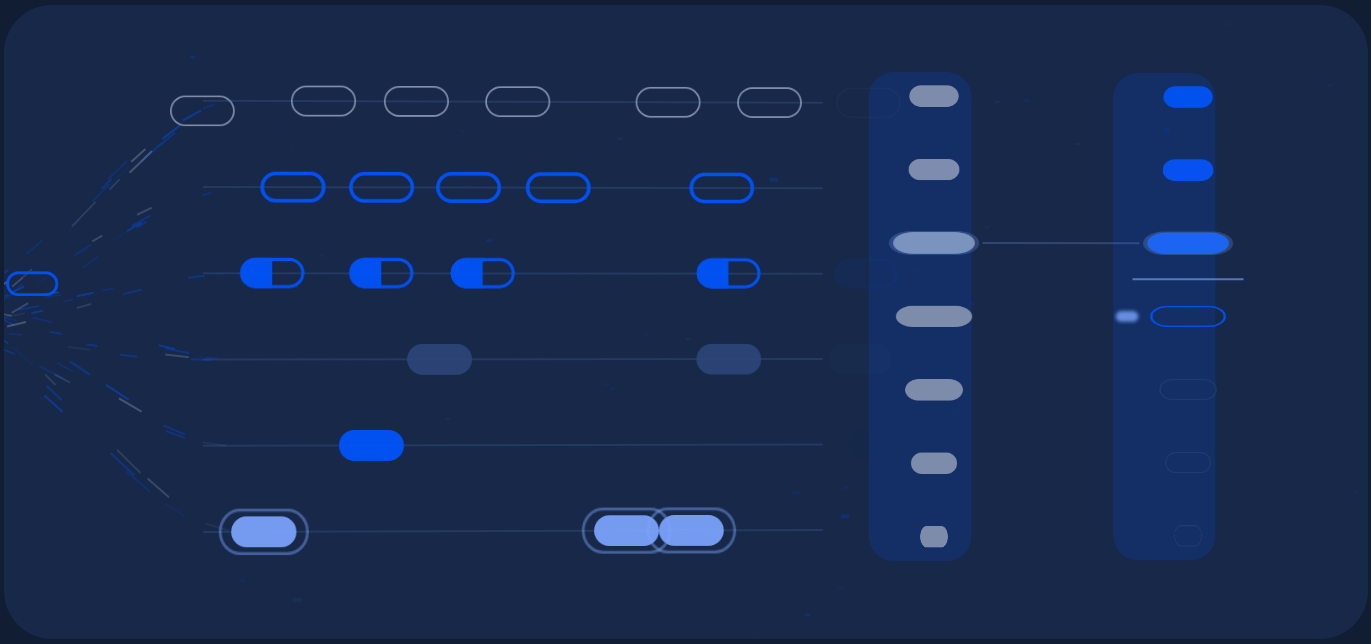
Rillor

METHODS NOTE

How to compare GPU system prices across configurations

Two GPU servers rarely match part for part. Comparing GPU server prices across configurations takes matched builds, line-item adjustment, price per unit of performance or hedonic regression.



Rillor

Contents

01 Summary

02 Background

03 Method

04 Findings

05 Implications

06 Limits of this analysis

07 Sources

Summary

Comparing GPU server prices across configurations takes two steps: bring the builds to one basis, then choose the unit to compare.

- On 8 November 2026, Exxact's configurator listed four eight-GPU HGX B300 servers at $529,958 to $588,841, a spread of $58,883 or 11.1% (derived) [21, 22, 23, 24]. All four share the GPU board, seller, capture date and three-year warranty, and differ in CPUs, host memory, storage and network hardware.
- A price per GPU divides every other part of the system by the GPU count. A GB300 NVL72 rack has one host CPU for every two GPUs, while an NVIDIA DGX B200 server has one for every four (derived) [15, 16].
- Option prices from the same configurator narrow the gap between builds without closing it. Taken to the same CPUs, 3,072 GB of DDR5-6400 and no local drives, three Intel Xeon 6 builds still differed by up to $34,243, or 6.3% (derived) [22, 23, 24].
- Pricing the same build at two dates isolates price change, as a matched-model index does. One default HGX B300 build at Exxact listed at $439,982.40 on 16 February 2026 and $535,513 on 8 November 2026, 21.7% higher (derived) [21, 25].
- Price per unit of performance ranks systems differently by metric. At Lenovo's stated prices of 15 June 2026, an eight-GPU B300 server cost 42.7% more than a B200 server per GPU and per dense FP8 PFLOP/s, and 4.9% less per dense FP4 PFLOP/s (derived) [14, 27].
- Hedonic regression, which estimates a price for each product characteristic from many observations, is how the U.S. Bureau of Labor Statistics (BLS) adjusts server microprocessor prices [1]. A BLS working paper finds its error rising as samples shrink [4], and four builds carry more characteristics than observations (this note's reading).

