SECOND EDITION · FULL RE-VERIFICATION · 7 NOVEMBER 2026

NVIDIA HGX B300 in Asia: sellers, stock, pricing, export controls and tax

A re-analysis from scratch across 23 Asian markets and the GCC — every market re-researched independently, every claim traced to a source, every seller contact re-read from the vendor's own page.

What changed from the first edition. Every market was re-researched from primary sources rather than carried forward, and every load-bearing number was re-checked against the issuer's own document — NVIDIA's spec tables, BIS and the Federal Register, DGFT and MOTIE notices, exchange filings and vendor price pages. That produced **five corrections to my own earlier advice** (Indonesia duty, Malaysia permit scope, UAE country grouping, NVIDIA's own memory spec, and the Nebius price table), **two specification traps the first edition missed entirely** — INT8 falls to 3 POPS and FP64 to 10 TFLOPS on B300 — plus **roughly 90 named sellers with contact details** across 23 markets, **18 confirmed B300/GB300 deployments** including the largest orders in the world, and a materially different picture of *which product is actually being bought* — it is overwhelmingly the GB300 NVL72 rack, not the air-cooled 8-GPU baseboard. A read-back of the finished draft also surfaced **five internal inconsistencies in my own numbers** — KONST's price unit, the §30 utilisation arithmetic, the node price basis in the decision table, the Korea/India label on the national GPU programme, and the confidence level of the hardware anchors — all corrected here. Two sections were added after the second edition was first assembled: **§33, the used and secondary-market channel**, and **§34, the data-centre operators who can actually host a B300**.

NAMED SELLERS WITH CONTACTS

~90

23 markets, NPN / OEM / distributor / SI / neocloud

CONFIRMED B300/GB300 DEALS

18

Riyadh, Taipei, Chennai, Tokyo, Seoul, Jakarta

CHEAPEST LIVE ASIA ON-DEMAND

$5.00

Gata Cloud, Singapore — per GPU-hour

BIGGEST SINGLE ORDER FOUND

US$521M

10,000 B300, DataSection × Inventec (Sydney)

CONTENTS

1 · The product, decoded

2 · What changed vs B200

3 · Performance: claims vs evidence

4 · Cost per token & unit economics

5 · Supply, allocation & lifecycle

6 · The deployments (who bought what)
7 · Seller directory — Singapore
8 · Seller directory — Malaysia
9 · Seller directory — Thailand
10 · Seller directory — Indonesia
11 · Seller directory — Vietnam
12 · Seller directory — Philippines
13 · Brunei, Cambodia, Myanmar
14 · Seller directory — Japan
15 · Seller directory — South Korea
16 · Seller directory — Taiwan
17 · Seller directory — India
18 · Pakistan, Bangladesh, Sri Lanka
19 · Seller directory — UAE
20 · Seller directory — Saudi Arabia
21 · Qatar, Kuwait, Oman, Bahrain
22 · Israel
23 · Turkey
24 · China, Hong Kong, Macau
25 · GPU cloud price board
26 · Hardware price anchors
27 · Export controls & import permits
28 · Tax, duty & landed cost
29 · Data-centre siting & facility economics
30 · Verdict: buy, rent, or wait
31 · RFQ checklist
32 · What is unreliable
33 · Used, refurbished & secondary market
34 · Where to put it — data-centre operators

1 · The product, decoded

HGX B300 is a **baseboard**, not a server and not a rack. NVIDIA builds the 8-GPU board; OEMs (Dell, HPE, Lenovo, Supermicro, GIGABYTE, ASUS, Pegatron, Foxconn, Quanta, Wiwynn) wrap it in a chassis with CPUs, NICs and cooling. NVIDIA sells no baseboards directly — every legal quote in Asia comes from an OEM, an NPN partner, a distributor or an integrator.

SPECIFICATION	HGX B300 (8-GPU NODE)	GB300 NVL72 (RACK)	B200 / H200
CPU	x86 or OEM-supplied (DGX B300: 2× Intel Xeon 6776P)	36 × NVIDIA Grace (Arm Neoverse V2, 2,592 cores; 17 TB LPDDR5X)	—
GPU memory	**2.1 TB** (NVIDIA spec — 270 GB per GPU, 262 GB usable)	20 TB GPU memory + **37 TB fast memory**	180 GB / 141 GB per GPU
Memory bandwidth	8 TB/s per GPU (Blackwell Ultra HBM3e)	up to 576 TB/s	8 TB/s / 4.8 TB/s
FP4 Tensor Core	**144 PFLOPS sparse / 108 dense**	1,440 / 1,080 PFLOPS	B200: 72 dense
FP8 / FP6 Tensor Core	72 PFLOPS dense	720 PFLOPS dense	B200: 72 PFLOPS
INT8 Tensor Core	**3 POPS**	24 POPS	**B200: 72 POPS**
FP64 / FP64 Tensor Core	**10 TFLOPS**	100 TFLOPS	**B200: 296 TFLOPS**
Scale-up domain	**8 GPUs**, 14.4 TB/s NVLink	**72 GPUs**, 130 TB/s	8 GPUs, NVLink 5
Power	up to 1,400 W/GPU; **~14 kW/node** (DGX B300 spec)	fully liquid-cooled rack, ≈142 kW	1,200 W / 700 W
Networking	8× OSFP serving 8× ConnectX-8 VPI at 800 Gb/s + 2× dual-port BlueField-3 DPU at 400 Gb/s; 1.6 TB/s	ConnectX-8 SuperNIC, 800 Gb/s per GPU	ConnectX-7

Five spec traps, all live in 2026. (1) NVIDIA's blog says **15 PFLOPS dense NVFP4 per GPU**, but the HGX product page's **108 PFLOPS for 8 GPUs** implies **13.5** — the HGX B300 bin is de-rated ~10% versus the GB300 rack part; plan on 13.5. (2) NVIDIA's own HGX table gives **2.1 TB total** (270 GB × 8, 262 GB usable) — but Supermicro's AS-8126GS-NB3RT listing, KONST's cluster spec and several neocloud price cards publish **288 GB per GPU**. That is a *listing* discrepancy, not a second NVIDIA part: 2.1 TB is the number to design capacity against, and 288 GB is what sellers are quoting. (3) **INT8 collapses to 3 POPS on B300, from 72 POPS on B200 — a 24× cut.** Any INT8-quantised pipeline, most legacy inference stacks and a lot of enterprise video and speech models will run slower on a B300 than on the B200 it replaced. (4) **FP64 falls to 10 TFLOPS from 296 TFLOPS** — a 30× cut. A B300 is not a GPU for FP64/HPC work; if your model is scientific, seismic, CFD or molecular, the B300 is the wrong purchase and no amount of NVLink fixes it. (5) The **14 kW per node** figure is NVIDIA's own DGX B300 specification — but it is a DC-input figure for one configuration. NVIDIA's facility reference for 16 DGX B300 racks gives 820 kVA average / 1,228 kVA peak, which is the number to design electrical infrastructure against, not the nameplate.

> **The de-rates are the buying decision for two whole buyer classes.** If you are an HPC, scientific-computing or FP64 buyer, the B300 is a downgrade on your actual workload and a roadmap risk on everything else. If you run INT8-quantised inference — still common in production video, ASR and recommendation systems — check the INT8 path explicitly in an RFQ, because the FP4 headline improvement does not carry. The workloads the B300 genuinely wins are FP4/BF16 dense inference with large context, and training at FP4 precision.

2 · What actually changed versus B200

- **Capacity, not bandwidth.** 270 GB vs 180 GB (+50%) at the same 8 TB/s — and 2.1 TB vs 1.4 TB per node. A KV-cache and batch-size play for long-context and reasoning workloads — not a training-FLOPs upgrade.
- **Two real de-rates, both confirmed on NVIDIA's own spec table.** INT8 Tensor Core falls from **72 POPS (B200) to 3 POPS (B300)**, and FP64 from **296 TFLOPS to 10 TFLOPS**. FP8/FP6 is flat at 72 PFLOPS. The Blackwell Ultra generation bought NVFP4 density and attention throughput and paid for it in the legacy numeric formats.
- **+50% dense NVFP4** and **2× attention-layer throughput** from doubled SFU capacity. Test-time scaling is dominated by attention on long outputs, which is what the extra throughput targets.
- **+17% power** (1,200 → 1,400 W per GPU). Eight of these in a chassis is a liquid-cooling project, not a server refresh — the reason nearly every 2026 Asian deployment is a new-build rather than a retrofit.
- **Vendor headlines:** 11× LLM inference and 4× memory vs Hopper for B300 NVL16; 50× AI-factory output and 30× real-time video generation for GB300 NVL72. Projected, not reproduced.

