Everything You Wanted to Know About Compute Markets but Were Afraid to Ask

Why fintech innovation will shape the next phase of AI competition

Hivemind

Contents

Introduction

The US government has set a mandate to accelerate the maturation of a healthy financial market for compute. Let's drill into why fintech innovation will shape the next phase of AI competition.

The sums being committed to AI infrastructure are already staggering. Last year, Microsoft announced plans to spend approximately $80 billion on AI-enabled data centers. Meta increased its capital expenditure forecast to between $64 billion and $72 billion, driven largely by AI infrastructure investments. Collectively, Alphabet, Amazon, Meta, and Microsoft are projected to spend hundreds of billions of dollars on infrastructure in the coming years.

Besides the sheer balance sheet requirements, complex supply and demand dynamics have also emerged in the GenAI era and they have rapidly become top of mind. For many years, cloud computing created the perception that organizations could access as much computing power as they needed, when and where they needed. Capacity could be provisioned quickly and expanded alongside demand. With generative AI, any organization hoping to develop or deploy advanced AI systems depends on infrastructure that requires substantial investment long before commercial returns can be realized. In many cases, access to land, power, water and computing capacity is becoming just as important as advances in the models themselves. This means that organizations need adequate tools not just to finance, but also to plan, to hedge, and to speculate.

In other words, compute infrastructure has become more than a competitive advantage. It is reaching the stage of a critical commodity. While capital has so far scaled alongside the ever-growing demand for this infrastructure – and while there is still plenty of capital available to keep doing so – the conduits through which that capital is deployed need to become more standardized and sophisticated to ensure the trifecta of scalability, efficiency, and resilience is met.

This is why turning compute into a market is the defining financial engineering challenge of this decade.

The transition from bilateral cloud SLAs to exchange-traded capacity and standardized risk transfer will unlock hundreds of billions in trapped capital and dramatically accelerate the deployment of AI infrastructure.

How do we exactly do that?

Let's first go through how compute works today and how spending stacks up, to get a better sense of what priorities and opportunities look like within compute markets. We then look at those through the lens of what other commodity markets have done, notably surfacing any structural, technical, and regulatory hurdles. Finally, we derive a likely path forward for this new commodity market.

~$80B	$64–72B	Hundreds of billions
Microsoft, announced plans on AI-enabled data centers.	Meta, capital expenditure forecast, driven largely by AI infrastructure.	Alphabet, Amazon, Meta and Microsoft together, on infrastructure in the coming years.

1. The Make Up of Compute

We think it is most intuitive to decompose Compute into 4 layers:

Cloud & Middleware Fabric

All the software functions that run on racks, from low-level back-end to high-value API endpoints, from internal facing to external facing.

Specialized-purpose Data Ops	General-purpose Data Ops	Orchestration
OpenAI, Anthropic, Hugging Face, CoreWeave, AWS Bedrock	AWS, Snowflake, MongoDB, Oracle	Docker, Kubernetes, HashiCorp, Vercel

Datacenter Fabric

All components going into racks and across racks

Compute	Memory & Storage	Orchestration & Comms
NVIDIA, AMD, Google (TPU), AWS Graviton/ Trainium	Samsung, Micron, SK Hynix, Western Digital, Seagate	Supermicro, Foxconn, Quanta, Asus

Physical Site Fabric

All components interfacing the racks with the outside world

Power Delivery & Management	Power & Energy	Cooling
Schneider Electric, Eaton, Vertiv, Cummins	Regional utilities	Vertiv, Supermicro, Munters, Municipal Water

Raw Materials Fabric

Where everything starts, the bare bare-metal(s)

Copper, Lithium, Cobalt, Nickel, Silicon - rare earths: neodymium, praseodymium, dysprosium, terbium

For the purpose of our research, we focus on the subset of components across each of these fabrics that we think are front of mind for investors, operators, and governments. For each subset, we provide a qualitative evaluation across 4 salient criteria for prioritizing compute financialization: their cost contribution, their price stability, their supply-chain stability, and their substitution ability

LAYER	FUNCTION	COMPONENTS	COST CONTRIBUTION	PRICE STABILITY	SUPPLY-CHAIN STABILITY	SUBSTITUTION ABILITY
Cloud & Middleware	Specialized Purpose Data Ops	Training	++	~	~	--
		Inference, Tokens	++	-	-	~
	General Purpose Data Ops	**[NO CONCERN]** Storage buckets, Reserved & spot instances, databases, desktop services				
	Orchestration	**[NO CONCERN]** Containers, lambda scaling				
Datacenter	Compute	GPU, TPU, ASIC	+++	--	--	--
		CPU, FPGA	+	-	~	-
	Memory & Storage	HDD	~	~	~	+
		RAM, SSD	++	--	--	-
	Orchestration & Communication	**[NO CONCERN]** Motherboards, IO bus (e.g. Supermicro, Foxconn, Quanta, Asus)				
Physical Site	Power Delivery & Mgmt	Generators, Transformers, Power Managers, PDU, UPS	+	~	-	+
	Power & Energy	Electricity	+	+	~	--
	Cooling	Water	~	+	-	--
		Liquid Cooling loops, Air HVAC	–	+	~	++
Raw Materials	Electronics	Copper, Lithium, Cobalt, Nickel, Silicon	++	+	--	--

How enterprises produce and consume compute resources across each of these fabrics and across each of their underlying functions has drastically evolved in the last 4 years with the spread of Generative AI. It is thus worth understanding a few intertwined dynamics of that evolution to better appreciate the most salient entry and expansion points from a financialization perspective.

