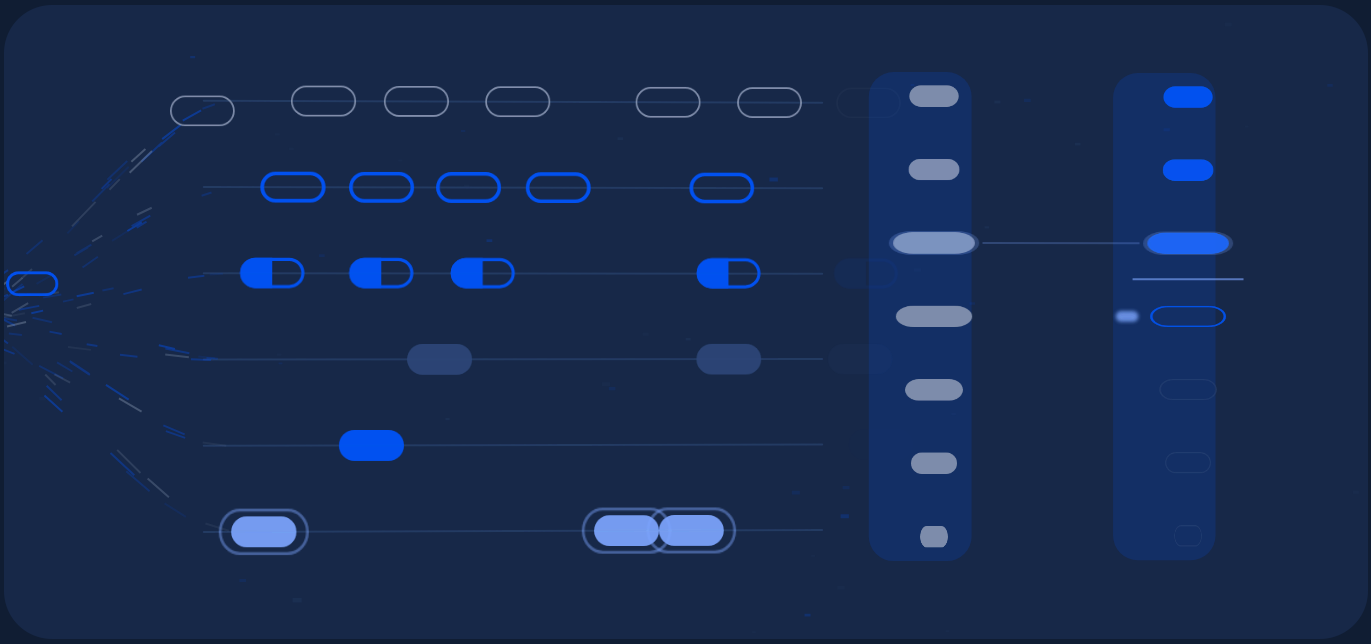
Rillor

EXPLAINER

NVL72 rack prices: what a GB200 or GB300 rack quote covers

NVIDIA has confirmed no GB300 NVL72 price. Dated figures for one GB300 rack run from $3.7 million to $6.5 million, and each covers different parts, terms and services.



Rillor

Contents

01 Summary

02 Background

03 Method

04 Findings

05 Implications

06 Limits of this analysis

07 Sources

Summary

An NVL72 rack price is a figure for one bundle of 72 GPUs, 36 CPUs, switch trays, power shelves and cooling parts, and the public figures for it differ in scope, date and terms [1, 2, 3].

- NVIDIA's GB200 NVL72 and GB300 NVL72 product pages show no price, and Tom's Hardware reported in March 2026 that NVIDIA has never confirmed NVL72 list prices [1, 2, 12]. Every public rack figure reviewed is an analyst estimate, a press report, a newsletter's figures from purchase orders or a figure in a securities filing [10, 11, 12, 13, 26].
- Dated figures for one GB300 NVL72 rack run from $3.7 million to $4.0 million in a March 2025 analyst note to $6.0 million to $6.5 million in a March 2026 press report [11, 12]. That is about $51,400 to $90,300 per GPU (derived).
- The Data Gravity newsletter reviewed GB300 NVL72 purchase orders in August 2026 and, with its figures adjusted for the buyer's anonymity, put the rack just under $5.0 million. In its teardown the compute trays are about 87% of the value, and power shelves, bus bar, cold plates, steel, cabling and integration together under 8% [13].
- Liquid cooling is a small line inside a GB300 NVL72 rack. Morgan Stanley valued the cooling parts at $49,860, as Tom's Hardware reported [14], and a 250 kW in-rack cooling unit added about $32,000 in the orders Data Gravity reviewed [13].
- A rack quote can leave out the delivery term, freight, duty and sales tax, liquid cooling outside the rack, the scale-out network, storage, installation, spares, software licenses and warranty years four and five [6, 8, 13, 19, 26, 28].
- Cluster purchases in filings are not rack prices. HIVE's expected spending, as of August 2026, on a cluster of 2,016 GB300 GPUs with InfiniBand, storage and warranties comes to about $91,800 per GPU, and its GB200 cluster price to about $71,600 (arithmetic on [23, 24]). IREN's filed Dell agreements redact the quantities behind its $5.8 billion order [16, 19].
- NVIDIA AI Enterprise software lists at $4,500 per GPU for one year, or $324,000 for the 72 GPUs of one rack (arithmetic on [27]), and NVIDIA states that Blackwell DGX systems need these licenses bought separately [28].