3 · Performance: claims versus evidence

NVIDIA's marketing numbers are *rack-scale*. Independent benchmarks are mostly *node-scale*. That gap is the single most important thing to understand before signing anything.

EVIDENCE	RESULT	STATUS
MLPerf Inference v6.0 — NVIDIA platform	GB300 NVL72: DeepSeek-R1 server 2,907 → **8,064 tok/s/GPU** (2.77×) vs v5.1; Llama 3.1 405B server 170 → 259	V
MLPerf Inference v6.0 — **no published node-level 8-GPU B300 submission was located**	MLPerf v6.0 does list Blackwell Ultra systems and NVIDIA's DGX B300 page cites v6.0 results, but no node-scale (8-GPU) B300 cost-per-token result was found. All node-level cost-per-token data here comes from InferenceX, not MLPerf	U
vLLM, 13 Feb 2026	GB300, DeepSeek-V3.2 FP4: 7,360 tok/GPU/s prefill-only; 2,816 tok/GPU/s at ISL 2k / OSL 1k	V
SGLang/LMSYS, 20 Feb 2026	GB300 NVL72 up to **25× H200** on DeepSeek-R1 at 50 tok/s/user	V
AMD MI355X, MLPerf v6.0 Open	Llama 2 70B **100,282 tok/s**, 3.1× MI325X; multi-node >1M tok/s at 97–98% scaling efficiency	V
InferenceX, MiniMax M3 428B @ 48 tok/s/user	B300 **$0.19/M** vs GB300 NVL72 $0.29/M — the **8-GPU node is 57% cheaper**	M
InferenceX, GLM 5	B300 $0.25/M vs GB200 NVL72 $0.08 — GB200 ~3× cheaper	M

EVIDENCE	RESULT	STATUS
NVIDIA-published economics	GB300 NVL72 $0.123/M @ 116 tok/s/user; DGX B300 $0.24/M @ 102 tok/s/user; H200 $4.20/M → GB300 $0.12/M (35× lower)	V vendor

> **The planning error to avoid.** Per-token cost is not a property of the GPU; it is a property of the deployment shape. On sparse-attention and wide expert-parallel serving, the 72-GPU NVLink domain can be 2–3× cheaper per token — *but only if it is tuned correctly* (see next section). NVIDIA's headline $0.12/M, 35× and 65× figures are rack-level GB300 NVL72 versus HGX H200, re-published by NVIDIA; no published *node-level* MLPerf submission for an 8-GPU B300 was located. **And every one of these numbers is FP4.** Your workload's cost per token depends on the precision it runs in — and on B300 that choice is a much bigger decision than it was on B200, because INT8 is 24× slower and FP64 is 30× slower than the part it replaced.

4 · Cost per token & unit economics

An independent TCO study (userecurrence.ai, 28 Oct 2026, updated 2 Nov) modelled a DeepSeek V4.1 Flash workload at 100 output tok/s/user against a market price of $0.019/M blended:

CONFIGURATION	THROUGHPUT (TOK/S)	RENT $/M TOKENS AT FULL LOAD	BREAK-EVEN UTILISATION	GROSS MARGIN @50% / @70%
H200 8× node	24,124	$0.0459 (2.4× market)	244%	negative
B300 8-GPU node	315,201	**$0.0065 (35%)**	**35%**	31% / 51%
GB300 NVL72, standard config	267,304	$0.0158 (84%)	84%	−68% / −20%
GB300 NVL72 + Dynamo	471,578	$0.0090 (48%)	48%	5% / 32%
GB300 NVL72, vLLM rack-tuned	954,000	**$0.0044 (24%)**	24%	53% / 66%

> **The sharpest finding in the study.** A rack-tuned GB300 NVL72 ($0.0044/M) is ~1.5× cheaper per token than the B300 8-GPU node ($0.0065/M) — **but an untuned GB300 is 2.4× worse** than the same B300 node. If you buy the rack, you must buy the Wide-EP / prefill-decode disaggregation tuning as well, or you will pay more for worse throughput. The B300 node is the safer buy precisely because it does not depend on rack-level software discipline. Also note the break-even utilisations: a standalone 8-GPU B300 node needs **35% sustained utilisation** to cover its rent, a standard GB300 rack needs **84%** — and most of Asia's announced capacity is being bought by operators who still have to find that demand.

Where the money actually goes

- **Power is not the swing factor.** ~14 kW ≈ $1.1–1.4/hour at $0.08–0.10/kWh. Utilisation and financing are.
- **Facility economics diverge sharply by market.** Wholesale colocation ranges from **$188/kW-month in Sydney to $403 in Singapore** (CBRE 2026); electricity from $0.063/kWh in Indonesia to **$0.270 in Singapore**. Singapore is simultaneously the easiest place to import and the most expensive place to switch on.

- **Rack density is the real constraint.** A DGX B300 SuperPOD reference is 4 systems per rack ≈ **56 kW/rack**; a GB300 NVL72 rack is **≈142 kW**. The industry average rack is 8 kW and average PUE is 1.56. Existing Asian colo floors are not grandfathered into this class.

5 · Supply, allocation & lifecycle

SIGNAL	FIGURE
NVIDIA Q2 FY2027 (ended 26 Jul 2026)	Revenue **$96.221bn** (+106% YoY, +18% QoQ); Data Center **$89.0bn** (+117% YoY); hyperscale DC $49bn (+13% QoQ); ACIE $40bn (+25% QoQ, +138% YoY)
Vera Rubin share	**≈20% of Data Center revenue in Q3 FY2027**; production shipments began Aug 2026; POs from every major hyperscaler, neocloud and OEM; Rubin quoted at 30× throughput/MW and 35× lower token cost vs Grace Blackwell Ultra
Supply guidance	FY2028 ≈ +70% YoY; **"supply to remain a bottleneck at least through the end of fiscal year 2028"**; neocloud partners to exit 2026 with ~8 GW installed vs ~3 GW at end-2025
Sovereign AI	+35% QoQ, >3× YoY. APAC neoclouds named by NVIDIA on the call: **GMI Cloud (Taiwan), Yotta and Neysa (India), Firmus (Australia), YTL AI Cloud (Malaysia)**; Japan: Noetra DSX; Korea: LG, Hyundai
Older-accelerator pricing	H100 on-demand fell to **$1.80–3.50/GPU-hr** in Q2 2026 (down 60–77% from the ~$8 peak); 1-year reserved H100 bottomed $1.70 (Oct 2025) then **rose to $2.35 by Mar 2026 (+40%)** — demand recovered faster than capacity
Price direction at the front	Nebius **raised** on-demand prices effective 1 Oct 2026: B300 $7.85 → **$9.50 (+21%)**, B200 $7.15 → **$8.50**, H100 $3.85 → $4.50, H200 $4.50 → $5.40. 90-day median on-demand across 38 providers: **$9.50, +14%**
Lead times	New data-centre GPU capacity **36–52 weeks** (mid-2026); HBM3e 2026 contract prices +~20%; TSMC CoWoS fully allocated through mid-2027; B300 systems 8–14 weeks; GB300 4–8 weeks where locally stocked
B300 revenue share	**Not separately disclosed by NVIDIA.** Blackwell Ultra's contribution sits inside the "sustained strength in Blackwell" hyperscale datapoint

> **Net read for a buying decision.** Two forces pull in opposite directions and the market is currently pricing the first one: genuine scarcity (a bottleneck through FY2028) and a generational handover (Rubin already at ~20% of data-centre revenue, racks live at CoreWeave, Google Cloud, Azure, OCI and Nebius). The observed behaviour is **price rises, not price cuts** — Nebius raising B300 21% in October is the clearest signal. A 6–12 month commitment with a price-review clause is defensible; a 24–36 month flat-rate lock-in is a bet against NVIDIA's own roadmap, and against the H100 precedent of a 60–77% decline followed by a 40% recovery in reserved pricing.