Historically, all is well in compute land from the Datacenter Fabric layer and up, and all is ok for the layers below that. Compute cards, racks, even datacenters are not that expensive, and supply-chain runs fairly smoothly outside of periodical hiccups in Raw Materials.

Accelerated computing and very-large-scale workloads fundamentally altered the math, resulting in:

- Revenue generated on top of these datacenters having scaled by **~4x**, while service providers operating on top of those datacenters as well as their clients regularly face capacity constraints (unmet demand).

- Total initial costs going up **~3-8x** with GPU-heavy servers now accounting for **~60%** of this initial cost.

- An explosion of (mainly off-balance sheet) private-credit financings to support those much larger, much more expensive build-outs.

- Multi-year backlogs across several components, as production lines were not built to sustain the current demand.

- Regular disruptions of the supply chain, as a lot of it was not built to sustain the current demand.

There are consequences worth highlighting for the make-up of compute. All these high-performance chips require a lot more power, and that increase in power consumption and distribution requires more than just equipment:

- It requires larger versions of pre-existing equipment such as transformers and uninterrupted power supplies (UPS). These larger versions not only take longer to produce, but they also consume more critical minerals (copper, nickel, lithium, cobalt, silicon) and rare earths (neodymium, praseodymium, dysprosium, terbium), whose supply chains are bottlenecked and vulnerable to disruption, with China as the dominant supplier.

- Moreover, those racks run very hot (power of 50kW to 100kW+ per rack implies temperatures of 85C+), so advanced cooling is no longer optional. Implementing Direct Liquid Cooling (DLC) or immersion cooling adds significant capital. expenditure ($15,000 to $50,000 per rack), pushing the combined cost of physical site infrastructure (Power Delivery + Cooling) to roughly 20-25% of the total build.

- It will soon require an extensive overhaul of Power & Energy Components to accommodate much higher voltage and amperage, a push by hyperscalers that again has deeper implications than swapping a part for a larger version of it.

Thus, financialization of compute has to do not just with managing financing, rising and fluctuating costs, but also with managing production delays and fluctuations in capacity.

Compute now has all the makings of a commodity: dependable, for which we need scalable, repeatable and efficient means of financing, planning, hedging, and hence speculating too.

2. How Have Other Commodity Markets Done It?

a. Defining and Using Standards

Liquidity is the lifeblood of any financial market. In commodity markets, that liquidity is impossible to scale without standardization. Standardization creates fungibility, meaning one contract is interchangeable with another and one contract has some form of continuity over time, allowing financial participants not only to trade without needing to physically inspect the underlying goods, but also to implement scalable and repeatable financial processes around their planning, hedging, speculation, and financing.

Defining these standards requires intense collaboration between industry consortiums, producers, consumers, and exchange committees (such as the CME or LME). They must agree on exact chemical, physical, and logistical parameters. For example:

> **Crude Oil:** The NYMEX WTI standard mandates specific API gravity (between 37° and 42°) and a maximum sulfur content of 0.42%.

> **Gold:** The COMEX standard requires gold to be cast in bars weighing either 100 troy ounces or 1 kilo, with a minimum purity of 99.5% (fineness of 995.0).

Agreeing on these standards often involves years of negotiation to ensure the specification is broad enough to guarantee sufficient physical supply, yet narrow enough to be an effective hedging tool. Indeed, a standard does not need to be perfect to be useful. To be viable as a multi-billion-dollar market, a hedging instrument needs to remove enough risk at scale to be worth its cost. A 70% correlation between a standard and a portfolio can still reduce that portfolio's variance in half.

This characteristic of commodity standards is critical to adapting what we have already done in traditional commodities to compute markets. Just as airlines hedge jet fuel using crude oil futures (despite chemical differences), or bond desks hedge rate risk in credit with treasuries, or long-short equity portfolios hedge market beta with S&P 500 futures, imperfect cross-hedges predominate in the real US economy, and are very useful, even transformative while still being imperfect. AI labs can successfully hedge their hardware costs using a standardized public GPU index, provided the correlation remains tight.

b. Using Basis to Deal with Deviations from Standards

Because physical commodities are natural products, they rarely perfectly match the standard, and AI commodities are no different. An "H100 hour" could represent many different goods: SXM or PCIe, spot or reserved, hyperscaler or neocloud, US or Asia.

To deal with these deviations in a scalable and repeatable manner, markets use "Basis"—a pre-agreed system of premiums and discounts adjusting the contract price based on quality or location differences.

Specific Examples of Basis:

Quality Basis: If a producer delivers crude oil with a slightly higher sulfur content (sour crude) against a WTI contract, a pre-agreed discount is automatically applied to their payout.