Background

The NVL72 is sold as a rack. NVIDIA describes the GB300 NVL72 as a fully liquid-cooled, rack-scale platform of 72 Blackwell Ultra GPUs and 36 Grace CPUs, and the GB200 NVL72 as 36 Grace CPUs and 72 Blackwell GPUs in one 72-GPU NVLink domain [1, 2]. Makers deliver it built and tested. Dell's Integrated Rack Scalable Systems program delivers its GB300 NVL72 product, the PowerEdge XE9712, "as fully built and tested sets of racks" [7]. Lenovo integrates its water-cooled racks at the factory and retests them at rack level so that a rack can be deployed directly at the customer site [6].

A widely quoted figure for such a rack, about $3 million, comes from a May 2024 analyst estimate for the GB200 NVL72 [10]. Data Gravity, a newsletter, wrote in August 2026 that the $3 million rack "described GB200 and answers the wrong question" [13]. Later GB300 figures differ by millions of dollars, and each has its own scope: the rack alone, the rack with a cooling unit, the rack deployed, or a cluster with its network, storage and warranties [11, 12, 13, 23, 26].

This report is for buyers checking a rack quote, finance teams budgeting a purchase, lenders valuing racks as collateral and analysts building price series. It covers the rack's parts, the dated figures and their evidence, the lines inside and outside a rack price, cluster purchases and rental rates. Rack power, cooling plant and colocation are a separate subject and appear here only where a price line depends on them.

Method

What was examined. Four groups of sources.

- Specifications: NVIDIA's product pages, its user guide for DGX GB racks, its GB300 SuperPOD reference architecture and a technical blog post [1, 2, 3, 4, 5, 8, 9], Lenovo's product guide for its GB300 NVL72 rack [6] and Dell's PowerEdge XE9712 specification sheet [7].
- Price reports: three Tom's Hardware articles [10, 12, 14], 9to5Mac's report of an analyst note [11], the Data Gravity newsletter's analysis of purchase orders [13], CoreWeave's rental price list [15] and CNBC's report of a chip price given by NVIDIA's chief executive [33].
- Filings with the U.S. Securities and Exchange Commission (SEC): IREN's Dell agreements, statement of work, annual report and releases [16, 17, 18, 19, 20, 21, 22], three HIVE releases and its quarterly report [23, 24, 25, 34], a Forum Markets investor presentation [26], NVIDIA's quarterly report [31] and a Boost Run service agreement [32].
- Terms: NVIDIA AI Enterprise pricing and licensing, NVIDIA's page on what a DGX purchase includes [27, 28, 29] and the International Trade Administration's guide to Incoterms [30].

How. Each figure was recorded with its date, the rack it names, what it covers in the source's own words, the route by which it reached print and its evidence class. The classes are the six of Rillor's GPU systems pricing method: specification, listing, quote, indicative price, reported sale and verified transaction. Per-GPU figures divide a rack figure by 72, or a cluster figure by the GPU count its filing states, and are labelled as derived.

Rules. Figures are never averaged, and no single figure stands for the price of a rack. A cluster figure is never presented as a rack figure. Where a filing redacts quantities, no per-GPU figure was derived. Blogs that repeat estimates without a primary source were left out, and so was a broker's undated rack ballpark. Two reseller listings and two maker datasheets refused automated access and were left out too. All sources were viewed on 8 November 2026, except [34], viewed on 14 November 2026.

Findings

1. What one NVL72 rack holds

NVIDIA's user guide for DGX GB racks, last updated 17 September 2026, describes the complete rack as "compute trays with one or two compute boards, NVLink switch trays, an NVLink passive copper cable backplane, power shelves, a bus bar, and liquid cooling manifolds" [3]. Each NVL72 rack holds 18 compute trays of one rack unit, nine NVLink switch trays of one rack unit, two top-of-rack switches for management and the power shelves [3].

Part	Per compute tray	Per rack of 18 trays (derived)
CPUs and GPUs	2 Grace CPUs, 4 Blackwell or Blackwell Ultra GPUs	36 CPUs, 72 GPUs
Cluster network	4 single-port OSFP cards: ConnectX-7 at 400 Gb/s (DGX GB200) or ConnectX-8 at 800 Gb/s (DGX GB300)	72 cards
Storage and management network	2 BlueField-3 DPUs, dual port 400 Gb/s	See the note on counts below
Data cache	4 x 3.84 TB E1.S NVMe drives, software RAID 0	276.48 TB
Boot drive	1 x 1.92 TB M.2 NVMe drive	18 drives

Sources: [3]. Rack totals are the tray figures times 18.