Background

Buyers checking a quote, finance teams valuing systems in service and analysts building price series all compare prices for systems that differ. Eight-GPU HGX servers share a board and little else. NVIDIA describes HGX as "a single baseboard with eight NVIDIA Rubin, NVIDIA Blackwell, or NVIDIA Blackwell Ultra SXMs" [14], and server makers choose the rest. Dell's PowerEdge XE9780 takes two 6th Generation Xeon Scalable processors with up to 86 cores, while the older XE9680 in the same table takes 4th or 5th Gen Xeon with up to 56 or 64 cores [19].

Public prices come as listings, stated prices and reported purchases, each with its own scope [21, 27, 29]. On 8 November 2026, Exxact's B200 default build TS4-124903244 listed at $236,076.50 and its B300 default build TS4-131188947 at $529,958, 2.24 times as much (derived) [20, 22]. The B200 build carried 1,536 GB of memory on two Xeon Platinum 8468 CPUs and no network card, while the B300 platform listed eight 800 Gb/s ports and 2,048 GB on Xeon 6 [20, 22]. The ratio compares two builds, each with its own host.

Statistical agencies have long handled this problem. The Producer Price Index (PPI) is "a constant quality fixed-input output price index" that measures "'pure' price change based on market factors" [7]. Before 1985, the Bureau of Economic Analysis (BEA) deflated computers by an index "equal to 1 for all periods" [12]. BEA and IBM then developed quality-adjusted indexes, introduced in December 1985, and computer and peripheral equipment prices fell an average 16.9% a year from 1959 to 2005 [12]. Eurostat's manual reports that many national statistical institutes use hedonic methods for computers and other digital equipment [13].

This note sets out those methods for GPU systems and works them on public listings. What an HGX price covers is set out in what an HGX B200 or B300 server price includes (https://rillor.com/insights/what-an-hgx-server-price-includes). The question here is narrower: how to put two builds on one basis before reading the gap between their prices.

Method

What was examined. Five ways to compare prices for products that differ: division by a unit count, matched models, adjustment with line-item (option) prices, hedonic regression and price per unit of performance, each described from its source and applied to public GPU system prices where the data allow.

Sources. Thirty-five sources, each opened. Thirteen cover price index methods: pages and papers from BLS, BEA, Eurostat and four National Bureau of Economic Research (NBER) working papers. Six are manufacturer specification pages or technical guides. Ten are listings, stated prices or reported purchases, two of them Internet Archive captures. Six cover software licensing, price-performance trends, a deflator, benchmark rules and a memory price forecast.

Captures. Exxact's five configurator pages were captured on 8 November 2026 between 07:31 and 07:45 UTC [20, 21, 22, 23, 24]. Two Internet Archive captures of TS4-126572371 date from 16 February and 16 May 2026 [25, 26]. Lenovo's prices are stated as of 15 June 2026 [27]. Each Exxact page shows three tier prices, Essential, Professional and Premium, and this note reads the Essential price as the price of the page's default build [20, 21, 22, 23, 24].

How. Adjustments add each page's option price differences to its default price, and remove the default storage at the value the page gives for removing it. Performance figures are NVIDIA's dense values unless marked sparse [14, 15]. "Derived" and "this note's arithmetic" mark calculations on a source's own numbers. No regression was run and no index was computed. No figure here is a Rillor Compute Index value, a quote or a transaction price.

Where sources disagree. NVIDIA lists 1.4 TB of GPU memory for HGX B200 and 1,440 GB for DGX B200 [14, 15], and this note uses 1,440 GB. For B300, NVIDIA's HGX page gives 2.1 TB per eight GPUs, Dell and Exxact label each GPU 288 GB, and NVIDIA's GB300 NVL72 page gives 20 TB for 72 GPUs, about 278 GB each (derived) [14, 16, 19, 21]. Exxact's tier prices run out of order on two pages: Premium below Professional on TS4-124903244 and Professional below Essential on TS4-104366747 [20, 23]. Each figure is quoted as its source prints it.

Findings

1. Why dividing by GPU count misleads

Per-GPU figures carry the host. Epoch's price-performance data includes "all costs of a DGX server node with eight GPUs" divided by eight, which "accounts for server and interconnect hardware costs" [31]. Its table runs from $16,000 for P100 to $44,000 for H100, and Epoch finds that hardware at a given precision and fixed performance level "becomes 30% cheaper each year" [31]. The divisor keeps the CPUs, memory, storage, networking and chassis inside each figure.

System types divide different bills. An NVIDIA DGX B200 carries two Xeon Platinum 8570 processors with 112 cores in total, 2 TB of system memory configurable to 4 TB, eight 3.84 TB NVMe drives, eight ConnectX-7 adapters and two BlueField-3 DPUs, with three years of support [15]. That is 14 host CPU cores per GPU (derived) [15]. A GB300 NVL72 rack integrates 72 GPUs with 36 Grace CPUs and 2,592 Arm cores [16]. That is 0.5 host CPUs and 36 cores per GPU, against 0.25 CPUs per GPU in DGX B200 (derived) [15, 16]. A per-GPU figure from a rack carries more host per GPU than one from a server.

Fleet figures carry fit-out. In July 2025 IREN announced the purchase of about 2,400 B200 and B300 GPUs for about $130 million "including fit-out costs", which covers servers, storage, labor, InfiniBand, racks and cabling [29]. That is about $54,167 per GPU (derived) [29], a figure of a different scope from a server listing.