Location Basis: The Brent-WTI spread is a classic location basis, representing the price difference between seaborne crude in the North Sea and landlocked crude in Cushing, Oklahoma. Another example is the Waha Hub vs. Henry Hub natural gas basis swap, which prices the localized oversupply in the Permian basin against the national benchmark.

c. Enabling and Ensuring the Delivery of Commodities

While only 1-2% of commodity derivatives result in physical delivery, this mechanism is the vital anchor tethering the "paper" price to physical reality.

For WTI crude for instance, delivery revolves around Cushing, Oklahoma - a hub with ~90 million barrels of storage capacity. Upon contract expiration, the exchange matches open short and long positions. The seller transfers electronic pipeline receipts or storage certificates representing exactly 1,000 barrels per contract in an approved Cushing facility to the buyer. The buyer remits the multimillion-dollar cash equivalent. This seamless transfer of custody within a highly networked infrastructure ensures the financial derivative remains a true proxy for the physical barrel.

Beyond the commodity itself, there is a massive, opaque derivative market for the logistics of delivery of Basis Swaps and Financial Transmission Rights (FTRs in power markets) to hedge the price differential between standards and operating reality:

When physical supply outpaces regional transport infrastructure, pipeline space becomes the actual scarce asset.

When physical supply deviates from the standard but remains useful and needed, participants still need a common reference, and a repeatable medium to adjust deviations from it.

For instance, the cost to move gas from the oversupplied Permian Basin to the high-demand Gulf Coast represents tens of billions of dollars in notional volume annually. In major US power markets (PJM, ERCOT), the market for congestion revenue rights handles over $2-$3 billion in annual revenues. Hundreds of specialized energy traders, utilities, and infrastructure funds participate daily.

Understanding these logistical derivatives is foundational. As we transition to digital infrastructure, the framework of trading "capacity" and "transit rights" over physical constrained networks (pipelines) maps perfectly onto the emerging need for derivatives pricing compute power access and data transmission bandwidth.

d. What Are the Different Types of Market Participants?

The CFTC's Commitments of Traders (COT) report outlines these groups:

40-45%
OF OPEN INTEREST

Commercial Hedgers (Producers/Consumers = CH)
Such as oil majors, airlines, food conglomerates, miners, construction conglomerates need to finance, plan and hedge. They hold roughly 40-45% of Open Interest (OI). They are structurally short (producers locking in prices) or long (consumers securing costs). For compute markets, these are the chip and memory designers, the hyperscalers, the neoclouds and everyone running services on top of datacenters

30-35%
OF OPEN INTEREST

Managed Money (Speculators/Hedge Funds = MM)
Need to speculate and hedge. They hold 30-35% of OI. They provide the necessary liquidity for hedgers to offload risk. They dominate trend-following strategies across energy and metals. The same firms are relevant to compute markets

~15-20%
OF OPEN INTEREST

Swap Dealers (Banks = SD)
They account for ~15-20% of OI. They act as intermediaries, packaging bespoke OTC trades for clients and offsetting that risk on public exchanges. The same firms are relevant to compute markets

e. Getting Exposure through Spot Markets and Derivatives

INSTRUMENT	FUNCTION & USERS	VOLUME & MARKET SHARE	EXECUTION (EXCHANGE VS. OTC)	CLEARING (CENTRAL VS. BILATERAL)
Spot Market	Planning (CH): secure immediate physical inventory	Fraction of paper markets (derivatives outstrip physical by 20:1 to 30:1)	Entirely OTC	Bilateral (direct counterparty risk)
Futures	Used to lock in prices (CH) or bet on trends (MM) and hedge other speculative positions (MM)	Dominates with 65–70% of derivative volume	Exchange-traded (standardized contracts with public order books)	Centrally cleared
Options	Used to insure against price swings (CH) or bet on volatility (MM) and hedge other speculative positions (MM)	~20–25% of volume; heavily utilized in agricultural and energy tail-risk hedging	Primarily Exchange-traded, with some bespoke OTC activity	Centrally cleared for exchange products; some bilateral for OTC
NDFs & Swaps	Cash-settled forwards to transfer financial risk where physical delivery is impossible (MM, SD)	5–10% of volume; concentrated in specialized energy logistics and base metals	OTC (custom dates, exotic grades, bespoke locations)	~80% centrally cleared post-Dodd Frank; remainder is bilateral
Depositary Receipts	Bearer documents representing legal ownership of physical commodities stored in approved warehouses (CH)	Low turnover compared to derivatives; tied directly to actual physical inventory levels	Exchange-registered (e.g., LME warrants) or OTC	Centrally tracked by exchange depositories or held bilaterally
Exchange for Physical (EFP)	Off-market agreement to swap a physical commodity position for a futures contract position (CH)	Niche but critical subset of futures volume; heavily used by major physical merchants	Negotiated OTC (bilaterally) but immediately reported to the Exchange	Futures leg is centrally cleared; physical leg is settled bilaterally
Basis Swaps & Financial Transmission Rights (FTR)	Hedges against spatial price differences (congestion costs) between two locations on an electricity grid (CH, MM)	Highly concentrated and strictly limited to regional power markets (e.g., PJM, ERCOT)	Acquired via ISO/RTO grid operator auctions or traded OTC	Centrally cleared by the regional Independent System Operator (ISO)

Taking a top-down view, exchanges command ~60% of the market share by notional value (over 8 billion contracts annually), while OTC markets hold ~40%, favored by physical merchants (Trafigura, Vitol) for highly specific logistics, like airline jet fuel delivery to specific airports. Following the 2008 Dodd-Frank regulations, ~80% of all commodity derivatives are now centrally cleared. The remaining ~20% of bilateral trades are largely reserved for highly capitalized entities executing illiquid deals (e.g., 10-year off-take agreements for new mines, or Nvidia hardware-swap agreements with major partners).