Each NVLink switch tray carries two NVLink switches with 72 NVLink ports each [3]. Liquid runs up and down the rack through manifolds and then through cold plates on the CPUs and GPUs, while the trays' networking and storage devices are air cooled [3]. Lenovo's guide for its GB300 rack adds cold plates for the ConnectX-8 adapters and all NVSwitch components, and puts the heat split at about 90% to liquid and 10% to air [6].

The two generations differ in GPU memory and network speed. NVIDIA lists 13.4 TB of HBM3E for the GB200 NVL72 and 20 TB of GPU memory for the GB300 NVL72, about 1.49 times as much (derived) [1, 2]. The GB300 NVL72 provides 800 Gb/s of network connectivity per GPU through ConnectX-8 [1], against 400 Gb/s ConnectX-7 cards on the DGX GB200 [5]. Both racks list 17 TB of LPDDR5X CPU memory and 130 TB/s of NVLink bandwidth [1, 2].

The power shelf changed between generations. NVIDIA added electrolytic capacitors to the GB300 NVL72 power shelves, about 65 joules of energy storage per GPU, and measured a 30% cut in peak grid demand while training a Megatron LLM [9]. NVIDIA's user guide puts rack power at about 120 kW, from eight power

shelves of six 5.5 kW supplies [3]. Lenovo states 135 kW TDP for its GB300 rack and up to 155 kW at peak, depending on workload [6].

Makers let buyers vary the rack. Lenovo's customers choose six or eight power shelves, depending on configuration and redundancy, pick from several cooling unit options, and can fit up to eight E1.S drives per compute tray where the reference design populates four [6]. Dell builds the XE9712 into its IR9048, a 48-unit Open Rack version 3 rack whose 33 kW power shelves hold six 5,500 W supplies [7]. The counts also differ between NVIDIA documents: the DGX GB300 page lists 18 dual-port BlueField-3 cards for the rack, the user guide's tray table gives two per tray, and the DGX GB200 page lists 36 [3, 4, 5]. A quote that states its part counts settles such differences for one order.

2. NVIDIA publishes no list price

NVIDIA's GB300 NVL72, GB200 NVL72, DGX GB300 and DGX GB200 pages list specifications and no price [1, 2, 4, 5]. Tom's Hardware reported in March 2026 that "Nvidia has never confirmed the list prices of its NVL72 or NVL144 products" [12]. Lenovo's product guide and Dell's specification sheet carry no price either [6, 7].

Among the sources reviewed, the one price from NVIDIA's leadership is for a chip. In March 2024 NVIDIA's chief executive told CNBC that a Blackwell GPU would cost between $30,000 and $40,000, and later added that the cost is also about designing data centers and integrating into other companies' data centers [33]. A rack price covers more than its chips: CPUs, switch trays, power shelves, cooling, the rack itself, assembly and test [3, 6, 7].

Every public NVL72 rack price found therefore belongs to one of three evidence classes: indicative price, quote or reported sale. None is a verified transaction, a sale confirmed by invoice, payment and delivery records. The one listing reviewed is a rental price for a GB200 NVL72 instance, in Finding 7 [15].

3. The dated public figures, rack by rack

Date	Source and route	Rack	Figure	What it covers	Per GPU (derived)	Evidence class
14 May 2024	HSBC analysts, via a Barron's writer's post, via Tom's Hardware [10]	GB200 NVL72	About $3.0 million	The rack, analyst estimate	About $41,700	Indicative price
25 Mar 2025	Loop Capital client note, quoted from Investor's Business Daily by 9to5Mac [11]	GB300 NVL72	$3.7 to $4.0 million each	A reported order of about 250 systems for about $1 billion	$51,400 to $55,600	Indicative price

Date	Source and route	Rack	Figure	What it covers	Per GPU (derived)	Evidence class
24 Mar 2026	Unnamed source, Tom's Hardware [12]	GB200 NVL72	$2.8 to $3.4 million	The rack, by configuration	$38,900 to $47,200	Indicative price
24 Mar 2026	Unnamed source, Tom's Hardware [12]	GB300 NVL72	$6.0 to $6.5 million	The rack, by configuration	$83,300 to $90,300	Indicative price
31 Aug 2026	Data Gravity, purchase orders reviewed in August 2026 [13]	GB300 NVL72	Just under $5.0 million, about $32,000 more with a 250 kW cooling unit	The rack, Ex Works, small orders	About $69,400	Reported sale
31 Aug 2026	Data Gravity [13]	GB300 NVL72	About $5.7 million	The rack deployed in an existing facility	About $79,200	Indicative price
5 Oct 2026	Forum Markets investor presentation, SEC filing [26]	GB300 NVL72	About $6.4 million per 72-GPU rack	Rack supply, "all-in", from an unnamed vendor	About $88,900	Quote

Per-GPU figures divide each rack figure by 72.