6 · The deployments — who has actually bought what

Eighteen confirmed B300/GB300 commitments across Asia and the near neighbourhood, ordered by scale. This table is the best available proxy for who in Asia can actually source at volume today.

BUYER / PROGRAMME	COUNTRY	SCALE	STATUS	DETAIL
HUMAIN + NVIDIA	Saudi Arabia	**84 MW / ~27,000 GPUs.** Phase 1: 14 MW ≈ **6,000 HGX B300** on Quantum-X800; 70 MW Vera Rubin from Jun 2027	Q4 2026	50 MW to a US client, 20 MW to HUMAIN Cloud Jul/Aug 2027. Earlier phase: up to 18,000 GB300
G42 / Stargate UAE	UAE	**35,000 GB300 chips** authorised; 1 GW campus, first 200 MW	2026	UAE government agencies plus G42 and Core42 named in 15 CFR 740 Supp. 8 as Approved Entities — their licence-free status **expires after 270 days unless they become "U.S. companies" or obtain a BIS Advisory Opinion** (~6 Apr 2027)
Foxconn / Big Innovation	Taiwan	**US$1.4bn**, 10,000 Blackwell GPUs, GB300 NVL72 hybrid cooling	Live	Plus Visionbay.ai sovereign AI centre on GB300 NVL72, online 1H 2027
Yotta (IndiaAI Mission)	India	**20,736 Blackwell Ultra**, >10,000 of them B300	Live Aug 2026	60 MW D2 Noida → 250 MW; Navi Mumbai → 2 GW target. MeitY-empanelled
L&T / Vyoma.AI → Together AI	India	**10,000 B300** on a single NVLink fabric, Chennai	Phase 1 250 MW	Deal value **up to US$1.6bn**; 150 MVA
Ooredoo → Zankore	Indonesia	**~200 MW GB300 NVL72**; 1 GW DSX target	H1 2027	Ooredoo takes a **49% stake (~US$800M)**; US$3.1bn senior term loan with Citi as exclusive debt adviser
MSIT AI Highway	South Korea	KRW **20.8tn**; 9,704 GPUs = **2,016 Vera Rubin + 7,688 B300**	POs placed Jun 2026	Naver Cloud 1,008 VR + 3,112 B300; Samsung SDS 1,008 VR + 2,016 B300; Elice Group 2,560 B300
DataSection × Inventec	Australia (APAC)	**1,250 servers = 10,000 B300, US$521M**	Delivered Dec 2025–Feb 2026	Executed 11 Dec 2025, JPY 81,451m at 156.15. For a Sydney data centre — **the largest single B300 order anywhere, and Japanese-owned**, though the destination is not Asia
DataSection × Compal	Japan	**635 B300 servers = 5,080 GPUs, US$325M**	Ops May–Jul 2026	Executed 7 May 2026, JPY 50,935m at 156.57. US$252M unpaid at signing. AI DC in Inzai, Chiba. Contact Yuichi Nozawa +81-50-3649-4858

BUYER / PROGRAMME	COUNTRY	SCALE	STATUS	DETAIL
XMax / DSH HYPER on kt cloud	South Korea	64 nodes = **512 B300 + 560 TB**	Live 29 Oct 2026 (60 nodes / 480 GPUs)	5-year agreement signed 28 Oct 2026 with Singapore-based DSH HYPER; capacity sourced from Korea's kt cloud. **~US$122.5M scheduled service charges**; >US$100M revenue over five years; contingent on board approval and SBLC by 17 Nov 2026
Unity Group / Multipolar	Indonesia	**1,024 HGX B300 8-GPU units**, 3-year lease	HKEX-filed Mar 2026	**US$123,789,312** = US$3,438,592/month, 99.5% availability. Sourced from an NVIDIA Cloud Partner
YCO Cloud × Aolani	Philippines	**>10,000 Blackwell Ultra / GB300 NVL72**	Phase 1 Q1 2027	Dedicated capacity from 2,000 GPUs; among the first GB300 NVL72 rollouts in the Philippines. enquiries@aolani.com
Bitdeer AI	Malaysia	A102 9.5 MW liquid-cooled GB300 NVL72 fully committed; A201 21.7 MW >70% contracted; A901 67 MW campus engineered for GB300	Q1 2027 / Q2 2027	>$800m expected 5-yr revenue on A102; >$1.7bn on A201; A103 60 MW Q1 2028
KONST Group	Malaysia	**128 nodes × 8 = 1,024 B300 GPUs** at Cyberjaya (with ADATA; 800G InfiniBand, direct-to-chip liquid cooling, ~294.9 TB)	Live 2027 Q1	From **$6.00/GPU-hr** on a 5-year term, billed monthly. Total 3,856 GPUs across B300/B200/H100/H200/L40S. Plus 15 MW liquid-cooled lease at SMX01 Jakarta, move-in 26 Dec 2026
GMO Internet	Japan	**¥26.1bn, up to 116 B300 servers**; Q2 material adds 64 B300 / ¥10.1bn	Service Aug 2027	New DC in Ishikari, Hokkaido. B300 in production since 25 Mar 2026
Indosat ZanKore	Indonesia	**~200 MW GB300 NVL72**; 1 GW DSX target	H1 2027	DSX MaxLPS, +40% compute. Also 5.5 MW NeutralDC BTM-1 Batam with B300 from Sept 2026
Giga Computing GAIFA	Taiwan	700 kW: 2 × GB300 NVL72 racks + 1 HGX B300 rack	Announced 22 Sep 2026	New Taipei. Also DL83-GP6 liquid-cooled 8 × B300
SuperX → Woodman	Japan	~US$38.8m B300 order	Delivery mid-Nov 2026	Signed ~25 Aug 2026, ~20% advance paid. Yokohama

> **The structural finding.** Across all 18 confirmed deployments, **the unit actually being ordered is the GB300 NVL72 rack** — 72 GPUs, liquid-cooled, Quantum-X800 fabric. The air-cooled 8-GPU HGX B300 baseboard is what *capacity resellers and integrators* buy, and it dominates only where the buyer wants a small deployable unit: India's 10,000-GPU build, Taiwan's local rack, and the SEA reseller channel. If your use case is frontier training or wide expert-parallel serving, the market is telling you that a rack is the relevant unit of purchase. If your use case is a single tenant with 8–64 GPUs of long-context inference, the node is still the right product — but that is now the minority transaction.

Secondary and partial deployments

- **South Korea — National GPU programme (KAIT / NIPA):** ~10,000 GPUs; self-pay **H200 ₩400,000/GPU-month, B200 ₩800,000/GPU-month**; 50% subsidy for youth firms, free for academia.
- **Israel — doubleAI (Tel Aviv):** ~1,000 NVIDIA B300 bought, roughly half of a $220m raise (≈$110m) — U secondary sourcing.
- **Israel — Anan × Crusoe (Afula):** $1bn deal, 40 MW server farm to 100 MW, ~12,000 NVIDIA processors planned — chip type unconfirmed.
- **Israel — Nebius (Modi'in):** 4,000 HGX **B200** national supercomputer; B300 is *not* in the me-west1 region.
- **Thailand — Datasection:** US$235.2m high-performance GPU server infrastructure, Bangkok + Pathum Thani. **Thai AI Cloud (TAC):** 5 HGX B300 nodes (40 GPUs) in Bangkok — 19 of the 40 GPUs still unallocated.
- **Thailand — BOI:** seven DC approvals >USD 3.1bn (True Internet DC USD 1.4bn/223 MW; GSA DC 05 USD 1.2bn/120 MW; Stellar DC USD 0.3bn/25 MW; Freyr USD 0.2bn).
- **Vietnam — FPT AI Factory:** first HGX B300 in Vietnam (30 May / 2–5 Jun 2026). **Viettel IDC:** Vietnam's first DGX B200 supercomputer, 6 Feb 2026 (1.5 ExaFLOPs FP8).
- **Singapore — Singtel RE:AI:** first APAC DGX GB300 deployment, initial capacity fully committed ahead of deployment (28 Oct 2026).
- **Indonesia — RAIA Grid / IFT:** local B300 *assembly plant*, ~US$22m capex, 140–150 units, first production Dec 2026.
- **Japan — NTTPC + Getnets + Fixsters:** water-cooled GPU container DC (HPE ProLiant XD685, HGX B300), from ¥300M capex, ¥750k/month opex, **minimum 8-month lead time**.
- **Israel — AI Research Compute Center:** 300 B200 GPUs at 80% subsidy, **$0.70 + VAT per GPU-hour** including 5 TB storage; 1,000 B200 overall, 70% industry / 30% academia.