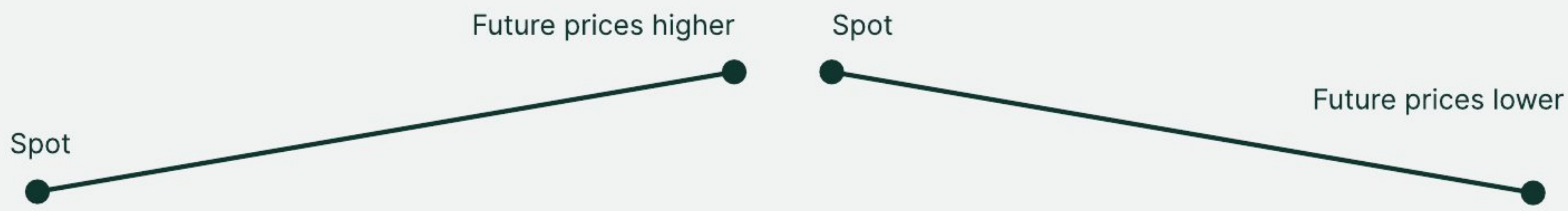
~60% / ~40%
Exchange against OTC market share by notional value.

8 billion+
Contracts annually across exchanges.

~80%
Of all commodity derivatives are now centrally cleared, following Dodd-Frank.

f. Term Structure: Why Do Contango and Backwardation Matter?

The term structure represents prices across maturity dates. With commodities, there are two noteworthy and generally transient patterns worth emphasizing, as they will quite key in how to think about bootstrapping and scaling liquidity in compute markets.



Contango (Upward Sloping)

Future prices are higher than spot prices. This reflects the "cost of carry" (storage, insurance, financing). It encourages liquidity providers and physical arbitrageurs to buy spot, store it, and sell futures risk-free. For example, gold is almost perpetually in contango due to storage/interest costs and lack of immediate consumption utility.

Backwardation (Downward Sloping)

Spot prices exceed future prices, reflecting a "convenience yield"—a premium paid for immediate access due to scarcity. It incentivizes market participants to aggressively draw down inventories and sell into the spot market, effectively rebalancing physical shortages. For example, crude oil frequently falls into steep backwardation during supply shocks or post-recession demand surges (e.g., post-COVID recovery).

Equipped with all this background knowledge on commodity markets, let's now turn our attention to how they can be adapted to compute without requiring reinventing the wheel.

3. Turning Compute into a Market

Compute markets are not “that” new. Certain adaptations from traditional commodities are required, but a lot of existing structures and frameworks can be reused. That is not to say that barriers to entry are low, rather to say that this is a product and distribution challenge, not a research & technology breakthrough one.

a. Genesis of Compute Markets: The Cloud "OTC" Era

Compute markets are not a new phenomenon; they already exist in the form of massive, bilateral Over-The-Counter (OTC) markets hosted by the dominant hyperscalers. The purchasing mechanisms offered by Amazon Web Services (AWS), Microsoft Azure, and Google Cloud Platform (GCP) are, in economic terms, primitive financial instruments:

- **Spot aka Pre-Emptible Instances:** These represent the Spot Market. Purchasers bid for excess, interruptible capacity at a dynamic clearing price for immediate consumption.
- **Reserved Instances / Committed Use Discounts:** These function identically to Forward Contracts. A buyer commits to a specific volume of compute over a 1-to-3-year term at a fixed, typically discounted rate but not always discounted as there are higher guarantees on availability and stability than spot.
- **Residual / On-Demand Instances:** These act as Basis Risk Instruments, bridging the gap between committed baseloads and unexpected peak or drop in demand, carrying premiums for high flexibility (on-demand) or discount for reduced flexibility (residual).

To understand the scale of this "OTC" market, we can look at the latest Q1 2026 run rates for the Big Three. While hyperscalers do not break out exact billing types, industry analysts estimate that approximately 40-50% of infrastructure revenue is tied to long-term commitments (Forwards), while 5-10% is transacted via interruptible instances (Spot).

Estimated 2026 Global Cloud Market Sizing by Instrument Proxy

PROVIDER	2026 EST. ANNUALIZED RUN RATE	EST. COMMITTED (FORWARDS)	EST. INTERRUPTIBLE (SPOT)
AWS	$150B	~$65B	~$10B
Microsoft Azure	$139B	~$60B	~$9B
GCP	$80B	~$35B	~$5B

Source Data: Q1 2026 Earnings / Synergy Research Group (Run rates extrapolated. Commitment/Spot allocations are industry estimates).

b. Exchange-Based Markets Complement Bilateral OTC Markets

What people are really discussing when they talk about compute markets nowadays is the expansion of this $500B+ market from a (somewhat) narrow, (triopolies) concentrated, and opaque bilateral structure to an expansive, diversified, exchange-traded ecosystem upstream of it, concurrent to it, and downstream from it.

