CHAPTER 24 / 3 MINUTE BRIEF

Markets for Compute

Read the full chapter → / Read this brief online →

The main idea

This book has treated capacity as decisions made inside one organization: what we measure, promise, reserve, and commit to pay. This closing chapter looks outward. Compute already has markets; what 2026 added was an attempt to list futures on the rental price of a GPU, so that buyers and sellers could pass price risk to someone else. As of October 4, 2026, none of those contracts was trading on a U.S. exchange. The work they ask of a buyer, defining exactly what is priced and comparing our own price with an outside one, can start now.

Treat the chapter as advanced reading and a direction of travel; nothing here changes the practice of the previous twenty-three chapters.

Explore this section in the full chapter →

A market appears

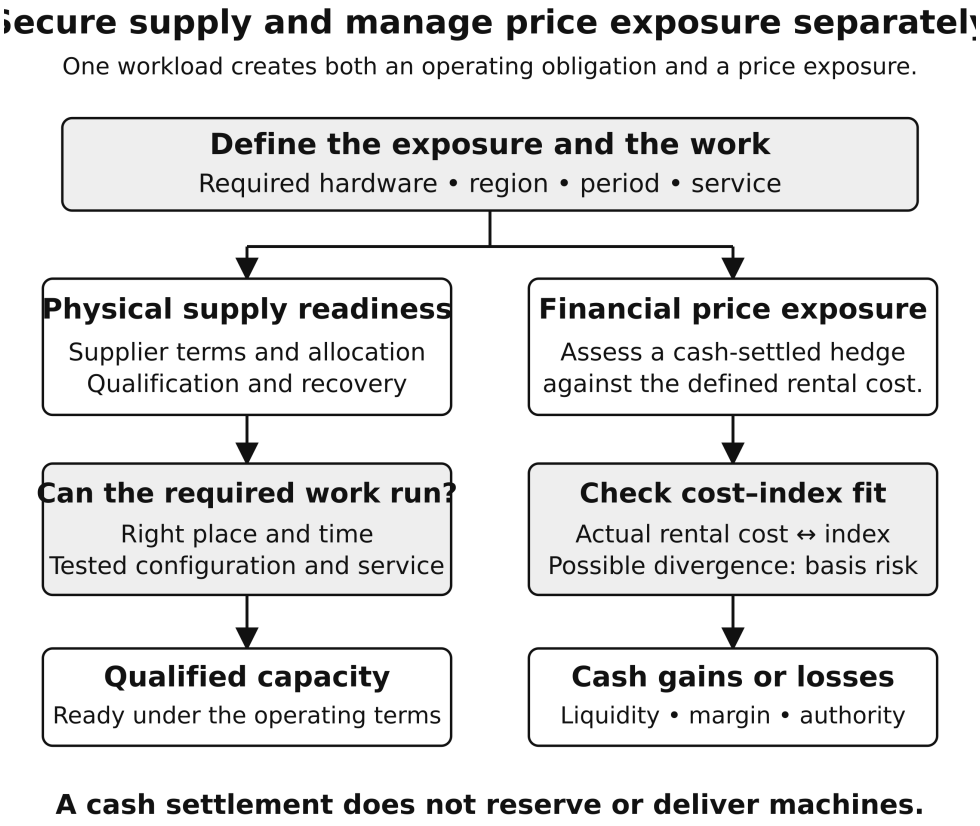


Diagram 24.1. The operating plan secures and qualifies physical supply; the financial plan evaluates price exposure and cost–index fit. A cash-settled hedge can leave basis risk and does not reserve or deliver machines.

Explore this section in the full chapter →

Why exchange trading has been difficult

The first obstacle to exchange trading is heterogeneity. An H100 and a B200 are different resources, and the same chip in two regions or on two networks may support different work. A token buyer is buying a service, not a GPU-hour. A standard contract has to say which differences it includes, which it adjusts for, and which stay with the buyer; Architect's proposed delivery network, for example, would publish basis tables by SKU, memory configuration, and location.[1]

Explore this section in the full chapter →

[1]Architect Financial Technologies, "Architect Financial Technologies Partners with Compute Desk to Launch Exchange-for-Physical Market for GPU Compute", July 8, 2026: the American Innovation Exchange is "forthcoming" and "pending regulatory review"; ComputeConnect would publish "standard basis tables for different SKUs, memory configurations, and locations." Company announcement, checked October 4, 2026.