Each figure carries its own warning. Tom's Hardware presented the HSBC numbers as "merely analyst estimates" and wrote that actual prices depend on individual contracts, the volume ordered and other negotiations [10]. The same estimate put one GB200 superchip at $60,000 to $70,000, so the 36 superchips of a rack would come to $2.16 million to $2.52 million of the $3 million (arithmetic on [2, 10]). The Loop Capital figure is an analyst's reading of a reported order, relayed by two publications [11]. Tom's Hardware's 2026 source described the GB200 figure as one for a rack optimized for training and HPC and the GB300 figure as one for an inference-optimized rack, and warned that many quotes "represent only the ODM price without a proper warranty" [12].

Data Gravity's figures rest on purchase orders and rack bills of materials in two configurations, with and without an in-rack cooling unit, with the figures "adjusted modestly to preserve the buyer's anonymity" [13]. It calls them "real transacted numbers for small orders, not list prices", and expects hyperscale volume buyers to do better [13]. Its figure of about $69,000 per GPU slot compares with about $40,000 to $50,000 reported for the GPU module alone [13].

Forum Markets' presentation states that rack cost and supply were established with "Vendor B" at about $6.4 million per 72-GPU rack, about $88,000 per GPU all-in, and compares that with a quotation for B300 server pods at about $100,000 per GPU [26].

Among the dated GB300 rack figures, $6.5 million is 1.76 times $3.7 million, a year apart and from different sources (derived). The sources do not separate how much of that gap comes from time, configuration, warranty or order size.

4. What sits inside the rack price

Data Gravity's teardown puts the compute trays at about 87% of the rack's value, and the 72 GPU modules inside them at about $3.4 million [13]. With the NVLink fabric added, NVIDIA's content exceeds 92%, and what an infrastructure investor might picture as the rack, the power shelves, bus bar, cold plates, steel, cabling and integration, comes to under 8% [13]. Divided by 72, $3.4 million is about $47,200 per GPU module (derived), inside the $40,000 to $50,000 range the newsletter cites for the module alone [13]. Tom's Hardware wrote in March 2026 that the fully assembled L10 compute trays NVIDIA reportedly plans to ship for Vera Rubin, with CPU, GPUs, memory, networking, power delivery and liquid cooling, could account for about 90% of a server's cost [12].

Morgan Stanley's report, as Tom's Hardware described it in November 2025 after a social media account obtained it, valued the liquid cooling parts of a GB300 NVL72 rack at $49,860 [14]. That total is about $2,260 per compute tray, $40,680 for 18, and about $1,020 per switch tray, $9,180 for nine [14]. Cold plates for CPUs and GPUs cost $300 each and those for NVSwitch chips $200, and the cooling content of a GB300 rack is about 20% above the GB200's [14]. Against Data Gravity's $5.0 million rack, $49,860 is about 1.0%, and against the low end of Tom's Hardware's GB300 range about 0.8% (arithmetic on [12, 13, 14]).

The in-rack cooling unit is a separate choice. Data Gravity priced one configuration with a 250 kW unit at about $32,000 above the rack without it [13], about 0.6% of the rack (derived). Lenovo lists "Various CDU Options" in its cooling solution, together with the rear manifold, quick disconnects and cold plates [6]. Whether a quote includes a unit, and of what capacity, is a line to check.

5. What a quote can leave out

Delivery term. Data Gravity describes a purchased rack as Ex Works (EXW), "your problem from the loading dock" [13]. In its account, putting the rack into service in an existing facility adds five line items and brings it to about $5.7 million, while data center construction, cluster-level core switching and storage stay outside as campus decisions [13]. The International Trade Administration explains that Incoterms rules specify who pays for and manages shipment, insurance, documentation and customs clearance, and when the risk of loss or damage passes from seller to buyer [30]. EXW (Ex Works) and DDP (Delivered Duty Paid) are two of those rules [30].

Freight, duty and tax. The B300 quotation summarized in the Forum Markets filing, issued 1 September 2026 and valid to 15 September 2026, did not include liquid cooling, freight, duty or sales and use tax, and did not price the accelerator apart from the server [26]. That quotation covers B300 server pods. It still shows the exclusions a written quote can carry.

Export terms. NVIDIA's quarterly report filed on 26 August 2026 lists the GB200 NVL72 among products covered by U.S. licensing requirements, announced in October 2023, for exports to China and to Country Groups D:1, D:4 and D:5, including Saudi Arabia, the United Arab Emirates and Vietnam and excluding Israel [31]. A rack's destination is part of its terms.

Network and storage outside the rack. NVIDIA's DGX SuperPOD design for GB300 groups eight DGX GB300 racks, 576 GPUs (derived), in each scalable unit, with a thermal design power of 1.2 MW per unit, and scales to 128 racks with 9,216 GPUs [8]. Around the racks it adds an InfiniBand compute fabric of Q3400 switches with four Quantum-X800 connections per system, Spectrum-4 SN5600 Ethernet switches for storage and in-band management, a Unified Fabric Manager appliance, management nodes and storage from a vendor certified in NVIDIA's DGX Storage Program [8]. Lenovo's guide lists two top-of-rack Ethernet switches among its rack components and says the top-of-rack switches depend on the customer's network topology [6].