This expansion is inevitable now that compute, particularly AI compute, has become a strategic and critical resource.

Capital Efficiency

Currently, AI labs and enterprises must hoard GPUs or sign multi-billion-dollar, multi-year cloud commitments to guarantee capacity. A liquid market allows participants to secure forward capacity using margin, freeing up working capital.

Price Discovery

Bilateral NDAs obscure the true market-clearing price of a GPU hour. An exchange provides a transparent, standardized tradable ticker for each relevant component of compute.

Resilience and Risk Transfer

AI labs, datacenters, hyperscalers, neoclouds hold massive exposure to hardware depreciation, while financiers and end-accounts of alternative asset managers hold exposure to utilization rates, to delays to bring capacity to production, and to compressions in equity multiples of these AI companies. An exchange allows pure risk-transfer, where speculators and hedgers can exchange this risk seamlessly.

Liquidity requires two sides. Small GPU clusters and alternative cloud providers (alt-clouds) are highly motivated to provide volume to an exchange, as it grants them a distribution channel and price transparency they currently lack. Conversely, the Big Three hyperscalers have little incentive to participate initially; opacity protects their margins. However, as consumers (enterprise AI labs) demand transparent benchmark pricing to hedge their compute costs, hyperscalers will eventually be forced to engage.

Liquidity requires two sides

Motivation to participate, by provider type

HIGHLY MOTIVATED

Small GPU clusters and alt-clouds

An exchange grants them a distribution channel and price transparency they currently lack

LITTLE INCENTIVE, INITIALLY

The Big Three hyperscalers

Opacity protects their margin

EVENTUALLY FORCED TO ENGAGE

Hyperscalers, pulled in by Consumers

As enterprise AI labs demand transparent benchmark pricing to hedge their compute costs

c. What Are Good Standards and Basis in Compute?

In compute, there are two elephants in the room when defining the “deliverable grade”:

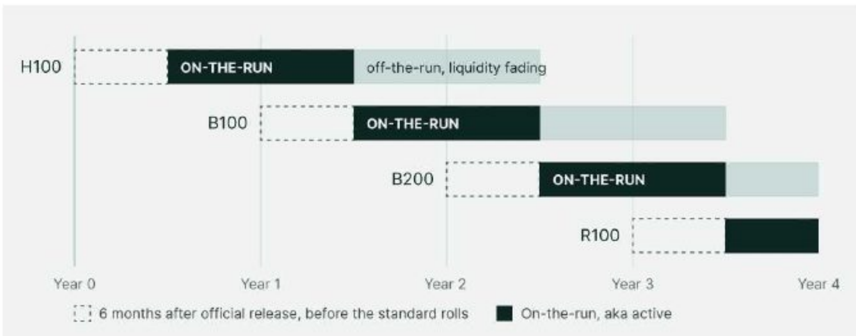
First, unlike traditional commodities such as gold or oil, compute components become obsolete within 3 years and in many cases within a year, including software components such as LLM tokens. As such, standards tied to a specific component architecture will have a finite shelf-life. A H100 was “the thing” 3 years ago, then we moved to B100, B200 and soon R100. Meanwhile, 2026 releases of LLM by top AI labs have significantly increased token consumption to deliver virtually the same outcomes on a broad range of tasks.

Second, compute components are very heterogenous. An “H100 hour” could represent many different goods: SXM or PCIe, running on a spot or reserved instance, hosted by a hyperscaler or neocloud, located in the US or Asia. 1M tokens could be from claude fable 5 or gpt 5.6, and again either under a spot or reserved instance, hosted by a hyperscaler or neocloud, located in the US or in Asia.

The good news is that these problems are not new in financial markets.

Limited shelf-life benchmarks have been a staple of structured products (CDX, CMBX, IOS) going back 20 years, and a tenet of fixed-income futures (TU, FV, TY, US, WN with the shortest one effectively changing benchmark no less than every 2 years). Deliverables that are broadly comparable to one another but with some variations have also been a staple of securitized finance (TBA market) and fixed-income futures (cheapest-to-deliver option) going back more than 20 years.

Therefore, there is no question about the ability to define good standards and basis in compute: participants agree on a methodology for choosing a standard for each component, and a duration over which that standard remains on-the-run aka active. For instance: once a year, take the latest Nvidia GPU starting 6 months after its official release, with price averaged across SXM and PCIe, and across the top twenty providers by volume in the preceding year, volume-weighted. Define basis as any deviation from the standard along a dimension considered as a primary discriminant, again based on participants agreement. For instance: GPU hour basis will exist as a function of chip model (H100 vs B100 vs B200), port type (SXM vs PCIe), instance type (spot vs reserved), host, and region. When a new year comes, roll to the new standard, and the previous one becomes off-the-run (still quoted, but presumably with liquidity on top of it fading out).



The more relevant and complicated questions are about the priority and complexity level of each compute components to introduce standards and basis. To us, priority and complexity are entirely driven by the cost / price / supply-chain / substitution analysis we conducted earlier. Iterating on it yields the following ranking (hereafter 1 being top priority / higher complexity).