7 · Seller directory — Singapore

Asia's most established entry point: 0% duty, 9% GST, English contracts, and the only market with a **full NVIDIA-authored DGX partner list carrying direct contacts**. Note that NVIDIA's *ja-jp* and *ko-kr* "where to buy" pages fall back to the global list — Singapore, Taiwan and Israel are the only markets with genuine country-specific rosters.

COMPANY	ROLE	CONTACT	AI COMPUTE OFFER	NOTES
Jardine OneSolution (2001) Pte Ltd	NPN Elite Solution Provider	+65 6551 9611 enquiry@jos.com.sg	DGX/HGX systems, deployment	On NVIDIA's where-to-buy page
NovaGlobal Pte Ltd	NPN Elite Solution Provider	+65 9615 5098 / +65 9721 1968 sales@novaglobal.com.sg	DGX/HGX, GPU servers	On NVIDIA's where-to-buy page
Symmatrix Pte Ltd	NPN Elite / systems reseller	+65 6776 6786 sales@symmatrix.com	DGX, GPU servers	On NVIDIA's where-to-buy page
Fujitsu Asia Pte Ltd	NPN / OEM-integrator	+65 6512 7555 info@sg.fujitsu.com	DGX/HGX, AI infrastructure	On NVIDIA's where-to-buy page
Tech Data Advanced Solutions (Singapore)	NVIDIA Marketplace **Distributor** — Compute, DGX Cloud, DGX AI Compute Systems, NVIDIA Enterprise Software	No public contact found — marketplace profile	DGX systems and DGX Cloud	Competencies include DGX AI Compute Systems, so **can quote HGX/DGX**

COMPANY	ROLE	CONTACT	AI COMPUTE OFFER	NOTES
Freyr Technology AI Pte Ltd	NVIDIA Cloud Partner (Preferred) — Compute, NVIDIA AI	No public contact found (contact page 404s) — freyr.ai	GB300 / B300 capacity	Its Thailand BOI project is the live build
Core42 (G42)	Sovereign AI cloud	+971 2 699 2700 / +966 112 486 551 aicloud.core42.ai	GB300 "On Request"; B200 from $5.00/hr; H200 from $4.00; H100 from $2.50	UAE-origin with Saudi and Qatar presence; see §19
SuperX AI Technology (NASDAQ: SUPX)	Neocloud / integrator	superx.sg enquiry form	HGX B300 8-GPU (270 GB vRAM, 4096 GB RAM, 61.44 TB) at **$35.84/spot-hr for the 8-GPU instance** (≈$4.48/GPU-hr) — **SOLD OUT**; GB300 NVL72 $28.00/GPU-hr	Singapore HQ; Japan subsidiary in Yokkaichi. Also a US$38.8m B300 order from WOODMAN Inc. for Yokohama, and a 128-unit B300 order from Ezisight for Australia
Gata Cloud	Neocloud	No public contact found (Telegram enquiry only)	B300 288 GB; 100+ B300 nodes across APAC; pins Vietnam / HCM City / Indonesia — **$5.00/GPU-hr on-demand**	Cheapest live Asia on-demand found
Aolani	NVIDIA Cloud Partner	enquiries@aolani.com 391B Orchard Road #22-01, Ngee Ann City Tower B, Singapore 238874	>10,000 Blackwell Ultra / GB300 NVL72 via YCO Cloud (Philippines)	Announcement 28–29 Sep 2026 U

> **Singapore-specific rules.** Customs duty 0% on IT hardware; GST 9%, recoverable if GST-registered. No AI-chip import control. But the data-centre gate is now real: **DC-CFA2** requires ≥200 MW, PUE ≤1.25 and ≥50% green energy for new large facilities (METI confirmed the green-energy requirement on 7 Oct 2026), and the Digital Infrastructure Bill has passed Parliament with PUE and water-efficiency standards. At ~14 kW per node, a 1,024-GPU cluster is 13.4 MW of IT load before PUE — right at the threshold where DC-CFA2 starts to bind.

8 · Seller directory — Malaysia

COMPANY	ROLE	CONTACT	AI COMPUTE OFFER	NOTES
VSTECS Berhad	Distributor (NPN)	pr@vstecs.com	Compute and DGX distribution	Regional hub for TH/MY/SG/ID
Glocomp Systems (M) Sdn Bhd	NVIDIA Marketplace Distributor	+60 3 7652 1188 glocomp.com/contact-us	AI infrastructure distribution	Petaling Jaya

COMPANY	ROLE	CONTACT	AI COMPUTE OFFER	NOTES
MTC Communications (M) Sdn Bhd	GPU cloud / SI	sales@mtcc.my 4F Menara Lien Hoe, No.8 Persiaran Tropicana, 47410 Petaling Jaya	**B300 8-GPU preset, live — RM120/GPU-hr** (RM86,400/month at 720 h); 3/6/12-month discounts of 5/10/15%	RM120 ≈ US$29/GPU-hr — a **6× outlier** versus Singapore
KONST Group	Neocloud / colo	Site enquiry — konst.com.my	**1,024 B300 GPUs at Cyberjaya** (128 nodes, 288 GB, 800G InfiniBand, direct-to-chip liquid cooling); from **$6.00/GPU-hr**, 5-year term, billed monthly	Live 2027 Q1. Also SMX01 Jakarta 15 MW liquid-cooled lease, move-in 26 Dec 2026; Taiwan New Taipei 4 nodes pre-sale
PT Synnex Metrodata Indonesia	Distributor (MY-listed, ID ops)	+62 21 29345 800 Info.Metrodata@metrodata.co.id	Distribution, incl. NVIDIA	APL Tower 42F, Jl Letjen S Parman Kav 28, Jakarta
Bitdeer AI	AI data-centre operator	Site enquiry	A102 9.5 MW liquid-cooled GB300 NVL72; A201 21.7 MW; A901 67 MW campus; A103 60 MW	See §6
YTL AI Cloud	Neocloud (named by NVIDIA on its call)	Site enquiry	GB200 at RM68.80/GPU-hr (+ RM0.60 object storage)	GB200 not B300 — but a live Malaysian Blackwell operator

> **Malaysia — the permit position, corrected.** MITI's Strategic Trade Controller Directive No. 1/2025 (in force 14 July 2025, under STA 2010 §12) governs **export, transit, transshipment and brokering** — it does *not* cover plain domestic import, and MITI stated in June 2025 that servers with Nvidia chips are "not classified as controlled goods". **But enforcement is real and escalating:** customs seized **72 servers (RM52.92m) on 5 June 2026** and **30 servers (RM49.5m / US$12.11m) on 15 October 2026** at KLIA. Where the export, transit or broker path is engaged, the AICA form goes to **admin.sts@miti.gov.my at least 30 days in advance**; permits are single-use and valid 6 months; broker registration runs 1 year; records are retained 6 years; penalties reach 10 years and RM10m for individuals / RM20m for corporates. Controlled codes include 3U090, 4U090 (fully built servers containing controlled ICs) and 4A003.u / 4A004.u / 4A005.u. **Action:** confirm with MITI before shipping, and make sure the seller — not you — is the exporter of record for any re-export leg.