Installation, support and warranty. Dell's non-cancellable, non-returnable (NCNR) agreement with IREN for GB300 racks refers to Dell's ProSupport One for Data Center support, treats rack integration and field deployment services as professional services, and lets the buyer elect an additional limited warranty for the XE9712 in years four and five, priced on its own line [19]. Lenovo's standard warranty is three years of customer-replaceable unit and onsite service at next business day, 9x5, with three-year and five-year terms, upgrades to 4-hour or 2-hour response and a 6-hour fix, and one-year or two-year extensions [6]. NVIDIA's DGX GB300 and DGX GB200 pages list three years of business-standard hardware and software support [4, 5]. Deployment is sold in named tiers, from Lenovo's Standard Deploy to Premier Deploy Plus and from Dell's ProDeploy suite [6, 7].

Spares. Lenovo recommends onsite spare compute and switch trays of at least 2% of the order to support a next-business-day service target, and states that spares are required for that service [6]. For one rack of 18 compute trays, 2% is 0.36 of a tray (derived), and spares come as whole trays.

Software. NVIDIA AI Enterprise is licensed per GPU, with a license required for every GPU installed [28]. Its list prices are $4,500 per GPU for one year, $13,500 for three years and $18,000 for five, or $22,500 for a perpetual license with five years of support [27]. For one rack's 72 GPUs that is $324,000 for one year and $1,296,000 for five at list (arithmetic on [27]). NVIDIA's licensing guide states that these licenses "need to be purchased separately for NVIDIA DGX systems using the NVIDIA Blackwell architecture" [28], while the DGX GB300 and DGX GB200 pages list NVIDIA AI Enterprise under software, with Mission Control, DGX OS and Ubuntu [4, 5]. NVIDIA's DGX resources page says a purchase may include Mission Control or Base Command Manager, that Base Command Manager is free without support for up to eight accelerators per system, and that DGX OS is included with a system [29]. A quote should state which of these it covers.

Payment and cancellation. Dell's order with IREN cannot be cancelled, rescheduled or changed without Dell's written consent, accepts partial shipments and is invoiced per shipment, payable net 30 days from the later of the invoice date or shipping [19].

6. Cluster purchases in filings

The filings reviewed describe clusters more often than racks, and two of their figures name lines beyond the racks: IREN's deployment services, InfiniBand, cabling and software, and HIVE's associated hardware and service warranties [16, 18, 23].

Filing	System	Figure	What it covers	Per GPU (derived)
HIVE, 17 Aug 2026 [23]	2,016 GB300 GPUs, 28 racks (derived)	About $185 million of expected capital spending	Cluster with Quantum-X800 InfiniBand, storage and all associated hardware and service warranties	About $91,800
HIVE, 27 Oct 2026 [24, 25]	2,303 GB200 GPUs (2,304 in June)	$165 million	The GPU cluster, which the June release describes as interconnected with Quantum InfiniBand	About $71,600
IREN, 3 Nov 2025 [16, 18, 19]	GB300 racks from Dell, quantities redacted	About $5.8 billion	Deployment services, servers, InfiniBand, cabling, software and licensing	Not derivable

Per 72 GPUs, HIVE's GB300 cluster spending comes to about $6.61 million and its GB200 cluster price to about $5.16 million (arithmetic on [23, 24]). HIVE's GB300 figure sits above every dated GB300 rack figure in Finding 3 [11, 12, 13, 23, 26]. Its GB200 figure sits above the GB200 rack figures [10, 12, 24]. HIVE's GB300 spending covers the cluster and all associated hardware and service warranties, for GPUs linked by Quantum-X800 InfiniBand and supported by VAST Data storage [23]. HIVE called the GB200 figure "a very attractive price for NVIDIA GPUs in today's market", noting that NVIDIA hardware had risen substantially in price over the previous six months [24]. It financed about $129 million of it at a fixed 9.0%, secured by the GPUs and related assets [24]. In August 2026 HIVE expected the GB300 cluster to become operational later in 2026, at a facility in Merritt, British Columbia [23]. Its quarterly report filed on 9 November 2026 expects deployment in the first quarter of calendar 2027, with capital spending still put at about $185 million [34].

IREN's figure cannot be split. Its release says it agreed with Dell "to purchase the GPUs and ancillary equipment for approximately $5.8 billion", and its presentation's footnote says the estimate "Includes deployment services, servers, InfiniBand, cabling, software and licensing costs" [16, 18]. The 8-K schedules delivery in several tranches from March 2026, with payment in installments within 30 days of each tranche shipping [17]. The NCNR agreement redacts the racks per phase, the total order value and the cost of the added warranty years [19]. The statement of work with IREN's cloud customer redacts the GPU quantities and the price per GPU-hour [20]. No per-rack or per-GPU figure follows from these filings. IREN's annual report lists further Dell purchase and NCNR agreements dated 4 March 2026 and 19 May 2026 [21]. On 13 August 2026 IREN announced that the first of four 50 MW deployments for that customer had been delivered and accepted [22], about nine months after the November 2025 agreement (derived).