We therefore see time to market resembling the following table:

LAYER	FUNCTION	COMPONENTS	PRIORITY	COMPLEXITY
Cloud & Middleware	Specialized Purpose Data Ops	Training	3	3
		Inference, Tokens	1	1
	General Purpose Data Ops	[NO RELEVANCE AT THIS TIME]		
	Orchestration	[NO RELEVANCE AT THIS TIME]		
Datacenter	Compute	GPU, TPU, ASIC	1	2
		CPU, FPGA	3	3
	Memory & Storage	HDD	3	3
		RAM, SSD	2	3
	Orchestration & Communication	[NO RELEVANCE AT THIS TIME]		
Physical Site	Power Delivery & Mgmt	Generators, Transformers, Power Managers, PDU, UPS	2	2
	Power & Energy	Electricity	3	3
	Cooling	Water	3	3
		Liquid Cooling loops, Air HVAC	3	3
Raw Materials	Electronics	Copper, Lithium, Cobalt, Nickel, Silicon	1	1

d. Where is the Demand for Financial Instruments Coming from First?

Let's preamble this section with a reminder that in commodity markets, while cash settlement is standard, physical delivery is critical both for its tethering mechanism ensuring financial instruments remain effective at their stated purpose and because a subset of market participants needs the physical commodity, and not the cash payout resulting from price fluctuations. The same applies to compute, and we go so far as to state that in the early stages of development of compute markets, physical delivery is even more critical to its market participants. An AI lab buying a compute forward doesn't just want financial compensation if prices rise; they need the actual compute to train their models or to serve their expanding client base and expanding usage from existing clients. If an instrument cannot guarantee delivery of the underlying asset, it fails as a capacity planning tool.

Second, and by contrast to other commodity markets, compute markets need substantial debt financing to build out all layers to meet current and anticipated demand (notably because they are so early, and a lot more expensive per unit-of-raw output). That is why in our recent oped in the AI Journal, we introduced the concept of DMBS and DMBX, leveraging what has been created in commercial real-estate lending. In that sector, every year, large debt financings for office buildings, hotels, warehouses, and other commercial assets are pooled into standardized investment products, with different levels of risk and protection packaged for different investors. AI infrastructure needs an equivalent. Just as commercial mortgage-backed securities (CMBS) and the CMBX market standardized capital conduits for commercial real estate, AI infrastructure will require its own equivalent – a market for standardized, tradable infrastructure-backed securities. This "DMBS" and "DMBX" are how the next many trillions of dollars of AI infrastructure will succeed. We are adding these instruments to the typical list of commodity market financial instruments.

infrastructure will require its own equivalent – a market for standardized, tradable infrastructure-backed securities. This "DMBS" and "DMBX" are how the next many trillions of dollars of AI infrastructure will succeed. We are adding these instruments to the typical list of commodity market financial instruments.

We can now approach demand for financial instruments based on satisfying the 4 core needs of a compute market: financing, planning, hedging, and speculating, across the functions and components we previously identified as higher priority, and based on the specific properties of each of these financial instruments (hereafter 1 being higher / more urgent demand).

LAYER	FUNCTION	COMPONENTS	SPOT	FUTURES & OPTIONS	NDFS & SWAPS	DRS	EFPS	BASIS & FTRS	DMBS & DMBX
Cloud & Middleware	Specialized Purpose Data Ops	Training							
		Inference, Tokens	DONE	1	3	1	2	1	
	General Purpose Data Ops								
	Orchestration								
Datacenter	Compute	GPU, TPU, ASIC	DONE	1	3	1	2	1	1
		CPU, FPGA							
	Memory & Storage	HDD							
		RAM, SSD	DONE	2	3	2	3	2	
	Orchestration & Communication								
Physical Site	Power Delivery & Mgmt	Generators, Transformers, Power Managers, PDU, UPS	DONE	2	3	1	3	3	
	Power & Energy								
	Cooling								
Raw Materials	Electronics	Copper, Lithium, Cobalt, Nickel, Silicon	DONE	2	2	1	3	1	

1 · higher, more urgent demand   2   3   DONE

e. How Do We Bootstrap and Scale Liquidity?

Just because something is listed doesn't mean it is liquid. Liquidity needs to be brought, which requires enough market participants across at least 2 categories (hedgers, speculators, swap dealers) and enough market makers transacting in financial instruments seen by all of them as good-enough for their stated purpose on an exchange deemed robust, reliable, and scalable enough to support at least a few hundred millions in daily volume. This takes time, this is not specific to compute markets, and this is only within reach of a small number of teams that come from this very background.