Listings divided by eight.

Source and date	System	Price	Per GPU (derived)
Lenovo, 15 June 2026	8x H200, ThinkSystem SR675 V3	$397,801.60	$49,725.20
Lenovo, 15 June 2026	8x B200, ThinkSystem SR680a V3	$550,475.10	$68,809.39
Lenovo, 15 June 2026	8x B300, ThinkSystem SR680a V4	$785,606.50	$98,200.81
Exxact, 8 November 2026	8x B300, four builds	$529,958 to $588,841	$66,244.75 to $73,605.13

Sources: [21, 22, 23, 24, 27].

Four B300 builds at one seller span $7,360 per GPU on the same board (derived) [21, 22, 23, 24]. The Lenovo figures sit on three server models and carry no parts list [27].

National accounts measured the gap. BEA compared the average unit value of single-user computers, from Census shipments data, with its quality-adjusted price index for 2001 to 2005 [12]. Unit values fell 4.9% a year and the index 16.4% a year, a difference of 11.5 percent that the authors describe as the effect of quality adjustment [12]. A price per GPU is a unit value of the same kind (this note's reading).

2. Matched builds and the turnover problem

The method. "A matched-model price index uses prices on the same set of products every period to separate changes in price from changes in the average quality of the product set" [4]. For a minor change in a product, the BLS handbook compares the new price directly with the last price reported under the former specification [5].

One matched GPU server. Exxact's TS4-126572371 page kept one default build across three captures: two 64-core AMD EPYC 9554 CPUs, 24 x 96 GB of DDR5-5600 (2,304 GB), 1.92 TB of U.2 storage, the eight-GPU HGX B300 board, no add-in network card and a three-year warranty [21, 25, 26].

Capture	Essential price	24 x 96 GB to 24 x 128 GB
16 February 2026, 08:27 UTC	$439,982.40	+$7,920
16 May 2026, 11:23 UTC	$513,073.00	+$14,520
8 November 2026	$535,513.00	+$25,080

Sources: [21, 25, 26]. February and May from the Internet Archive.

The Essential price rose 16.6% from February to May, 4.4% from May to November and 21.7% across the period, and the memory step rose to 3.2 times its February price (derived) [21, 25, 26]. Because the build held still, those are price changes for that build at that seller, in the listing class. They describe one configurator, and the series ends when Exxact changes the default build. TrendForce projected on 30 September 2026 that conventional DRAM contract prices would rise 10% to 15% quarter on quarter in the fourth quarter of 2026 and that server DRAM would remain undersupplied [35].

Labels change while builds hold. The board's label read "2.3TB HBM3e" in February and May and "288 GB HBM3e" per GPU in November [21, 25, 26]. A match rests on the parts list.

Turnover. Matching fails when products leave or enter. Adams and Sawyer write that for microprocessors a strict matched-model method "would yield constant price levels, no matter how drastically prices or product qualities differ for new models" [4]. They cite Byrne and colleagues' finding that "only 20 percent of microprocessors introduced from 2010 to 2013 saw a price change" [4]. Byrne, Oliner and Sichel found that the matched-model PPI for microprocessors "likely started to be biased in the mid-2000s" after Intel changed its pricing behavior [9]. Their preferred index fell at an average 43% a year from 2008 to 2013, against 8% for the PPI [9]. For PCs, Pakes found that "on average, the matched model indices indicate no price fall at all", while the hedonic index showed steep declines in every year [8]. He attributes the gap to a selection problem: matched-model indexes omit goods that exit, whose prices typically fall [8].

Across dates, state the deflator. Epoch deflates its prices with what it calls a "data industry producer price index", normalized to October 2022 [31]. That link leads to the PPI for hosting, ASP and other IT infrastructure provisioning services (FRED series PCU5182105182105), which stood at 128.520 in September 2026 against 100 in December 2006 [33]. Any real comparison over time names its deflator (this note's reading).

3. Adjusting with line-item prices

What agencies do. When a product is replaced, BLS staff adjust for quality, and the handbook describes adjustments with cost data and with hedonic regressions. With no information, BLS may treat "any difference in price between the old and the new items" as quality [5]. Eurostat's manual defines direct price comparison as valuing the quality change at zero, and option pricing as valuing it at "some fixed proportion of the market price of features by which the two individual products differ" [13]. Option pricing applies where quality changes are changes of specific features whose prices "can be identified and collected separately", and "has also been used for desktop computers" [13]. Implicit methods, which read quality from the price gap in an overlap period, mislead when an old product is "deeply discounted to clear the remaining stock" [13].

GPU configurators publish option prices. Each Exxact page lists every option with its price difference from the current selection, and shows the HGX board as a selected line with no price [20, 21, 22, 23, 24]. Memory steps differ by page, module size and speed:

Page	Step	Price	Per added GB (derived)
TS4-124903244 (B200, DDR5-4800)	32 x 48 GB to 32 x 64 GB, +512 GB	+$33,440	$65.31
TS4-131188947 (B300, DDR5-6400)	32 x 64 GB to 32 x 96 GB, +1,024 GB	+$21,120	$20.63
TS4-198437339 (B300, DDR5-6400)	32 x 96 GB to 32 x 128 GB, +1,024 GB	+$33,440	$32.66
TS4-126572371 (B300, DDR5-5600)	24 x 96 GB to 24 x 128 GB, +768 GB	+$25,080	$32.66

Sources: [20, 21, 22, 24].