Forum Markets' presentation shows what a full pod adds beyond servers, for B300 rather than GB300 [26]. In its cost bridge for a B300 pod, GPU servers are 67% of the all-in cost per GPU, storage and management 11%, network and firewalls 8%, support and services 4%, rack, PDU and installation 4%, and site and contingency 6% [26]. The lines other than servers come to 33% (derived). The presentation gives an all-in equipment cost of $88,000 to $102,000 per GPU from GB300 to B300 [26]. The split applies to that B300 pod only.

7. Rack price against the rental rate

CoreWeave's public price list shows a GB200 NVL72 instance of four GPUs, built from two Grace Blackwell superchips, at $42.00 an hour on demand, with 30.72 TB of local storage [15]. That is $10.50 per GPU-hour, and 18 instances make 72 GPUs at $756 an hour, or about $6.62 million a year at the on-demand rate (arithmetic on [15]). The GB300 NVL72 line shows "Contact sales" [15].

A rack price and a rental rate do not compare like for like. Data Gravity estimated $240,000 to $410,000 a year to run an owned GB300 NVL72 rack, and a cost of $2.6 to $2.9 per utilized GPU-hour at 80% to 90% utilization [13]. That estimate covers a GB300 rack and CoreWeave's list price a GB200 instance.

8. Reading an NVL72 quote line by line

Line	What to ask	Sources that show it varies
Rack and build	GB200 or GB300, the maker's model, part counts per tray, drives fitted	[3, 4, 6]
Power shelves	Six or eight, and the redundancy	[6, 8]
Cooling	Whether an in-rack cooling unit is included, its capacity, what sits outside the rack	[6, 13]
Delivery	The Incoterm, the place, freight and insurance	[13, 26, 30]
Duty, tax and export	Duty, sales and use tax, licensing for the destination	[26, 31]
Network and storage	Top-of-rack switches, the scale-out fabric, storage	[6, 8]
Services	Rack integration, field deployment, deployment tier	[6, 7, 19]

Line	What to ask	Sources that show it varies
Warranty and support	Years, response time, years four and five	[4, 6, 19]
Spares	Trays held onsite	[6]
Software	NVIDIA AI Enterprise term, cluster manager, support	[27, 28, 29]
Validity and payment	How long the price holds, cancellation, invoicing	[19, 26]
Acceptance	The tests that define acceptance	[32]

Acceptance can be written into a contract. A Boost Run service agreement for GB300 NVL72 capacity, filed in October 2026, has Boost Run send a ready-for-testing notice once the racks are set up and configured [32]. Its validation report covers a review of conformance with NVIDIA's reference architecture for GB300 NVL72, single-node and multi-node NCCL all-reduce tests and single-node and multi-node LLM benchmark tests [32]. It is a cloud service agreement. Its tests still name what ready means for a rack.

Implications

For buyers checking a quote. Start from the part list and the delivery term, then the services and the software. A GB300 figure without its scope can sit anywhere from $3.7 million to $6.5 million in the dated public record, and HIVE's expected cluster spending per 72 GPUs higher still [11, 12, 23]. Ask which lines are excluded, in the way the B300 quotation in the Forum Markets filing listed them, and how long the price holds [26].

For finance teams and lenders. The compute trays carry most of the value, and software, warranty extensions, spares, network and installation arrive as separate lines [6, 8, 13, 19, 28]. At list price, five years of NVIDIA AI Enterprise for one rack comes to $1,296,000 (arithmetic on [27]). Payment tied to each shipment, as in Dell's terms with IREN, sets when cash is needed, and that order cannot be cancelled without Dell's written consent [17, 19].

For analysts and price series. Each figure needs its date, the rack, its scope and its evidence class. Indicative prices, quotes and cluster purchases describe different things, and averaging them describes no rack. Rillor's GPU systems pricing method models each configuration by GPU family, GPU count, memory, CPU, interconnect, cooling, condition, location and seller type, and each observation keeps its evidence class, so a listing is read as a listing and a quote as a quote.

Limits of this analysis

- No figure here is a verified transaction. The rack figures are indicative prices, a quote reported in an investor presentation and a newsletter's figures from purchase orders, adjusted for anonymity [10, 11, 12, 13, 26].
- Two figures rest on parties the reports do not name, the Tom's Hardware source and Forum Markets' "Vendor B" [12, 26]. The HSBC and Morgan Stanley figures passed through social media posts before the press reported them [10, 14].
- IREN's filed agreements redact quantities, values and the price per GPU-hour, so they show scope and terms and no unit price [19, 20].
- HIVE's two releases give 2,303 and 2,304 GPUs for the same GB200 cluster [24, 25]. The per-GPU figure moves by about $31 between them (derived).
- No source cited here gives freight, duty, tax or installation figures for an NVL72 rack, and Data Gravity's deployment line items appear here only as its total [13].
- No public rental price for the GB300 NVL72 and no price for a used NVL72 rack were found [15].
- NVIDIA's documents differ on BlueField-3 counts, and NVIDIA and Lenovo on rack power [3, 4, 5, 6]. Each is given as published.
- The export passage is restated as NVIDIA filed it [31]. This report describes what public sources show and is not legal, tax or investment advice.