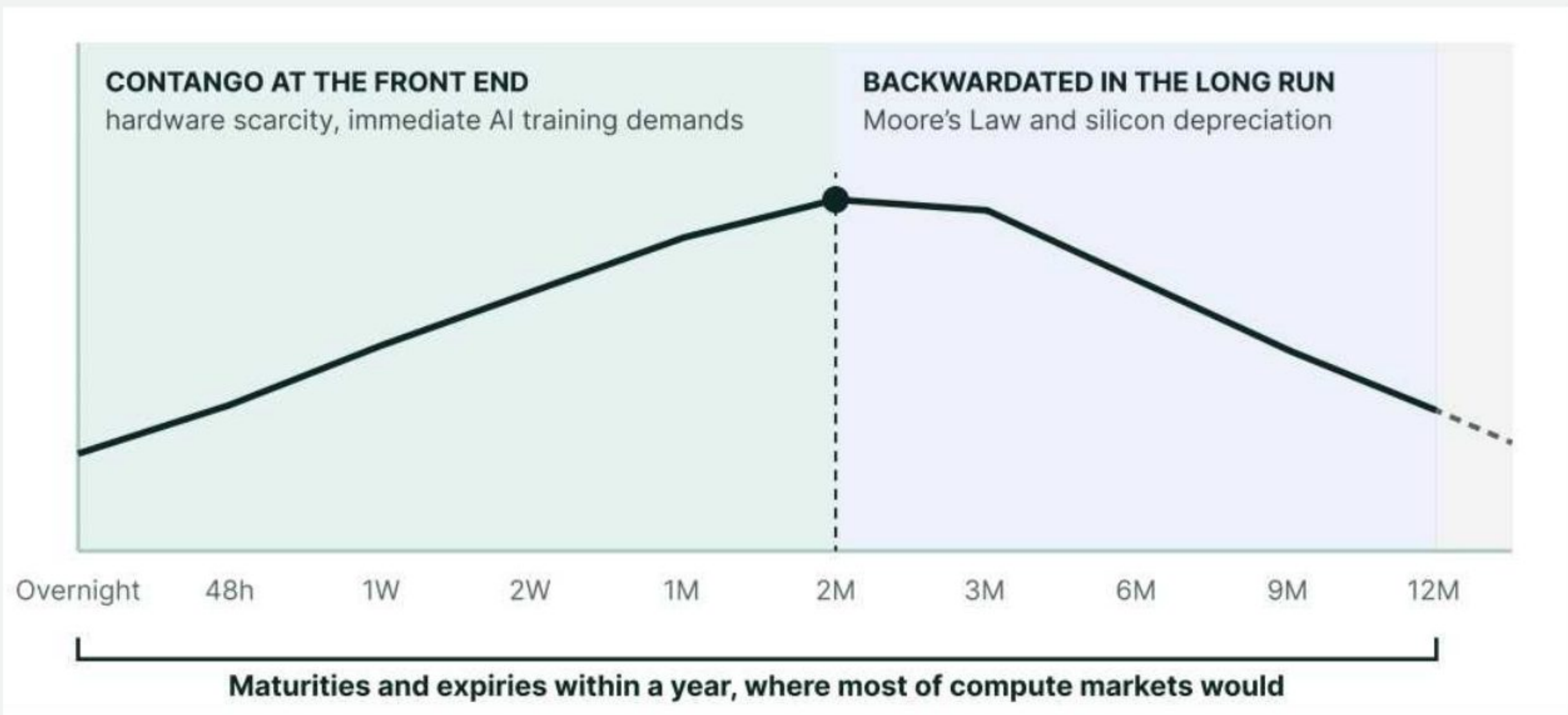
We think that speculators and hedgers are the two categories that startups building compute markets should focus on bringing onboard within their products.

Furthermore, there are challenges specific to compute markets, as they exhibit unique term-structure / curve dynamics. Because technology improves exponentially for several components, the long-term price of a unit of these components predominantly falls. However, in the short term, massive supply bottlenecks for new generation chips create aggressive near-term premiums. Market makers must navigate a curve that may be in steep contango at the front end (due to hardware scarcity and immediate AI training demands) but structurally backwardated in the long run (due to Moore's Law and silicon depreciation). While those dynamics exist in other commodities, their structural presence within several compute components is rather unique. To manage their risk (inventory and curve), market makers have two choices: widening bid-ask spreads (which tends to hinder liquidity), and shortening the range of contract maturities they are willing to actively quote.

Since the standards underpinning compute financial instruments are inherently short-lived, the latter does not appear (yet) to introduce trade-offs for market participants. It is therefore very likely that most of compute markets would concentrate on maturities and expiries within a year, with granular terms within that window similar to what we see in EuroDollars or interest-rate swaps (Overnight, 48h, 1W, 2W, 1M, 2M, 3M, 6M, 9M, 12M).

Steep contago at the front end, structurally backwardated in the long run

The curve market makers must navigate, drawn on the maturity ladder given in the text.



Price, Direction Only • Illustrative: No Observed Series

f. Regulatory Considerations To Keep In Mind

The classification of compute financial instruments dictates their regulatory treatment. Currently, compute is treated primarily as a physical commodity or service, akin to electricity or bandwidth, rather than a security, and thus the CFTC is the primary regulator. There are pockets of instruments that we think would be tremendously helpful for compute that would fall under the SEC's supervision (depositary receipts in particular).

The next table has all the details, but there isn't much in there that is new except for the role that the Commerce Department and the Committee on Foreign Investment in the US are expected to take on given the geopolitical sensitivity of some components within the compute stack. We expand on that further below.

Platforms facilitating the trade of compute instruments have distinct pathways to operate without registering as a full National Securities Exchange or Designated Contract Market (DCM), primarily dictated by the instrument's legal definition and the venue's trading volume. These registration exemptions allow for faster time to market and much lower operating costs, a great backdrop for startups looking to build out compute markets infrastructure, but with limitations for the market participants. We keep the details brief.