Malaysia siting note: non-AI data centres are being curbed (PM Anwar Ibrahim, TNB clip 25 Feb 2026), Johor sits at WUE 1.8 / PUE 1.8, and Selangor imposes water and power conditions (NADMA, 8 Sept 2026). Duty 0% under ITA; sales and service tax 8% / 10%. No published Malaysian list price for an 8-GPU B300 system — the only local anchors are MTC's cloud rate and KONST's lease.

9 · Seller directory — Thailand

COMPANY	ROLE	CONTACT	AI COMPUTE OFFER	NOTES
VST ECS (Thailand) Co Ltd	NVIDIA Marketplace **Distributor** — DGX AI Compute Systems, DGX Cloud, Compute, Networking, Enterprise Software	+66 2032 9999 pr@vstecs.co.th	Can quote HGX/DGX systems	275 Soi Lat Phrao 101, Bangkok 10310
CYN Communication	Systems integrator	contact@cyn.co.th	**GB300 NVL72** (72 GPU + 36 Grace CPU) quoted from ~THB 3m	Quoted, not yet live (GB300 page 2 Jul 2026)
Thai AI Cloud (TAC)	GPU cloud	sales@thaiaicloud.co.th	HGX B300 — 5 nodes / 40 GPUs, 10.5 TB HBM3e, 720 PFLOPS FP4, 2–6 ms — **$8.90/GPU-hr on-demand**; $30,900/node/month; $34,000/month private; **$0.09/M tokens**	node-01/02 Reserved, node-03 Partial, node-04/05 Available — **19 of 40 GPUs unallocated**
Freyr Technology	BOI-promoted AI infrastructure	No public contact found — boi.go.th	US$0.2bn Rayong + Samut Prakan; US$35m initial, US$160m phase Q1 2027	B300/GB300 + Spectrum-X
GreenNode (BKK-1A)	GPU cloud	Site enquiry	H100 $3.22/hr; H100 73,832,147 VND/30 days	Pre-Blackwell capacity
Gorilla Technology	DC developer	gorillatechnology.co.th	Korat 200 MW IT-load campus; NeutralDC BTM-1 Batam 5.5 MW with B300 from Sept 2026	~100,000–110,000 GPU-equivalents, illustrative

Thailand rules. 0% duty under ITA; **VAT 7%, time-limited to 30 September 2027**. BOI approval is required for data-centre projects, and a Royal Gazette notification of 5 June 2026 sets **PUE ≤ 1.3** for high-energy-efficiency data centres, with a draft data-centre construction law pending. No AI-chip import control. Foreign ownership: 49% cap under the Foreign Business Act versus 100% for BOI-promoted projects. Regional price ladder: Singapore $5.00 → Vietnam $6.99 → Thailand $8.90 → Malaysia US$29 per GPU-hr.

10 · Seller directory — Indonesia

COMPANY	ROLE	CONTACT	AI COMPUTE OFFER	NOTES
PT Synnex Metrodata Indonesia	Distributor (NVIDIA)	+62 21 29345 800 Info.Metrodata@metrodata.co.id	Distribution	APL Tower 42F, Jl Letjen S Parman Kav 28, Jakarta
Indosat ZanKore	DC / neocloud	indosatooredoo.com	**~200 MW GB300 NVL72 from H1 2027**; 1 GW DSX target; DSX MaxLPS, +40% compute	Also part of the Ooredoo Zankore consortium — see §6
KONST Group (SMX01 Jakarta)	Colo / neocloud	konst.com.my	15 MW liquid-cooling lease, move-in 26 Dec 2026; B300	Same operator as the 1,024-GPU Cyberjaya build
RAIA Grid / IFT	Assembly / integration	No public contact found U	**B300 assembly plant** — ~US$22m capex, 140–150 units, first production Dec 2026	Local assembly is a price and tariff advantage
Gorilla Technology	DC developer	gorillatechnology.co.th	NeutralDC BTM-1 Batam 5.5 MW with B300 deployments from Sept 2026	Korat 200 MW campus
Lintasarta	Sovereign AI cloud — "GPU Merdeka"	business@lintasarta.co.id U	GPU capacity, Blackwell detail unconfirmed	Jakarta
Gata Cloud	Neocloud	Telegram enquiry only	B300 288 GB with an Indonesia pin — $5.00/GPU-hr	See §7
Biznet GIO NEO GPU	GPU cloud	Site enquiry	H200 Rp74,000/hr; 8 × H200 Rp959,000/hr; 7 × Rp839,000/hr	Pre-B300
Lintasarta Deka GPU	GPU cloud	Site enquiry	H100 XL / Cloudeka H100 Rp79,000/hr / Rp69,000/hr	Pre-B300

Indonesia — duty corrected. My earlier statement that Indonesian servers carry 0% duty under the ITA was wrong. **HS 8471.50 attracts 5% import duty** (HS 8517.62 10%), per trade.gov, despite ITA expectations. VAT is **11%** on an 11/12 DPP base — the 12% rate applies only to luxury goods. There is no AI-chip import control, but ownership matters: **KBLI 63102 (data centre / computer hosting) is 100% open to FDI**, and a PT PMA must hold paid-up capital of IDR 2.5bn, rising above IDR 10bn per project (IDR 50bn for some services). If you intend to own and operate rather than lease, the capital requirement is a real gate. The Unity Group lease — 1,024 B300 8-GPU units for US$123.8m over three years — implies roughly **US$11,250 per GPU-year**, the most transparent unit economics found in the region.

11 · Seller directory — Vietnam

COMPANY	ROLE	CONTACT	AI COMPUTE OFFER	NOTES
FPT Smart Cloud / FPT AI Factory	GPU cloud — first HGX B300 operator in Vietnam	Hotline 1800 6139 support@fptcloud.com	**1× / 2× / 4× / 8× B300 VM from $6.99/GPU-hour** (288 GB/GPU; 8× = 2.2 TB GPU memory, 1536 GB RAM, 224 cores, 24,000 NVMe) — 8× ≈ $55.92/hr	No.10 Pham Van Bach, Cau Giay, Hanoi. First in Vietnam, 30 May / 2–5 Jun 2026
CMC Telecom	Tier III colocation + DC	Site enquiry	Colocation, Hanoi / HCMC	General colo, not a B300 channel
Viettel IDC	DC / sovereign AI	Site enquiry	Vietnam's first DGX B200 supercomputer, Hoa Lac, 1.5 ExaFLOPs FP8 (6 Feb 2026)	B200, not B300
GreenNode (HCM-03)	GPU cloud	Site enquiry	H100 73,832,147 VND / H200 79,989,699 VND per 30 days	Pre-B300

Vietnam is the most permissive market in the region for owning the asset. HS 8471 MFN duty is 0%; VAT 10%. Law 90/2025/QH15 grants import-duty exemption for science-technology and digital-technology goods from 1 July 2025. The **Digital Technology Industry Law 71/2025/QH15 (effective 1 Jan 2026)** with Decree 353/2025 gives **full land-use-fee and rent exemption for AI chips** and 14 other key materials, plus a 4-year corporate income tax exemption with a 50% reduction for nine years for electronics manufacturers. The Law on AI took effect 1 March 2026 (35 articles). **100% foreign ownership for data centres has been permitted since January 2025** under the Telecom Law 24/2023/QH15. If you want to own hardware in Asia rather than lease it, Vietnam is currently the cleanest jurisdiction to do it in.

12 · Seller directory — Philippines

COMPANY	ROLE	CONTACT	AI COMPUTE OFFER	NOTES
VST-ECS Philippines / MSI-ECS	Distributor (NVIDIA)	marketing@msi-ecs.ph M. Eusebio Ave, Pasig	Distribution	Main NVIDIA-capable broadline channel
YCO Cloud × Aolani	Neocloud / NVIDIA Cloud Partner	enquiries@aolani.com	**>10,000 Blackwell Ultra / GB300 NVL72**; phase 1 Q1 2027; dedicated from 2,000 GPUs	Pricing not published
PEZA / DICT	Government anchor	dict.gov.ph	Clark–Bataan named "primary anchor for AI compute infrastructure"	PEZA AO 45 s.2026 lifted the 7-year Metro Manila IT-park moratorium

Philippines rules. Duty 0% under CITI/PEZA; VAT 12% (RA 12023, RR 3-2025). No AI-chip import control; DTI/BPS rules cover certain electronics, not servers. PEZA registration is required for incentives, and foreign ownership is generally permitted inside PEZA. No Philippine provider publishes a B300 $/GPU-hour rate.