Sources

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

2. NVIDIA GB200 NVL72. NVIDIA, viewed 8 November 2026. https://www.nvidia.com/en-us/data-center/gb200-nvl72/

3. Hardware, NVIDIA DGX GB Rack Scale Systems User Guide. NVIDIA, last updated 17 September 2026, viewed 8 November 2026. https://docs.nvidia.com/dgx/dgxgb200-user-guide/hardware.html

4. NVIDIA DGX GB300. NVIDIA, viewed 8 November 2026. https://www.nvidia.com/en-us/data-center/dgx-gb300/

5. NVIDIA DGX GB200. NVIDIA, viewed 8 November 2026. https://www.nvidia.com/en-us/data-center/dgx-gb200/

6. Lenovo NVIDIA GB300 NVL72 Rack Scale AI, Product Guide LP2357 (Marlon Aguirre and Ana Irimiea). Lenovo Press, updated 30 August 2026, viewed 8 November 2026. https://lenovopress.lenovo.com/lp2357-lenovo-nvidia-gb300-nvl72-rack-scale-ai

7. PowerEdge XE9712 Spec Sheet. Dell Technologies, file dated 9 April 2026, viewed 8 November 2026. https://www.delltechnologies.com/asset/en-us/products/servers/technical-support/poweredge-xe9712-spec-sheet.pdf

8. NVIDIA DGX SuperPOD: Next Generation Scalable Infrastructure for AI Leadership, Reference Architecture, DGX GB300. NVIDIA, file dated 7 November 2025, viewed 8 November 2026. https://docs.nvidia.com/pdf/dgx-spod-gb300-ra.pdf

9. How New GB300 NVL72 Features Provide Steady Power for AI (Rouslan Dimitrov). NVIDIA Technical Blog, 28 July 2025, viewed 8 November 2026. https://developer.nvidia.com/blog/how-new-gb300-nvl72-features-provide-steady-power-for-ai/

10. Nvidia's next-gen Blackwell AI Superchips could cost up to $70,000, fully-equipped server racks reportedly range up to $3,000,000 or more (Anton Shilov). Tom's Hardware, 14 May 2024, viewed 8 November 2026. https://www.tomshardware.com/pc-components/gpus/nvidias-next-gen-blackwell-ai-gpus-to-cost-up-to-dollar70000-fully-equipped-servers-range-up-to-dollar3000000-report

11. Apple is about to spend $1 billion on NVIDIA servers for AI: analyst (Ryan Christoffel). 9to5Mac, 25 March 2025, viewed 8 November 2026. https://9to5mac.com/2025/03/25/apple-is-about-to-spend-1-billion-on-nvidia-servers-for-ai-analyst/

12. Price of Nvidia's Vera Rubin NVL72 racks skyrockets to as much as $8.8 million apiece, but server makers' margins will be tight (Anton Shilov). Tom's Hardware, 24 March 2026, viewed 8 November 2026. https://www.tomshardware.com/tech-industry/artificial-intelligence/price-of-nvidias-vera-rubin-nvl72-racks-skyrockets-to-as-much-as-usd8-8-million-apiece-but-server-makers-margins-will-be-tight-nvidia-is-moving-closer-to-shipping-entire-full-scale-systems

13. How Much Does an NVIDIA NVL72 Cost? (Chris Zeoli). Data Gravity, 31 August 2026, viewed 8 November 2026. https://www.datagravity.dev/p/how-much-does-an-nvidia-nvl72-cost

14. Cooling system for a single Nvidia Blackwell Ultra NVL72 rack costs a staggering $50,000, set to increase to $56,000 with next-generation NVL144 racks (Anton Shilov). Tom's Hardware, 5 November 2025, viewed 8 November 2026. https://www.tomshardware.com/pc-components/cooling/cooling-system-for-a-single-nvidia-blackwell-ultra-nvl72-rack-

costs-a-staggering-usd50-000-set-to-increase-to-usd56-000-with-next-generation-nvl144-racks

15. Pricing. CoreWeave, viewed 8 November 2026. https://www.coreweave.com/pricing

16. IREN Secures $9.7bn AI Cloud Contract with Microsoft (Exhibit 99.1 to Form 8-K). IREN Limited, 3 November 2025, viewed 8 November 2026. https://www.sec.gov/Archives/edgar/data/1878848/000114036125040072/ef20058139_ex99-1.htm

17. IREN Limited, Current Report on Form 8-K (Commercial Agreement with Dell). U.S. Securities and Exchange Commission, filed 3 November 2025, viewed 8 November 2026. https://www.sec.gov/Archives/edgar/data/1878848/000114036125040072/ef20058139_8k.htm

18. IREN investor presentation (Exhibit 99.2 to Form 8-K). IREN Limited, 3 November 2025, viewed 8 November 2026. https://www.sec.gov/Archives/edgar/data/1878848/000114036125040072/ef20058139_ex99-2.htm