Option prices are differences inside one page. Some options run against the parts they name. On TS4-124903244, the 56-core Xeon 8480+ costs $3,300 less than the 48-core 8468 default, and 2,048 GB on 128 GB modules adds $20,240 against $33,440 for 32 x 64 GB [20]. On TS4-131188947, the 86-core Xeon 6787P adds $22,000 and the 64-core 6767P adds $26,400, and 2,048 GB on 128 GB modules costs $30,800 less than the 32 x 64 GB default [22]. Eurostat's proportion is "a positive number no larger than one" applied to the feature's market price [13].

Storage and network. Drive quantities are not printed, so storage is valued here by the price of removing it: $1,375 to $7,920 across the four B300 defaults [21, 22, 23, 24]. The B200 page's default has no network card, and a ConnectX-7 400GbE single-port card adds $1,815 [20]. The B300 platforms list eight 800 Gb/s ports as part of the platform [21, 22, 24]. TS4-104366747 integrates eight ConnectX-8 adapters and also carries an Intel E810 25GbE card, and its network group offers no option to remove the card [23]. Integrated ports carry no option price, so line items cannot separate them.

4. Hedonic regression as statistical agencies use it

The idea. BLS describes hedonic adjustment as "decomposing an item into its constituent characteristics" and estimating a value for each with regression [6]. Each coefficient is "a measure of the proportional change in price associated with a unit change in the characteristic" [6]. When a product is replaced, the "previous period price of the old version (which was observed) is adjusted using the value of the change in characteristics" [6]. In the CPI's dryer example, a 71.5% raw price gap becomes a 15.5% quality-adjusted change [6].

Two forms. Eurostat's manual writes the regression as the log of price on a constant plus each characteristic times its coefficient, and names two main methods: hedonic imputation and dummy hedonic, which includes the time dummy [13]. A time-dummy model is one equation "assumed to hold for both the reference period and the current period simultaneously", with a dummy for the period [13]. Diewert, Heravi and Silver show "exactly why the results may differ" between imputation and time-dummy indexes, with desktop PCs as the illustration [11].

Server microprocessors at BLS. With its July 2018 data, released on 9 August 2018, the PPI began measuring server microprocessors with "a two-period time dummy hedonic model", the two periods being adjacent quarters [1]. "The time dummy coefficient gives an estimate of price change between the two quarters" [1]. The candidate characteristics were cores, threads, base and turbo frequency, cache per core and thermal design power, plus the SPEC CPU2017 benchmark [1]. BLS kept the benchmark in every model because benchmarks "can capture improvements in microprocessors that are not captured by the other characteristics" [1]. It compared specifications by repeated cross-validation, and "PPI selected the model with the lowest MSE" [1]. In the July 2018 results the model kept SPEC, cores, cache per core and squared log cores with the quarter dummy, and the page says PPI "plans to re-estimate the hedonic server microprocessors model quarterly" [1].

Retransformation bias. From its January 2024 data, BLS adds a bias adjustment term to the hedonic estimates for desktop, notebook and server microprocessors, because prices converted to logs and back give results "typically different than if price relatives are formed directly from prices in levels" [2]. The term is a ratio of mean exponentiated residuals in the two periods [2].

Cloud services. Sawyer and O'Bryan fitted hedonic models to public price lists of three cloud providers, quarterly from the second quarter of 2017 to the second quarter of 2019 [3]. Each package is "essentially a virtual version of a blank computer", and in the AWS model the per-hour price is explained by virtual CPUs,

memory, storage, an SSD indicator, a compute unit rating and a Windows indicator [3]. "The time dummy variable gives the quality adjusted price change between the two quarters" [3]. Byrne, Corrado and Sichel built hedonic indexes for AWS compute, database and storage services from 2009 to 2016 that "fall rapidly", with double-digit declines for all three from early 2014 [10].

What a GPU system model would use. A configurator prints the characteristics such a model would need: GPU board, CPU model and cores, memory capacity and speed, storage, network ports, power supplies and warranty (this note's reading of [20, 21, 22, 23, 24]). Cooling also differs by system: Supermicro's HGX B200 server is air cooled and the GB300 NVL72 is fully liquid-cooled [16, 18]. The constraint is the number of observations, as the next finding shows.

5. Small samples and bias

Error rises as samples shrink. Adams and Sawyer test hedonic specifications in Monte Carlo simulations built from Intel's wholesale microprocessor data: a price for a 1,000-chip direct shipment, seven characteristics and the PassMark score [4]. The root mean squared error (RMSE) of both time-dummy variants "rises as the sample size decreases", the variant with exit and entry indicators "having better performance until about a sample size of 500 observations" [4]. "The time-dummy always has a lower variance for a given sample size", and bias against the population target grows as samples shrink, "but the size of the bias is relatively small" [4]. One of their regressions, 2017 Q3 to Q4, has 183 observations and an R-squared of 0.921 [4].

