alcuni rapporti aiutano a leggere i numeri. A 1.400 miliardi la valutazione è circa 17 volte i ricavi annualizzati di ottobre (1.400 / 81); a 1.500 miliardi circa 18,5 volte. Lo strumento di calcolo del FT arriva a 1.400 miliardi ipotizzando ricavi di 270 miliardi nel 2028 e un multiplo di 5,2 volte, inferiore a quello attuale di Microsoft. Le aspettative dei sostenitori avevano puntato a 2.000 miliardi, poi ridimensionate. Intanto alcuni investitori con un'assegnazione la stanno già rivendendo a prezzi superiori a quello di IPO tramite veicoli dedicati, una pratica che la società ha cercato di contrastare.

5.2 Il collegamento esiste, ed è sostanziale

Esistono cinque nessi concreti tra IPO di Anthropic ed i finanziamenti ai data center dedicati al calcolo per l'IA.

a) **Equity dove i bilanci bancari non arrivano più.** Tre finanziatori di Anthropic riferiscono che la società punta a raccogliere il più possibile per sostenere la spesa in calcolo, più che a massimizzare la valutazione. È coerente con il quadro della Sezione 4: se le banche sono vicine ai limiti di concentrazione e devono pagare cara la protezione, il capitale azionario diventa la principale fonte alternativa. Le proporzioni chiariscono anche i limiti di questa fonte: oltre 100 miliardi rappresentano circa l'11% dei 900 miliardi di spesa pianificata. Il resto dovrà arrivare da flussi di cassa, partner e, inevitabilmente, ancora da debito, che sia della società o dei data center che lavorano per essa.

b) **Il debitore finale dei data center.** Anthropic è uno dei pochi laboratori che generano gran parte della domanda di calcolo. I prestiti che le banche erogano ai data center sono rimborsati con i canoni pagati da clienti come Anthropic. Un'IPO da oltre 100 miliardi rafforza il patrimonio di uno dei principali clienti finali e quindi, indirettamente, la qualità del credito di tutta la catena. Specularmente, una delusione sul titolo dopo la quotazione verrebbe letta come un segnale sulla tenuta dei canoni.

c) **Il ruolo doppio delle banche d'investimento.** Goldman Sachs e Morgan Stanley guidano il collocamento. Questo illustra il motivo per cui le banche accettano margini sottili sui prestiti ai data center: l'ecosistema AI

genera commissioni di collocamento e consulenza molto più ricche, e la presenza come finanziatore è parte della relazione commerciale.

d) **Fornitori che finanziano i clienti.** Nel collocamento privato investono Nvidia, Google e Microsoft, cioè il produttore dominante di chip e due dei principali fornitori di capacità. È il fenomeno della circolarità: chi vende il calcolo investe in chi lo compra. È la stessa concentrazione che preoccupa la CFTC per i futures e costituisce il punto debole del mercato.

e) **Il prezzo del calcolo come variabile di valutazione.** Il calcolo è la principale voce di costo di un laboratorio di AI. Il prezzo dell'ora-GPU incide quindi sui margini di Anthropic, sul valore delle infrastrutture e sulla solvibilità dei finanziatori. Un laboratorio è un potenziale acquirente di copertura. Per gli investitori azionari, una curva futures pubblica diventerebbe un indicatore in tempo reale dei costi attesi, con l'effetto "segnale" già descritto.

5.3 Elementi di prudenza

Il contesto è fragile: la quotazione arriva pochi giorni dopo un ribasso di 3.200 miliardi dei titoli tecnologici, e un grande investitore ha osservato che in questo mercato un fine settimana è un'eternità. La forchetta di Morningstar, da 500 a 3.000 miliardi, esprime l'enorme incertezza sulla velocità di adozione degli strumenti AI. Il confronto con OpenAI (10 miliardi di ricavi nel terzo trimestre, +50%) mostra una crescita rapida in tutto il settore, ma la sostenibilità del rapporto tra ricavi e investimenti resta il nodo che il mercato dovrà valutare.

6 Il possibile impatto del quantum computing sull'AI

6.1 Il fatto

Secondo l'Agence France-Presse (AFP), la start-up australiana Diraq sta conducendo il primo test sul campo al mondo di un piccolo computer quantistico basato su chip di silicio all'interno di un data center commerciale, quello di Equinix a Sydney. Il fondatore Andrew Dzurak sostiene che il quantum potrà ridurre l'enorme fabbisogno energetico dei data center AI e sarà pronto a farlo entro cinque anni, perché una parte dei problemi affrontati dall'AI sarebbe più adatta al calcolo quantistico.

Tappa	Obiettivo Diraq
Oggi (novembre 2026)	8 qubit su chip di silicio, consumo inferiore a 20 kW
2029	150.000 qubit
2031	2 milioni di qubit, soglia DARPA raggiunta con due anni di anticipo; consumo di alcune centinaia di kW
2033	Termine DARPA per dimostrare un computer quantistico utile a costi sostenibili

Tabella 6. La tabella di marcia di Diraq. Fonte: AFP, 10 novembre 2026.

Il contesto australiano è vivace: Diraq e PsiQuantum (sostenuta da un investimento pubblico di 1 miliardo di dollari australiani) sono tra le sei società nella terza fase del programma DARPA, Silicon Quantum Computing è stata premiata da Time per un chip AI potenziato dal quantum, e il paese conta oltre 50 imprese del settore. Il limite riconosciuto è la commercializzazione: lo stesso Dzurak precisa che tutto dipende dalla capacità di costruire macchine a costi competitivi.