19. Non-Cancellable Non-Returnable (NCNR) and Supplemental Terms Agreement, IE US Hardware 3 Inc. and Dell Marketing L.P. (Exhibit 10.5 to IREN Form 10-Q). U.S. Securities and Exchange Commission, filed 5 February 2026, viewed 8 November 2026. https://www.sec.gov/Archives/edgar/data/1878848/000187884826000015/exhibit105-nonxcancellab.htm

20. Partner Statement of Work, IE US Hardware 3 Inc. and Microsoft Corporation (Exhibit 10.3 to IREN Form 10-Q). U.S. Securities and Exchange Commission, filed 5 February 2026, viewed 8 November 2026. https://www.sec.gov/Archives/edgar/data/1878848/000187884826000015/exhibit103-partnerstatem.htm

21. IREN Limited, Annual Report on Form 10-K for the fiscal year ended June 30, 2026. U.S. Securities and Exchange Commission, filed 27 August 2026, viewed 8 November 2026. https://www.sec.gov/Archives/edgar/data/1878848/000187884826000052/iren-20260630.htm

22. IREN Delivers Horizon 1 to Microsoft and Achieves NVIDIA Exemplar Cloud Status on GB300 NVL72 (Exhibit 99.1 to Form 8-K). IREN Limited, 13 August 2026, viewed 8 November 2026. https://www.sec.gov/Archives/edgar/data/1878848/000114036126032638/ef20080141_ex99-1.htm

23. HIVE's BUZZ HPC Signs $350 Million AI Cloud Services Agreement with Investment-Grade Enterprise Customer (Exhibit 99.1 to Form 8-K). HIVE Digital Technologies Ltd., 17 August 2026, filed 19 August 2026, viewed 8 November 2026. https://www.sec.gov/Archives/edgar/data/1720424/000106299326004522/exhibit99-1.htm

24. HIVE's BUZZ HPC Secures Approximately $129 Million at 9.0% Fixed-Rate Financing at 80% Loan-to-Cost to Expand Sovereign AI Infrastructure (Exhibit 99.1 to Form 8-K). HIVE Digital Technologies Ltd., 27 October 2026, filed 29 October 2026, viewed 8 November 2026. https://www.sec.gov/Archives/edgar/data/1720424/000106299326005585/exhibit99-1.htm

25. HIVE's BUZZ HPC Closes USD $220 Million Sovereign AI GPU Contract with Bell AI Fabric (Exhibit 99.1 to Form 8-K). HIVE Digital Technologies Ltd., 18 June 2026, filed 22 June 2026, viewed 8 November 2026. https://www.sec.gov/Archives/edgar/data/1720424/000106299326003293/exhibit99-1.htm

26. Forum Edge AI: Inference Infrastructure Opportunity, investor presentation (Exhibit 99.2 to Form 8-K). Forum Markets, Inc., October 2026, filed 5 October 2026, viewed 8 November 2026. https://www.sec.gov/Archives/edgar/data/1690080/000121390026106655/ea030608001ex99-2.htm

27. NVIDIA AI Enterprise Pricing, NVIDIA Enterprise Licensing Guide. NVIDIA, last updated 2 September 2026, viewed 8 November 2026. https://docs.nvidia.com/ai-enterprise/planning-resource/licensing-guide/latest/pricing.html

28. NVIDIA AI Enterprise Licensing, NVIDIA Enterprise Licensing Guide. NVIDIA, last updated 2 September 2026, viewed 8 November 2026. https://docs.nvidia.com/ai-enterprise/planning-resource/licensing-guide/latest/licensing.html

29. DGX Resources. NVIDIA, viewed 8 November 2026. https://docs.nvidia.com/dgx-resources/index.html

30. Know Your Incoterms. International Trade Administration, U.S. Department of Commerce, viewed 8 November 2026. https://www.trade.gov/know-your-incoterms

31. NVIDIA Corporation, Quarterly Report on Form 10-Q for the quarter ended July 26, 2026. U.S. Securities and Exchange Commission, filed 26 August 2026, viewed 8 November 2026. https://www.sec.gov/Archives/edgar/data/1045810/000104581026000075/nvda-20260726.htm

32. Boost Run LLC Service Agreement, Version 1.6.4 (Exhibit 10.1 to Form 8-K). Boost Run Inc., filed 6 October 2026, viewed 8 November 2026. https://www.sec.gov/Archives/edgar/data/2090646/000149315226045861/ex10-1.htm

33. Nvidia's latest AI chip will cost more than $30,000, CEO says (Kif Leswing). CNBC, 19 March 2024, viewed 8 November 2026. https://www.cnbc.com/2024/03/19/nvidias-blackwell-ai-chip-will-cost-more-than-30000-ceo-says.html

34. HIVE Digital Technologies Ltd., Quarterly Report on Form 10-Q for the quarter ended September 30, 2026. U.S. Securities and Exchange Commission, filed 9 November 2026, viewed 14 November 2026. https://www.sec.gov/Archives/edgar/data/1720424/000106299326005842/form10q.htm

About this report

Published by Rillor on 15 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. NVL72 rack prices: what a GB200 or GB300 rack quote covers. 15 November 2026. https://rillor.com/insights/nvl72-rack-prices

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.

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