The conclusion is a trade-off. For "small sample sizes (like those often found in the PPI)", the pure time-dummy model "has lower variance and lower MSE than other specifications", and a population benchmark "reveals a bias-variance tradeoff" [4].

Small samples misbehave. In the cloud study, a "limited number of models makes it difficult to have enough variation in microprocessor characteristics", and third-party benchmark results were "too small to support a hedonic model" [3]. The vCPU variable was dropped as "strongly correlated with cache and TDP", some variables had "counterintuitive signs on their coefficients in some of the models", and one provider no longer published its compute unit rating as of 2021 [3]. Its models used 128 to 236 observations for one provider, 132 to 152 for another and 152 for the third [3].

Data before regression. Eurostat's manual warns that ensuring a hedonic model's procedures and results are satisfactory "can be demanding", and calls the method "of limited use in the HICP since it requires a reasonably large and detailed data set" [13].

GPU systems sit at the thin end. The four B300 builds in Finding 7 share a board and differ in at least six characteristics: CPU model, memory capacity, memory speed, storage, network hardware and power supplies [21, 22, 23, 24]. Four observations cannot identify six or more coefficients, so this note runs no regression (this note's reading). This research found no public hedonic study of GPU server prices. The evidence that matched-model and hedonic indexes drift apart comes from microprocessors and PCs [4, 8, 9], and it gives no direction for GPU systems.

6. Price per unit of performance

NVIDIA's figures. NVIDIA's HGX table gives FP4 Tensor Core performance of 144 PFLOPS sparse and 108 dense for HGX B300, against 144 and 72 for HGX B200 [14]. FP8 is listed at 72 PFLOPS sparse on both boards, and the footnote puts dense at half, 36 PFLOPS (derived) [14]. FP64 is 10 TFLOPS on B300 and 296 TFLOPS on B200, GPU memory 2.1 TB and 1.4 TB, and networking bandwidth 1.6 TB/s and 0.8 TB/s [14].

Five metrics, one pair of systems. Lenovo stated a "usual sale price to the customer" as of 15 June 2026 for configurations it built to be comparable to cloud instances [27].

Metric	8x B200, $550,475.10	8x B300, $785,606.50	B300 against B200
Per GPU	$68,809.39	$98,200.81	+42.7%
Per dense FP8 PFLOP/s (36 and 36)	$15,291	$21,822	+42.7%
Per dense FP4 PFLOP/s (72 and 108)	$7,645	$7,274	-4.9%
Per TB of GPU memory (1.44 and 2.1)	$382,274	$374,098	-2.1%
Per FP64 TFLOP/s (296 and 10)	$1,860	$78,561	About 42 times

Sources: [14, 15, 27]. This note's arithmetic.

The B300 server costs more per GPU and per FP8 operation, and less per FP4 operation and per terabyte of GPU memory. The page itemizes no CPUs, memory or network cards, and the two configurations use different server models, the B300 one "designed for the B300 and Intel Xeon 6 processors" [27]. So the premium mixes the board with the host.

Charts mix bases. AMD lists the MI355X at 5 PFLOPS for MXFP8, 5 for OCP-FP8 or 10.1 with structured sparsity, and 10.1 for MXFP6 and MXFP4 [17]. Its comparison chart shows B200 at 9 for FP8, labeled sparsity, and 4.5 for FP6 [17]. Per GPU, NVIDIA's HGX B200 table gives 9 PFLOPS of sparse FP8 and 4.5 of dense FP6 (derived) [14]. The chart's footnotes add that "Server manufacturers may vary configurations, yielding different results" [17]. A price per FLOP across vendors needs the precision and sparsity of every figure stated.

Trends carry their own bases. Epoch's 2022 study of "470 models of graphics processing units (GPUs) released between 2006 and 2021" found FLOP/s per dollar doubling about every 2.5 years, every 2.95 years for top GPUs and every 2.07 years for GPUs used in machine learning research [32].

Delivered performance. Peak FLOPS are specifications. A price per unit of measured throughput, such as an MLPerf® result, is a derived metric. MLCommons requires that any comparison on a derived metric such as cost "must make the basis for comparison clear in the text and in a footnote", and that derived metrics "must not be presented as official or verified MLPerf metrics" [34]. This note uses no throughput result.

7. A worked comparison from public listings

Step 1: one seller, one date, one class.

Build	CPUs	Host memory	Storage value	Network	Power	Listed price	Per GPU (derived)
TS4-131188947	2x Xeon 6747P	2,048 GB DDR5-6400	$4,400	8x 800 Gb/s on platform	12x 3000W	$529,958	$66,244.75
TS4-126572371	2x EPYC 9554	2,304 GB DDR5-5600	$1,375	8x 800 Gb/s on platform	6+6x 3000W	$535,513	$66,939.13
TS4-198437339	2x Xeon 6740P	3,072 GB DDR5-6400	$1,980	8x 800 Gb/s InfiniBand	12x 3000W	$564,718	$70,589.75
TS4-104366747	2x Xeon 6747P	3,072 GB DDR5-6400	$7,920	8x ConnectX-8 integrated, plus E810 card	6x 6600W	$588,841	$73,605.13

Sources: [21, 22, 23, 24]. Essential prices captured 8 November 2026.