The main route leverages CFTC Exclusions, Swap Execution Facility (SEF) and Swap Deal Exemptions which allow the incumbent OTC compute market to never worry about regulation or reporting (the Physical Forward Contract Exclusion) and allow bespoke and custom transactions, typically among large institutional hedgers or physical merchants, to remain private and bilateral (the Bilateral Swap Exemptions) and allow small swap dealers to operate de-minimis The second route only applies to securities and leverages SEC Regulation ATS (Alternative Trading Systems) with exemptions for order transparency and fair execution for small venues (the 5% ADV threshold) and exemptions from reporting for platforms only matching buyers and sellers.

Prediction Markets, while offering a lot of flexibility and light regulatory overhead, have too many limitations to be a significant pillar of compute markets: no physical delivery, limited payoff and structure flexibility. We think they are a useful retail product to gain exposure to compute markets, and can support early bootstrapping of liquidity on more adequate venues.

Regulatory Treatment by Instrument

INSTRUMENT	PRIMARY DERIVATIVES REGULATOR	PHYSICAL / SECONDARY OVERSIGHT
Spot Market	CFTC (Limited to anti-fraud and anti-manipulation authority over cash commodities).	FTC & State Commercial Law (UCC) for general commerce and contract disputes.
Futures	CFTC strictly oversees the Designated Contract Markets (DCMs) and Derivatives Clearing Organizations (DCOs) where futures trade.	N/A (Fully financialized).
Options	CFTC oversees options on commodity futures. SEC oversees options if tied to a security-based underlying.	N/A
NDFs & Swaps	CFTC oversees commodity swaps under the Dodd-Frank Act (traded on SEFs or bi-laterally).	N/A
Depositary Receipts	SEC regulates them if structured as securities/investment contracts. CFTC oversees them if they function strictly as commodity warehouse receipts.	Warehousing standard bodies or data center auditors.
Exchange for Physical (EFP)	CFTC regulates the futures leg and maintains anti-manipulation authority over the physical exchange.	BIS/Commerce Dept. (if the physical leg involves cross-border hardware/compute export).
Basis Swaps & FTRs	CFTC retains concurrent jurisdiction over the derivative nature of the swaps.	FERC (Primary regulator for physical grid/network congestion, assuming compute transmission mimics energy FTRs).

4. National Security: Export Controls, Bureau of Industry & Security, and CFIUS

The rapid militarization and strategic importance of AI computing power have triggered aggressive national security regulations. However, the legal frameworks explicitly bifurcate physical technology transfers from purely financial risk transfers.

The Commerce Department's recent rulings heavily restrict the export of physical hardware and AI model weights based on Total Processing Power (TPP) and DRAM bandwidth (e.g., controlling chips like the NVIDIA H200 or AMD MI325X).

Cash-Settled Derivatives

Export controls govern the transfer of physical items, software, and technology—not financial exposure. A cash-settled compute derivative (like a compute NDF or basis swap) transfers financial risk without transferring actual computing power or data. Therefore, purely cash-settled compute derivatives operate outside the scope of BIS export controls.

Physical Delivery Triggers

If a compute derivative (such as a physically settled futures contract or an EFP) results in the actual delivery of server access, compute cycles, or hardware to a foreign entity, the transaction immediately falls under BIS jurisdiction. The physical delivery leg must fully comply with all export licensing requirements, KYC checks, and end-user verifications.

CFIUS reviews foreign investments in U.S. businesses to assess national security risks, with a hyper-focus on critical technologies like AI and data centers.

Derivative Positions vs. Equity

CFIUS jurisdiction is triggered by equity investments, mergers, or acquisitions that grant a foreign person control or access to non-public technical information of a U.S. business. Purchasing compute derivatives, options, or futures contracts does not confer equity ownership, voting rights, or operational control over a U.S. tech company. Consequently, CFIUS has no jurisdiction over the trading of compute derivatives.

Infrastructure M&A

If a foreign entity attempts to acquire the trading venue itself (e.g., a compute ATS, a SEF, or a U.S. data center backing the depositary receipts), CFIUS would have full jurisdiction to review, block, or force the unwinding of the transaction.

CONCLUSION

Where does this leave us?

01 Compute markets are a matter of when, not if, and the projects best positioned to succeed in that space have deep financial backgrounds

02 Attention in the near term will be on chips, inference, tokens, memory and futures, options, and depositary receipts will be the most adequate instruments

03 A standardized and stratified risk market for datacenter financing is critical in our view, best expressed through DMBS and DMBX

04 Certain technology breakthroughs, generalized heterogenous computing in particular, also have a very substantial role to play in simplifying the functioning, scaling and resilience of compute markets by streamlining the implementation of a "cheapest to deliver" for listed derivatives

References

- Microsoft plans to invest $80 billion on AI-enabled data centers in fiscal 2025
- Meta raises AI data center capex forecast to up to $72bn, blames Trump tariffs for increased cost
- Meta, Microsoft, Amazon, and Alphabet are about to spend a shocking amount of money to dominate the AI era
- Q1 2026 Earnings and Synergy Research Group for cloud market sizing; the CFTC Commitments of Traders report for participant shares.