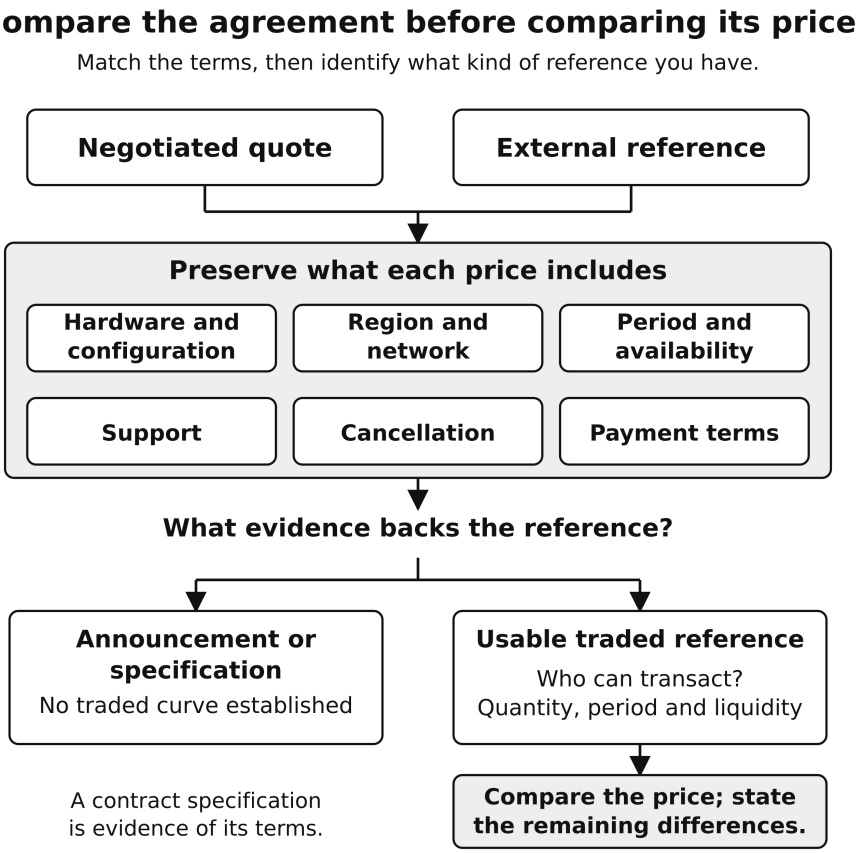


Diagram 24.2. A quote and an outside reference become comparable only after their hardware, configuration, location, period and commercial terms are stated. An announced specification describes terms; a usable traded reference also needs evidence of quantity, period and liquidity.

Explore this section in the full chapter →

What is available and what is announced

Status as of October 4, 2026:

Instrument	What is bought	Price set by	Example (status)
Provider spot	Interruptible capacity now	Provider, without bidding	EC2 Spot (operating)[2]
Prepaid capacity block	Named instances for a future window, prepaid	Provider, from supply and demand at purchase	AWS Capacity Blocks (operating)[3]
Commitment	Lower rates for a spending or usage obligation	Published or negotiated terms	Chapter 20 (operating)
Procurement marketplace	Competing quotes, then a direct supplier contract	Supplier offers	Compute Exchange (operating)[4]
Physical forward	Delivery in a future period at a price fixed now	The two parties	Compute Exchange (offered); OneChronos–Auctionomics (announced)[5]
Resale rights	Permission to sublease contracted time	Other buyers	SF Compute (operating)[6]
Cash-settled index future	Exposure to a monthly average rental index	Exchange trading	CME, ICE, Nodal, Architect (none trading)[7]
Perpetual future	H100 price exposure, no expiry	Crypto-venue trading	Seven venues (trading, thin)[8]

Explore this section in the full chapter →

[2]AWS, "Amazon EC2 Update — Streamlined Access to Spot Capacity, Smooth Price Changes, Instance Hibernation", November 28, 2017, on the change from bidding to smoother provider-set prices.

[3]AWS, Capacity Blocks and pricing and billing documentation, checked October 3, 2026, for the discount and billing rules described.

[4]Compute Exchange, marketplace site, checked October 3, 2026: quotes through a "live auction across verified providers," a services agreement signed directly with the chosen provider, and forward contracts on bilateral terms. Vendor descriptions.

[5]Auctionomics and OCX Group, "Auctionomics and OneChronos Partner on First Tradable Financial Market for GPU Compute", Business Wire, July 29, 2025: price discovery "through bi-lateral forwards"; "first" is the companies' claim. OneChronos, "Compute Markets", checked October 4, 2026: "Coming soon."