The four share a board, seller, date, evidence class and warranty term, and the raw spread is $58,883, or 11.1% (derived) [21, 22, 23, 24].

Step 2: adjust with each page's own options. The target is two Xeon 6747P CPUs, 3,072 GB of DDR5-6400 and no local drives.

Build	Steps	Adjusted (derived)
TS4-131188947	$529,958, less $4,400 storage, plus $21,120 for 32 x 96 GB	$546,678
TS4-198437339	$564,718, less $1,980 storage, plus $7,040 for 6747P	$569,778
TS4-104366747	$588,841, less $7,920 storage	$580,921

Build	Steps	Adjusted (derived)
TS4-126572371	$535,513, less $1,375 storage, plus $29,040 for 24 x 128 GB DDR5-6400	$563,178, on EPYC 9554

Sources: [21, 22, 23, 24].

Among the Xeon builds, the gaps to TS4-131188947 fall from $34,760 and $58,883 to $23,100 and $34,243, or 4.2% and 6.3% (derived) [22, 23, 24]. More than half of each raw gap remains. The residual sits in parts the configurators offer no option for: chassis, platform, power supplies, drive bays and integrated networking (this note's reading). TS4-198437339 has 12 NVMe bays against 8 on TS4-131188947, and TS4-104366747 carries 6600W power supplies and integrated ConnectX-8 adapters [22, 23, 24]. The EPYC build cannot reach the target CPUs, because its page offers no Xeon CPUs [21].

Step 3: check scope. Warranty terms differ by seller. Exxact's pages include three years of parts and labor, DGX B200 includes three-year Enterprise Business-Standard Support, and a Wiredzone listing of a Supermicro 10U GPU server, no longer available, gave three years of labor and one year of parts [15, 20, 28]. Software differs too. NVIDIA AI Enterprise is "licensed on a per-GPU basis", included with Hopper-based DGX systems and bought separately for Blackwell-based DGX systems [30]. Exxact's B300 page offers an "NVIDIA AI Enterprise Essentials Subscription per GPU" at $4,950 for one year, $14,300 for three and $19,800 for five, and does not say whether the line covers one GPU or eight [22].

Step 4: across sellers, compare builds. The B300-to-B200 ratio is 2.24 at Exxact's defaults and 1.43 at Lenovo's stated prices (derived) [20, 22, 27]. Exxact's B200 default carries 1,536 GB on two Xeon Platinum 8468 CPUs and no network card [20], Lenovo's configurations are not itemized [27], and the two sets of prices are dated nearly five months apart. The two ratios describe different builds and cannot be reconciled from the pages (this note's reading).

Step 5: keep classes apart. The Exxact figures are listings, Lenovo's are prices a maker states with no counterparty, and IREN's is a reported purchase including fit-out [21, 27, 29]. They answer different questions, as evidence classes in GPU pricing (https://rillor.com/insights/evidence-classes-in-gpu-pricing) sets out.

Implications

For buyers checking a quote. Ask for the build line by line and compare it with listings of the same parts. Adjust only with option prices from the same page and date, and treat what remains as unexplained. Check warranty and software scope before comparing totals [22, 28, 30].

For finance teams and valuers. A per-GPU figure is a unit value. Use it between matched builds, and state the system type, since a rack and a server carry different host shares [15, 16].

For analysts building a price series. A matched build at one seller gives clean price change until the build changes [21, 25, 26]. For products with high turnover, the index literature favors hedonic methods [8, 11], and hedonic results depend heavily on sample size [4]. Peak performance metrics need their precision and sparsity stated, and measured throughput needs its basis stated [14, 34]. Recording each observation with its build, date, source and evidence class keeps each method open. Rillor's GPU systems pricing models configurations by GPU family, GPU count, memory, CPU, interconnect, cooling, condition, location and seller type, and every observation keeps its capture time, source and evidence class, so a listing is read as a listing and a quote as a quote.

Limits of this analysis

- Every price here is a listing or a stated price on its date. None is a sale, a quote or a market average.
- The matched build, the line-item adjustments and the four-build comparison come from one seller's configurator. Lenovo's prices come without a parts list [27].
- Adjustments assume that option prices add as shown. A quote may price combinations differently.
- No regression was run and no index computed. The index method evidence comes from microprocessors, PCs, cloud services and consumer goods [1, 3, 4, 6, 8, 9], and this note makes no claim about bias in any GPU system index.
- Exxact's pages print no drive counts and do not say whether the software line covers one GPU or eight [20, 22].
- Exxact prices are as captured on 8 November 2026 (UTC). Configurator prices change, as the matched build shows [21, 25, 26].
- Performance figures are peak specifications. No delivered-throughput result was used [34].
- This note describes methods and public figures. It is not investment advice and gives no view on when to buy.