13 · Brunei, Cambodia, Myanmar

- **Brunei — UNN × Vobile:** 25 MW AI data-centre capacity, online 2028, NVIDIA GPU systems; UNN/EVYD first national AI platform with AI GPU servers (8 Dec 2025). No B300/HGX quote found.
- **Cambodia — KIDZ AI:** GPU revenue expected Q4 2026. No B300/HGX quote found.
- **Myanmar:** no relevant NVIDIA HGX/B300 channel, permit or pricing signal found.

Treat all three as pre-commercial for Blackwell. None belongs on a shortlist for a 2026–2027 B300 purchase.

14 · Seller directory — Japan

Japan is the most active B300 buyer market in Asia and has the deepest published procurement record. Note that NVIDIA's *ja-jp* "where to buy" page carries no Japan-specific reseller block — it falls back to the global list — so the country-specific roster comes from the Asia-Pacific networking-reseller page plus NPN marketplace listings.

COMPANY	ROLE	CONTACT	AI COMPUTE OFFER	NOTES
SuperX Industries (Japan arm)	OEM / integrator	059-253-6630 superx.co.jp/pricing	B300 and GB300 SKUs (XI6150, NVIDIA GB300 rack) — quote only	Yokkaichi, Mie. Delivered US$38.8m to Woodman
Hitachi Vantara	OEM	Site enquiry — hitachivantara.com	**HGX B300 liquid-cooled in the iQ enterprise line**	Tokyo. Announcement jp260327
GDEP Advance	SI ("AI Factory")	NPN marketplace	DGX AI Compute Systems	Tokyo
ASK Corporation	Distributor + SI	NPN marketplace	Compute	Tokyo
Fujitsu	OEM / NPN	NPN marketplace	DGX/HGX systems	Tokyo
Artiza Networks	Reseller (networking)	+81-42-529-3494	DGX / networking	Tachikawa
RESTAR Communications	Distributor	+81-3-5781-5130	DGX/HGX	Shinagawa
Fuyo Video	Reseller	045-924-0514	Networking / DGX	Tokyo
Ikegami	Reseller	+81-3-5700-4114	Networking / DGX	Tokyo
MITSUI Knowledge Industry · NETWORLD · NEWTECH · VISUAL GRAPHICS	Reseller	No public contact found	Networking / DGX	Tokyo
DataSection	Buyer / operator	Yuichi Nozawa +81-50-3649-4858	635 B300 servers, Inzai (Chiba)	US$325m order via Compal

COMPANY	ROLE	CONTACT	AI COMPUTE OFFER	NOTES
NTTPC + Getnets + Fixsters	Container DC operators	Site enquiry	Water-cooled GPU container DC (HPE ProLiant XD685, HGX B300)	From ¥300M capex, ¥750k/month opex, min 8-month lead
GMO Internet	Neocloud	Internet service desk	HGX B300 managed Slurm (dedicated); shared ¥100/min per node; DDN ¥30,000/TiB/month	B300 in production since 25 Mar 2026
NTTドコモビジネス	GPUaaS	Site enquiry	B300 / A100 GPUaaS — monthly cap **¥1,450,000 tax-excl** (B300), ~¥50,000/day; A100 ¥400,000	Live 17 Sep 2026
Starseeds Cloud	GPU cloud	ssd-sales@starseeds-digital.co.jp 03-6435-4433	B300 288 GB, 7,500 TFLOPS FP8 — **$5.33/GPU-hr**, 1-second billing (B200 $4.25, H200 $3.51)	Beta
Clore.ai	Bare-metal GPU broker	Site enquiry	B300 8-GPU blocks — **$5.16/GPU-hr in Japan and the US**, $5.02 EU, 360-day minimum	Rates as of 21 Sep 2026

Japan — no import permit, and the cleanest customs path in the region. METI's import-approval list contains **zero entries** for HS 8471, HS 8542, integrated circuits or semiconductors; only 2号承認 for designated regions applies. HS 8471.50 servers carry a WTO rate of 0%. Consumption tax is 10% and **recoverable through the invoice system** for registered businesses. Export and re-export are controlled under METI's 輸出貿易管理令 plus the catch-all controls (大量破壊兵器 / 通常兵器), so the relevant question for a Japanese buyer is not import permission but end-use screening and the re-export posture if the hardware later leaves Japan. One operational note: NVIDIA's own facility reference gives 16 DGX B300 racks at 820 kVA average / 1,228 kVA peak — design your facility against that, not against the node nameplate. Also note the listing discrepancy: NVIDIA's own HGX table gives **2.1 TB total (270 GB per GPU)**, while GMO and some OEM sheets publish 288 GB per GPU, and Lenovo and ASUS publish 270 GB / 1,100 W. **Correction to my first edition, which attributed 288 GB to NVIDIA.**

15 · Seller directory — South Korea

COMPANY	ROLE	CONTACT	AI COMPUTE OFFER	NOTES
Samsung SDS (Cloud Platform)	CSP / OEM	Site enquiry — docs.e.samsungsdscloud.com	**Korea's first B300 service.** Instance type `g3c128b8_metal` = 8 × B300, 2.1 TiB, 128 vCPU, 1.8 TB/s NVSwitch, 13.5 PFLOPS FP4/GPU	Price not published; on-demand / 1-yr / 3-yr plans, 3-yr up to −50%
Super Solution (GIGABYTE Korea)	Distributor	1661-6560 sales@supersolution.co.kr	GIGABYTE TO46 8-GPU B300 baseboard, B300/GB300	Seoul (Geumcheon-gu)
Quantum Bricks	Reseller (whole machines)	48-hour quote — quantum-bricks.com	HGX B300, DGX B300, DGX GB300 NVL72 — **ex-VAT list: HGX B300 (8) ₩1.2bn, DGX B300 ₩1.2bn, DGX GB300 NVL72 (72) ₩12bn**	Seoul. The only published Korean hardware price card
kt cloud	CSP	Site enquiry	512 B300 GPUs via XMax/DSH (64 nodes + 560 TB)	Live 29 Oct 2026
NHN Cloud	CSP	dl_aiinfra@nhn.com	GPU instance g4 = A100 × 8 at ₩47,848/hr; max B200 fleet	Seongnam. No B300 SKU price
Daou IDC	Colo / hosting	1600-6511 Mapo 070-8795-0017 · Seocho 070-8795-0790 idc_sales@daou.co.kr	Hosting — no B300 product found	Seoul. **Not in the NPN roster**
Daou Technology	IT services	02-3450-4500	No GPU/B300 line found	Seoul. Weak as a B300 channel
13 NPN partners	Distributors / SIs	NPN marketplace, KR	DGX AI Compute Systems — Azwell AI, BayNex, BN I&C, Infra Information Technology, AimToG, Coupang (Cloud Partner), Elice, ICRAFT, ICOMER, Feelaware, GIT, DS&G, Dongkuk Systems	Korean roster is complete on NVIDIA's own page
Elice	SI / CSP	NPN marketplace	**2,560 B300** under MSIT AI Highway	Seoul

South Korea — the import regime has changed. MOTIE's revised strategic-goods notification took effect **1 September 2026**: classifications moved from 3A903 → **3A501.a.16 (TPP ≥ 6,000)** and 4A902 → 4A507, with new line 3A502.i and 3B901–3B904 → 3B501–3B504. The **Exporter's Written Pledge was abolished** and replaced by **Article 18-3 due-diligence duties carrying criminal penalties**; non-listed goods suspected of WMD diversion now need a special licence. Separately, Korea applies an **import purpose confirmation certificate (수입목적확인서)** under Foreign Trade Act Article 27 — roughly 7 days, valid one year. Whether B300 *servers* specifically trigger it is not confirmed, but the general regime applies to strategic-goods imports. HS 8471 is WTO duty-free; VAT is 10% on CIF plus duty. Practical note: Korea has **no published on-demand B300 $/GPU-hour** anywhere — Samsung SDS is quote-only, and the cheapest published Korean GPU pricing is the government's own self-pay B200 rate of ₩800,000/GPU-month.

