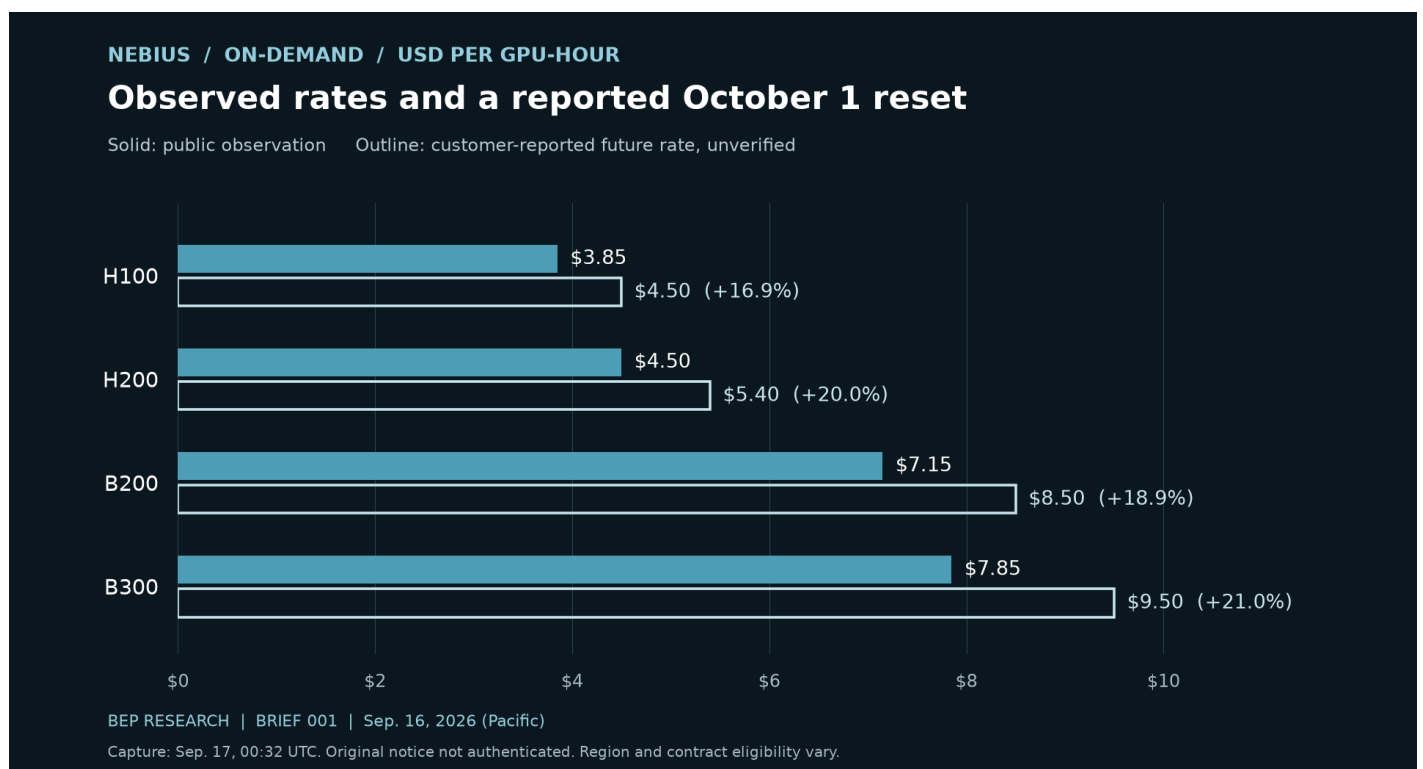




Nebius: the cost of a reported price reset.

A customer-reported October 1 increase puts GPU budgets under review. Verify the applicable quote before translating a headline into a procurement decision.

Research call: Budget for the possibility. The saved public capture still shows the existing rates; the higher figures come from an unauthenticated customer repost. This is a cost sensitivity, not a revenue or margin forecast.



THE BUDGET QUESTION

For B200, the reported change from \$7.15 to \$8.50 adds **\$1.35 per GPU-hour**. At 1,000 GPU-hours, that is **\$1,350**. Eight GPUs used for 125 hours equal 1,000 GPU-hours. The calculation assumes unchanged usage and excludes taxes, storage, networking, discounts, migration and differences in workload efficiency.

Observed public rates captured September 17, 2026 at 00:32:30 UTC (September 16 Pacific). Reported future rates are effective October 1 according to the repost; BEP has not authenticated the original notice or observed its implementation.



WHAT WOULD ESTABLISH PRICING POWER?

A higher on-demand list price can affect marginal buyers while reserved contracts and negotiated discounts behave differently. To evaluate pricing power, examine matched provider quotes, available capacity and realized contract economics together. One reported reset cannot establish fleet utilization, market-wide tightening or an earnings uplift.

REPORTED SCOPE AND COST SENSITIVITY

GPU	Observed	Reported Oct. 1	Change	Added / 1,000 hr
H100	\$3.85	\$4.50	+16.9%	\$650
H200	\$4.50	\$5.40	+20.0%	\$900
B200	\$7.15	\$8.50	+18.9%	\$1,350
B300	\$7.85	\$9.50	+21.0%	\$1,650

USD per GPU-hour, on-demand. Future column is customer-reported and unverified. Reported regions: H100 eu-north1; H200 eu-north1, eu-north2, eu-west1, us-central1; B200 us-central1, me-west1; B300 uk-south1, eu-west2, us-north1. The public landing-page rows do not specify regions. Verify eligibility and CPU/RAM inclusions against the actual offer.

WHAT TO CHECK NEXT

- 1. Contract:** Obtain an authenticated notice or official schedule; confirm region, GPU variant, term, included resources, effective date and existing-customer treatment.
- 2. Alternatives:** Obtain currently available, like-for-like offers. Catalog differences in memory, networking, host shape and commitment can dominate a headline rate.
- 3. Useful work:** Test the actual workload against a fixed quality and latency target. Count failures, retries and paid idle time in cost per successful task.

SOURCES AND PUBLICATION RECORD

1. Nebius public price page - Observed public list rates; dated capture.
2. Nebius compute pricing documentation - Regional scope, billing units and tax exclusions.
3. Customer repost of purported email - Reported future rates only; original email not authenticated.

Frozen edition; no corrections recorded at publication. Later observations may differ. [Online edition](#), [source record](#), [chart](#) and [interactive cost sensitivity](#). Full evidence stays behind subscriber access. Corrections: ben@bepresearch.com.

Commission a focused compute research pilot: sourced memo, inspectable model and a private briefing. Scope, timing and fee agreed before work begins. bepresearch.com/institutional