Sources

1. PPI Introduces Hedonic Price Estimation for Server Microprocessors. U.S. Bureau of Labor Statistics, last modified 16 September 2025, viewed 8 November 2026. https://www.bls.gov/ppi/quality-adjustment/ppi-introduces-hedonic-price-estimation-for-server-microprocessors.htm

2. PPI Modifies Hedonic Quality Adjustments for Microprocessors. U.S. Bureau of Labor Statistics, last modified 16 September 2025, viewed 8 November 2026. https://www.bls.gov/ppi/quality-adjustment/ppi-modifies-hedonic-quality-adjustments-for-microprocessors.htm

3. Exploring quality adjustment in PPI cloud computing (Steven D. Sawyer and Claire O'Bryan). Monthly Labor Review, U.S. Bureau of Labor Statistics, February 2023, viewed 8 November 2026. https://www.bls.gov/opub/mlr/2023/article/exploring-quality-adjustment-in-ppi-cloud-computing.htm

4. Hedonic Price Indexes under Static Pricing: An Application to PPI Microprocessors (Brian Adams and Steven D. Sawyer). BLS Working Paper 582, U.S. Bureau of Labor Statistics, December 2024, viewed 8 November 2026. https://www.bls.gov/osmr/research-papers/2024/pdf/ec240120.pdf

5. Producer Price Indexes: Calculation, Handbook of Methods. U.S. Bureau of Labor Statistics, last modified 19 March 2026, viewed 8 November 2026. https://www.bls.gov/opub/hom/ppi/calculation.htm

6. Hedonic quality adjustment in the CPI: questions and answers. U.S. Bureau of Labor Statistics, last modified 6 October 2026, viewed 8 November 2026. https://www.bls.gov/cpi/quality-adjustment/questions-and-answers.htm

7. Quality Adjustment in the Producer Price Index. U.S. Bureau of Labor Statistics, last modified 20 March 2026, viewed 8 November 2026. https://www.bls.gov/ppi/quality-adjustment/

8. A Reconsideration of Hedonic Price Indices with an Application to PC's (Ariel Pakes). NBER Working Paper 8715, National Bureau of Economic Research, January 2002, viewed 8 November 2026. https://www.nber.org/papers/w8715

9. How Fast are Semiconductor Prices Falling? (David M. Byrne, Stephen D. Oliner and Daniel E. Sichel). NBER Working Paper 21074, National Bureau of Economic Research, April 2015, viewed 8 November 2026. https://www.nber.org/papers/w21074

10. The Rise of Cloud Computing: Minding Your P's, Q's and K's (David Byrne, Carol Corrado and Daniel E. Sichel). NBER Working Paper 25188, National Bureau of Economic Research, October 2018, viewed 8 November 2026. https://www.nber.org/papers/w25188

11. Hedonic Imputation versus Time Dummy Hedonic Indexes (W. Erwin Diewert, Saeed Heravi and Mick Silver). NBER Working Paper 14018, National Bureau of Economic Research, May 2008, viewed 8 November 2026. https://www.nber.org/papers/w14018

12. The Role of Hedonic Methods in Measuring Real GDP in the United States (Dave Wasshausen and Brent R. Moulton). Bureau of Economic Analysis, paper P2006-6, viewed 8 November 2026. https://www.bea.gov/sites/default/files/papers/P2006-6_0.pdf

13. Harmonised Index of Consumer Prices (HICP) Methodological Manual, 2024 edition. Eurostat, European Commission, January 2024, corrected September 2024, viewed 8 November 2026. https://ec.europa.eu/eurostat/documents/3859598/18594110/KS-GQ-24-003-EN-N.pdf

14. NVIDIA HGX Platform. NVIDIA, viewed 8 November 2026. https://www.nvidia.com/en-us/data-center/hgx/

15. NVIDIA DGX B200. NVIDIA, viewed 8 November 2026. https://www.nvidia.com/en-us/data-center/dgx-b200/

16. NVIDIA GB300 NVL72. NVIDIA, viewed 8 November 2026. https://www.nvidia.com/en-us/data-center/gb300-nvl72/

17. AMD Instinct MI355X GPU. AMD, viewed 8 November 2026. https://www.amd.com/en/products/accelerators/instinct/mi350/mi355x.html

18. SYS-A21GE-NBRT GPU SuperServer, specifications. Supermicro, viewed 8 November 2026. https://www.supermicro.com/en/products/system/gpu/10u/sys-a21ge-nbrt

19. Dell PowerEdge XE9780 Technical Guide, Rev. A03. Dell Technologies, August 2026, viewed 8 November 2026. https://www.delltechnologies.com/asset/en-us/products/servers/technical-support/poweredge-xe9780-technical-guide.pdf

20. Configure Exxact TensorEX 10U HGX B200 Server TS4-124903244. Exxact Corporation, prices as captured 8 November 2026 (UTC), viewed 8 November 2026. https://configurator.exxactcorp.com/configure/TS4-124903244

21. Configure Exxact TensorEX 8U HGX B300 Server TS4-126572371 (AMD EPYC). Exxact Corporation, prices as captured 8 November 2026 (UTC), viewed 8 November 2026. https://configurator.exxactcorp.com/configure/TS4-126572371