16 · Seller directory — Taiwan

Taiwan is the only market where you can buy at the factory. NVIDIA's Taiwan page names authorised sellers explicitly, and the NPN roster lists 51 partners.

COMPANY	ROLE	CONTACT	AI COMPUTE OFFER	NOTES
Zero One Technology (零壹科技)	**NVIDIA authorised agent**	02-2656-5678 nvidia@zerone.com.tw	DGX/HGX	Taipei. On NVIDIA's own where-to-buy page
Macnica Galaxy Taiwan (茂綸)	Authorised distributor	02-8913-2200 NVIDIA.galaxy@macnica.com	DGX/HGX	Taipei. On NVIDIA's own where-to-buy page
Giga Computing (GIGABYTE subsidiary)	OEM	+886-2-8912-6699 marketing@gigacomputing.com	**GAIFA 700 kW: 2 × GB300 NVL72 racks + 1 HGX B300 rack** (G4L4-SD3-LAX7 DLC); racks DLB2-CB3 (GB300 NVL72), DL83-GP6 (8 × B300 liquid-cooled)	New Taipei (Xindian)
ASUS	OEM / NPN	servers.asus.com	**XA NB3I-E12** (9U, 8 × B300)	Taipei
Pegatron	OEM	svr.pegatroncorp.com	**AS402-2T1-8H2** HGX B300 8-GPU	Taipei
Foxconn / Big Innovation	NVIDIA Cloud Partner	Site enquiry	**US$1.4bn AI cloud supercomputing centre, 10,000 Blackwell GPUs, GB300 NVL72**	Plus Visionbay.ai sovereign AI centre, online 1H 2027
51 NPN partners	Distributors / OEMs	NPN marketplace, TW	DGX AI Compute Systems — ASRock Rack, Compal, Arrow Electronics Taiwan, Acer, ADLINK, AAEON, AIC, Advantech, Accton, Axiomtek and others	Largest NPN roster in the region

COMPANY	ROLE	CONTACT	AI COMPUTE OFFER	NOTES
Computeify	GPU cloud	computeify.ai/gpus	**$6.10/GPU·hr ex-tax**, 3-year lease, delivery Mar 2027 (5-month lead); hall lease $3.3–4.5M/month at 30 MW ($110–150/kW/month)	Taiwan DC, 32 machines
KONST (Taiwan)	Neocloud	konsttech.ai/en/compute	New Taipei, 4 nodes / 32 GPUs, 288 GB — from **$6.00/GPU-hr** (billed monthly)	Pre-sale, schedule TBD
GPUtw.ai	Reseller	gputw.ai/en	"HGX B300 XE9780", 270 GB — whole machine, formal quote only	Other SKUs priced (RTX PRO 6000 WS $1.97/hr)
NVIDIA Taiwan support	Support	+886-2-2656-3206	—	Taipei

> **Taiwan — the lowest-cost legal route, with one new risk.** HS 8471 servers and 8473.30 GPU/AI accelerators are **duty-free under Taiwan's WTO ITA**. Importation attracts a **business tax of 5%**, recoverable only where the registered business is the named duty payer and uses the goods for taxable supplies; customs duty itself is never recoverable. You need importer/exporter registration and a Unified Business Number, and restricted items need an MOEA ITA import permit (typically 1–10 working days) via the Customs-Port-Trade Single Window. **New risk:** Taiwan is moving advanced semiconductors and AI servers onto the Foreign Trade Act controlled list — a public hearing was held on 2 November 2026 and draft terms would jail unauthorised AI-chip exports to China. *Not yet in force*, but it directly affects any hardware you buy in Taipei for onward movement. Also note that Giga Computing publishes 288 GB while GPUtw publishes 270 GB for what is nominally the same board — pin the SKU in writing.

17 · Seller directory — India

India is the second-largest B300 market in Asia by scale and the only one with a **government-published B300 price card**. Note that NVIDIA's India networking-reseller page lists four firms (Hitachi Systems Micro Clinic, Marg Infotech, Team5 Technologies, Ninth Dimension IT Solutions) but they are **networking-only** — none shows evidence of quoting HGX/DGX systems.

COMPANY	ROLE	CONTACT	AI COMPUTE OFFER	NOTES
Yotta Data Services	NCP / sovereign cloud	+91 2192-400011 yntraacloud.ai/contact	**Shakti Cloud — 20,736 Blackwell Ultra**, >10,000 of them B300 committed to IndiaAI	Powai Mumbai, Greater Noida, Navi Mumbai. MeitY-empanelled
Netweb Technologies	**OEM** — NVIDIA Blackwell partner, "Tyrone" brand	+91 11 4324 0000 info@netwebindia.com	GB200 NVL4 Made-in-India; B300/GB300 in build. MLPerf v6.0 entry (Supermicro AS-8126GS-NB3RT + B300)	New Delhi, 18 offices. India's NVIDIA server OEM

COMPANY	ROLE	CONTACT	AI COMPUTE OFFER	NOTES
E2E Networks	NCP / neocloud	sales@e2enetworks.com +91-11-4084-4965	B200 procured Jan 2026; Blackwell on Trusted India Platform	New Delhi (Badarpur), Chennai
Locuz Enterprise	**NPN Elite SP** — the only India card on NVIDIA's DGX page	+91-9500076754 locuz.com	HPC / GPU cloud	Confirmed on nvidia.com/en-in/data-center/get-dgx
CtrlS Datacenters	Colo / GPU cloud	No public contact found — ctrls.com/contact-us	IndiaAI-empanelled, 1,004 GPUs	Hyderabad (Madhapur)
NxtGen DC & Cloud	NCP / cloud	marketing@nxtgen.com	IndiaAI-empanelled, 1,080 GPUs	Mumbai
Jio Platforms · Tata Communications · Sify	Telco / cloud	Via IndiaAI empanelled list	IndiaAI-empanelled GPU capacity	Mumbai
Hitachi Systems Micro Clinic	Reseller (networking)	+91-11-49250700 / +91-11-49250723	Networking only — **no HGX/DGX evidence**	E-44/2 Okhla Industrial Area Ph-II, New Delhi 110020
Marg Infotech	Reseller (networking)	+91-11-47095461	Networking only	302 Kumar Tower, Wazirpur Industrial Area, Delhi 110052
Team5 Technologies	Reseller (networking)	+91-9632900220	Networking only	No.63 OVH Road, Basavanagudi, Bangalore 560004
Ninth Dimension IT Solutions	Reseller (networking)	Via NVIDIA APAC reseller page	Networking only	Skywalk, Assaye Road, Bengaluru 560042
L&T / Vyoma.AI → Together AI	Buyer / integrator	Site enquiry	**10,000 B300 on one NVLink fabric, Chennai**, up to US$1.6bn	Phase 1 250 MW / 150 MVA

India — the only published B300 rate card, and a live import-policy contradiction. The **IndiaAI Compute portal** publishes **B300 SXM at 1×/2×/4×/8×, ₹351/GPU-hour on-demand** (8× = ₹2,808/hr), 1-month ₹342, 6-month ₹333, **12-month ₹319.5**; B200 SXM 1× is ₹290.7. Yotta's Shakti Cloud lists an 8× HGX B300 SKU but **leaves the price cell empty** (its 8× B200 sits at ₹473/GPU-hr), and its bare-metal card publishes **H100 reserved tiers** of ₹1,056 (1 yr), ₹864 (2 yr), ₹672 (3 yr) and ₹480 (4 yr) — **not B300 rates**, and easily misread as such. **The contradiction to be aware of:** the live price list carries B300 lines, but the IndiaAI empanelled-GPU allocation document still shows 8,192 B200 allocated entirely to Yotta and **no B300** — B300 is priced but not yet allocated under the programme. On imports: HSN 8471 servers are **Restricted** under the DGFT Import Management System, with applications open 22 Dec 2025 – 15 Dec 2026 and any Import Authorization valid only to 31 Dec 2026; a mid-term MeitY review is underway. BCD is 0%, SWS is 10% of BCD, IGST 18% (ITC-recoverable), and BIS CRS registration may apply to some equipment.