[6]San Francisco Compute Company, "About", checked October 3, 2026: clusters leased "on contracts that let our customers sublease." Vendor description.

[7]CFTC, product record 62544, "Approval Pending (90)," checked October 4, 2026; Division of Market Oversight, letter to CME Group, September 21, 2026, extending review of NYMEX Submission No. 26-370 to November 9, 2026, for "novel or complex issues." An extension is not a disapproval.

[8]Loris Tools, "H100 Perpetual Futures", checked October 4, 2026, 21:41 UTC: seven venues, $1.51 million of 24-hour volume, $80,800 of open interest, a $2.647 reference price, and 84.0% of volume on one venue. A third-party, unaudited aggregation of crypto-venue data; an order of magnitude only.

If suitable contracts start trading, I would take four questions to Finance and Treasury first. How closely has the index tracked our actual rate, month by month? How much trades in the months we would hedge? Who has authority to trade? Who funds the margin when the position moves against us? A hedged shortage is still a shortage.[9]

Explore this section in the full chapter →

[9] CME Group, "Learn about Basis: Grains", checked October 3, 2026, and *Self-Study Guide to Hedging with Grain and Oilseed Futures and Options*, on basis (cash minus futures), long hedges, and margin. Applying grain mechanics to compute still requires checking the actual contract.

If we all want the same GPUs

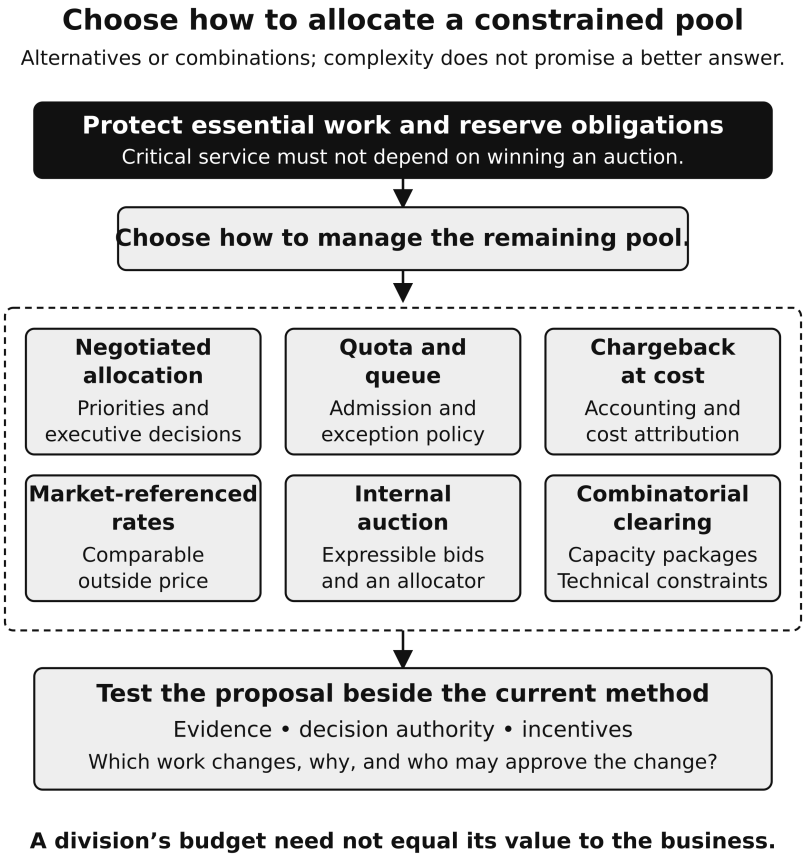


Diagram 24.3. Protect essential work, then choose among allocation and pricing mechanisms for the remaining constrained pool. These are alternatives or combinations; test the proposal against the current method with evidence, decision authority and incentives visible.

Explore this section in the full chapter →

The question the curve answers

None of this has to wait for an exchange. I would take the next GPU commitment due for renewal and write it as one row: hardware and configuration, region, term, price per GPU-hour, minimum spend, cancellation and transfer rights, and payment timing. Beside it go the nearest outside prices for the same period (the on-demand rate, a Capacity Block quote, a published index, a marketplace quote), each with what differs and roughly what the difference is worth. That page goes to procurement and Finance before the negotiation, with one question for Treasury: if a usable contract existed, could we hedge this exposure, and who would fund the margin? After signing, I would record our actual rate beside the index every month. That record is the basis estimate a future hedge will need, and it strengthens the next negotiation whether or not these exchanges ever become deep markets.

Explore this section in the full chapter →