22. Configure Exxact TensorEX 8U HGX B300 Server TS4-131188947. Exxact Corporation, prices as captured 8 November 2026 (UTC), viewed 8 November 2026. https://configurator.exxactcorp.com/configure/TS4-131188947

23. Configure Exxact TensorEX 8U HGX B300 Server TS4-104366747 (integrated ConnectX-8). Exxact Corporation, prices as captured 8 November 2026 (UTC), viewed 8 November 2026. https://configurator.exxactcorp.com/configure/TS4-104366747

24. Configure Exxact TensorEX 8U HGX B300 Server TS4-198437339. Exxact Corporation, prices as captured 8 November 2026 (UTC), viewed 8 November 2026. https://configurator.exxactcorp.com/configure/TS4-198437339

25. Configure Exxact TensorEX 8U HGX B300 Server TS4-126572371, capture of 16 February 2026. Exxact Corporation, via the Internet Archive Wayback Machine, viewed 8 November 2026. https://web.archive.org/web/20260216082726/https://configurator.exxactcorp.com/configure/TS4-126572371

26. Configure Exxact TensorEX 8U HGX B300 Server TS4-126572371, capture of 16 May 2026. Exxact Corporation, via the Internet Archive Wayback Machine, viewed 8 November 2026. https://web.archive.org/web/20260516112321/https://configurator.exxactcorp.com/configure/TS4-126572371

27. On-Premise vs Cloud: Generative AI Total Cost of Ownership (2026 Edition), LP2368 (Sachin Gopal Wani, David Ellison and Jarrett Upton). Lenovo Press, updated 24 July 2026, viewed 8 November 2026. https://lenovopress.lenovo.com/lp2368

28. Supermicro SYS-A21GE-NBRT-G1 10U GPU server powered by two Intel Xeon 8570 processors. Wiredzone, viewed 8 November 2026. https://www.wiredzone.com/shop/product/10032763-ready-to-ship-supermicro-sys-a21ge-nbrt-g1-10u-gpu-server-powered-by-two-intel-xeon-8570-processors-15460

29. IREN Expands AI Cloud with 2.4k NVIDIA Blackwell GPUs (Exhibit 99.1 to Form 8-K). IREN Limited, 3 July 2025, viewed 8 November 2026. https://www.sec.gov/Archives/edgar/data/1878848/000114036125024747/ef20051566_ex99-1.htm

30. NVIDIA AI Enterprise Licensing, NVIDIA Enterprise Licensing Guide. NVIDIA, viewed 8 November 2026. https://docs.nvidia.com/ai-enterprise/planning-resource/licensing-guide/latest/licensing.html

31. Performance per dollar improves around 30% each year (Robi Rahman). Epoch AI, 23 October 2024, viewed 8 November 2026. https://epoch.ai/data-insights/price-performance-hardware

32. Trends in GPU price-performance (Marius Hobbhahn and Tamay Besiroglu). Epoch AI, 27 June 2022, viewed 8 November 2026. https://epoch.ai/blog/trends-in-gpu-price-performance

33. Producer Price Index by Industry: Data Processing, Hosting and Related Services: Hosting, ASP, and Other IT Infrastructure Provisioning Services (PCU5182105182105). U.S. Bureau of Labor Statistics via FRED, Federal Reserve Bank of St. Louis, updated 15 October 2026, viewed 8 November 2026. https://fred.stlouisfed.org/series/PCU5182105182105

34. MLPerf® Results Messaging Guidelines. MLCommons, policies repository on GitHub, last changed 11 July 2025, viewed 8 November 2026. https://github.com/mlcommons/policies/blob/master/MLPerf_Results_Messaging_Guidelines.adoc

35. AI Server Demand Sustains Memory Contract Price Increases in 4Q26, While Consumer-Side Pressure Persists, Says TrendForce. TrendForce, 30 September 2026, viewed 8 November 2026. https://www.trendforce.com/presscenter/news/20260930-13258.html

About this report

Published by Rillor on 13 November 2026. Superintelligence (SI) tools assist research and drafting, and every figure cites its source. Rillor builds and secures SI systems, supplies the data they learn from and delivers the infrastructure they run on. We apply SI to financial and compute markets.

Cite as: Rillor. How to compare GPU system prices across configurations. 13 November 2026. https://rillor.com/insights/comparing-gpu-system-prices-across-configurations

Notices

Rillor is not a registered investment adviser or commodity trading advisor and does not provide investment or trading advice. Research, forecasts, data and software described on this site are for research and engineering use. Nothing here is an offer or recommendation to buy or sell any security, commodity interest or digital asset. Past or simulated results do not indicate future results.

Rillor Compute Index series are for research and planning use. They are not offered for settlement of any contract.

NVIDIA, AMD, Ceph, Kubernetes and related product names are trademarks of their owners and are used only to identify products. No affiliation or endorsement is implied.

Work with Rillor

Send the systems, configurations and conditions you need priced, and what the data has to support. Contact us at rillor.com/contact?type=index or contact@rillor.com. More on GPU systems pricing: rillor.com/markets/gpu-pricing#systems