6.2 Implicazioni

Energia. Il confronto è netto: alcune centinaia di kilowatt per un sistema quantistico avanzato contro data center da gigawatt. Se anche solo una parte dei carichi di lavoro migrasse, si allenterebbe il vincolo energetico che oggi genera nuove centrali, pressione sulle reti e l'opposizione locale documentata dalle 38 misure referendarie.

Domanda e prezzo dell'ora-GPU. Un sostituto tecnologico credibile ridurrebbe la domanda futura di GPU e quindi il prezzo atteso del noleggio. Questo riguarda direttamente la parte lunga della curva futures, che arriva a 36 mesi, cioè fino al 2029, proprio l'anno in cui Diraq punta a 150.000 qubit.

Valore residuo delle garanzie bancarie. Per i creditori il quantum è un rischio di obsolescenza che si somma a quello interno al mondo GPU (una H100 viene superata da una B200). Le scadenze dei finanziamenti ai data center e il piano di spesa di Anthropic si estendono oltre il 2030, quindi l'orizzonte 2029–2033 rientra nella vita dei prestiti di oggi.

Nuove applicazioni e possibile nuova commodity. Diraq indica usi in chimica, farmaci e batterie, stimando che nella sperimentazione dei farmaci il quantum possa aumentare del 10% il tasso di successo e ridurre di un terzo il costo di sviluppo. In uno scenario di successo, anche le ore di calcolo quantistico potrebbero un giorno diventare un sottostante con un proprio indice.

6.3 Una valutazione prudente

Le previsioni provengono dal fondatore dell'azienda, che ha un interesse evidente a presentarle in modo ottimistico, e non dimostrano ancora un vantaggio quantistico sui carichi tipici dell'AI. L'addestramento e l'uso dei

grandi modelli linguistici si basano soprattutto su moltiplicazioni di matrici, per cui le GPU sono estremamente efficienti; il consenso prevalente vede il quantum come acceleratore complementare per problemi specifici (simulazione molecolare, ottimizzazione), installato accanto ai sistemi classici, come nel test di Sydney, più che come sostituto generale. Lo scenario più probabile nei prossimi anni è quindi ibrido: un impatto limitato sulla domanda di GPU fino alla fine del decennio, ma un fattore di rischio che chi presta a lungo termine sulle GPU e chi valuta la curva dei futures dovrebbe iniziare a considerare negli scenari avversi.

7 Conclusioni

Il noleggio di potenza di calcolo soddisfa bene due delle tre condizioni per diventare una commodity con un mercato futures (volatilità e presenza di acquirenti naturali della copertura) e solo parzialmente la terza, la standardizzazione, che l'indice Silicon Data affronta con una metodologia dettagliata: conversione in prezzo per ora-GPU, aggiustamento econometrico verso una configurazione di riferimento, aggregazione a due livelli con tetti di concentrazione e una governance ispirata ai principi IOSCO. Il rinvio della CFTC dimostra che il punto debole resta la concentrazione e l'opacità del mercato sottostante.

Per le banche, il future sarebbe un utile strumento di valutazione e copertura, ma anche un segnale pubblico potenzialmente pro-ciclico. Arriva in un momento in cui il sistema bancario ha già raggiunto i propri limiti di esposizione creditizia verso l'AI e sposta rischio verso gli assicuratori a un prezzo che assorbe gran parte del margine. L'IPO di Anthropic si inserisce in questa stessa dinamica: è il modo in cui il settore attinge al capitale di rischio quando il credito è saturo, ma anche così copre solo una frazione della spesa programmata, e il suo esito sarà un indicatore della fiducia del mercato in tutta la catena di finanziamento. Il quantum, infine, è oggi un fattore di medio/lungo periodo con molte incertezze, ma tocca proprio le variabili che determinano il valore della catena (energia, domanda di GPU, valore residuo delle garanzie) ed entra nell'orizzonte temporale dei finanziamenti decisi oggi.

Indicatori da seguire nei prossimi mesi: la decisione della CFTC e la liquidità iniziale dei contratti CME e ICE; la convergenza o divergenza tra gli indici Silicon Data e Ornn; il prezzo finale e l'andamento post-quotazione di Anthropic; il costo dei CDS e dei premi assicurativi sui prestiti ai data center; i risultati intermedi del programma DARPA sul quantum.

Fonti

1. Financial Times, L. Harris e O. Aliaj, “Banks turn to credit insurance as they hit limits on lending to AI boom”, 11 novembre 2026.
2. Financial Times, M. Rubinstein, “The coming futures market in AI compute”, 3 ottobre 2026.
3. Financial Times, M. Chan e J. R. Shah, “US regulator postpones Wall Street push to trade AI computing futures”, 10 novembre 2026.
4. Financial Times, G. Hammond, A. Pollard e G. Steer, “Anthropic bankers prepare to increase IPO valuation to $1.5tn or more”, 11 novembre 2026.
5. AFP, “Quantum could solve AI energy drain, says tech founder”, 10 novembre 2026.
6. Silicon Data, “GPU Rental Index Series: Governance, Methodology & Control Framework”, silicondata.com.
7. Silicon Data, comunicato di lancio dell’indice SDH100RT, maggio 2025; pagine prodotto Silicon Index.
8. CME Group, Special Executive Report SER-9785, “Initial Listing of Two Compute Futures Contracts”, 11 agosto 2026.
9. Silicon Data, serie storica giornaliera del Silicon Data H100 Rental Index (SDH100RT), 1° settembre 2024 – 3 novembre 2026.

Il presente documento ha finalità informative e di ricerca; non costituisce consulenza finanziaria, raccomandazione di investimento né sollecitazione all’acquisto o alla vendita di strumenti finanziari. Le opinioni espresse sono quelle dell’autore a novembre 2026 e sono soggette a modifica senza preavviso.