18 · Pakistan, Bangladesh, Sri Lanka — thin or absent

MARKET	CHANNEL	CONTACT	ASSESSMENT
Pakistan	Data Vault Pakistan — integrator / GPUaaS	info@datavault.com.pk +92 21 3346 1111	Karachi (Korangi). B200/H200/A100 for NTC; **Supermicro study of B200/B300 + GB300 NVL72 only** — no confirmed purchase. IT equipment at 0% CD + 18% sales tax; PSW permits for controlled items; SIFC/SBP FX repatriation green channel applies; B300 not on the protected capital-goods list
Bangladesh	Potaka IT — retailer, **not an NVIDIA partner**	01841-995444 palash@potakabd.com	Lists "GB300 NVL72" at ৳770,000,000 marked "In Stock", 10–30 day delivery — **provenance unverified; treat as a listing, not a channel**. Officially no B300 deployment; public GPU cloud is ~20 Volta GPUs, 2,240 TFLOPS (launched Jan 2026). **Tax win:** NBR gazette 17 Aug 2026 exempts HS 8471.50.90 servers from CD/RD/SD/VAT, cutting incidence from ~24% to ~9.5% (including 7.5% advance tax). Draft National AI Policy 2026–2030 proposes duty-free GPU imports plus mandatory localisation
Sri Lanka	None found		No B300 channel, no GPU-cloud vendor, no deployment. The tax stack is punishing anyway: CID 0–100%, PAL 7.5%, CESS 2%, excise/SPD, SCL, SSCL 2.5%, then VAT 18% applied to the cumulative base

19 · Seller directory — United Arab Emirates

COMPANY	ROLE	CONTACT	AI COMPUTE OFFER	NOTES
Core42 (G42)	Sovereign AI cloud	+971 2 699 2700 / +966 112 486 551 aicloud.core42.ai	**GB300 "Price: On Request"**; B200 from $5.00/hr; H200 from $4.00; H100 from $2.50	Abu Dhabi + Dubai (Media1 Tower) + Riyadh + Doha. Named in 15 CFR 740 Supp. 8
PNY Technologies Middle East FZE	**NVIDIA distributor**	+971 4 881 4966 SALES@PNY.EU	DGX GB300 (72 × Blackwell Ultra, 20.1 TB HBM3e)	Jebel Ali Free Zone, Dubai
ITQAN AI Khaleej (Yas Holding)	NPN Elite Solution Provider	+971-2-4037953 mohammed.aslam@itqan.ae	DGX systems	Abu Dhabi. On NVIDIA's where-to-buy page
haink	Integrator / broker (export-screened)	No public contact found — haink.org/contact	B300 SXM / GB300 NVL72 "on allocation", 14–24 week lead	Dubai free zone. Price and lead-time table is vendor marketing, not a quote
Greenware Technologies	Reseller (networking)	+971 4 832 7552 info@greenware.ae	**Networking only**	Dubai

COMPANY	ROLE	CONTACT	AI COMPUTE OFFER	NOTES
Mindware FZ LLC	Reseller (networking)	+971 4 450 0600 sales@mindware.ae	**Networking only**	Dubai
servchip	Self-declared "AUTHORIZED NVIDIA DISTRIBUTOR – UAE"	No contact found — servchip.com/ae	H100/H200/B200 — **B300 not mentioned**	Sharjah free zone / New Delhi. The authorisation claim is self-declared
GPU.ai	Broker	Site enquiry	B300 288G, 2 regions — **$8.99/GPU-hr**	UAE-based
Khazna	DC developer	Site enquiry	QAJ1, Ajman — 100 MW under construction	Oct 2026

> **UAE — the export position changed materially in July 2026, and it is time-limited.** The **UAE moved to Country Group A:5** (removed from **Country Groups D:3 and D:4**) effective **10 July 2026** under the Federal Register. UAE government agencies plus **G42 and Core42** are named in 15 CFR 740 Supplement No. 8 as Approved Entities with licence-free access to advanced computing — but **that status expires after 270 days (~6 April 2027) unless the entity becomes a "U.S. company" or obtains a BIS Advisory Opinion**. Practically: the UAE is the easiest place in Asia to take delivery of a GB300 or B300 at scale, and the Approved Entity status needs re-checking before any delivery that lands after April 2027. Israel, Saudi Arabia, Qatar and Turkey have no equivalent listing — treat them case by case, and get a licence or an exemption reference in writing before committing. Taxes: 5% customs on CIF under the GCC Common Customs Law, 5% VAT, and free-zone zero-rating if the goods stay in-zone; UAE PDPL plus CBUAE/DOH in-country rules apply, though there is no general mandate that training compute sit in-country.

20 · Seller directory — Saudi Arabia

COMPANY	ROLE	CONTACT	AI COMPUTE OFFER	NOTES
HUMAIN (PIF)	Sovereign AI cloud	Site enquiry	**84 MW / ~27,000 GPUs.** Phase 1: 14 MW ≈ 6,000 HGX B300 on Quantum-X800 from Q4 2026; 70 MW Vera Rubin from Jun 2027	Riyadh and Dammam. The largest B300 commitment anywhere
Core42	Neocloud	+966 112 486 551	GB300 on request; B200 from $5.00/hr	Riyadh
MTC (Memory Technology Co., KSA)	**PNY's official Saudi distributor**	WhatsApp +966 53 077 5999 info@mtc-ksa.com	DGX Spark and next-generation Blackwell GPUs	Jeddah / Riyadh. Announced Sep 2026
e& / STC	Telco / cloud	Site enquiry	AI infrastructure	Kingdom-wide programmes

COMPANY	ROLE	CONTACT	AI COMPUTE OFFER	NOTES
Nebius (via DataVolt)	Neocloud	Site enquiry	Offtake of ≤50 MW at Oxagon	Agreement 21 Oct 2026

Saudi rules. VAT 15%; customs 0–15% by tariff line (the exact HS 8471 line is unconfirmed — verify with a Saudi broker); ZATCA/SABER conformity required. Saudi PDPL applies with consent or NDMO approval, and SAMA imposes core-data residency for financial data. Saudi is *not* in Country Group A:5 and has no STA listing, so shipments need case-by-case BIS screening even though the HUMAIN programme is a direct NVIDIA partnership.

21 · Qatar, Kuwait, Oman, Bahrain

COMPANY	ROLE	CONTACT	DETAIL
Ooredoo / Syntys	NVIDIA Cloud Partner	ooredoo.com contact	Qatar 30 MW DC operational; **GB300 via Zankore (Indonesia)** — see §6. Also Oman, Kuwait, Tunisia, Algeria expansion in 2026
Ooredoo–NVIDIA Sovereign AI Lab	National lab	Via Ooredoo	Kuwait, 10 MW — **H200 only** (Nov 2025)
Kuwait, Oman, Bahrain	No NVIDIA-published per-country reseller list		NVIDIA's EMEA networking-reseller page lists **no** Saudi, Qatar, Kuwait, Oman or Bahrain entries — only two UAE firms. Sourcing here runs through Ooredoo/Core42 regional contracts

GCC tax (outside the UAE): Qatar and Kuwait have **no VAT in force** and 5% customs. Oman has 5% VAT and 5% customs; Bahrain 10% VAT and 5% customs. Saudi has 15% VAT and 0–15% customs by line.

22 · Seller directory — Israel

COMPANY	ROLE	CONTACT	DETAIL
Harel Information Technologies	NPN Elite SP	9727658888 hit@harel.co.il	DGX/HGX systems — on NVIDIA's where-to-buy page
Bynet Data Communications	NPN Elite SP	972546683976 Ohad.Dudai@bynet.co.il	DGX/HGX plus data-centre integration
Matrix IT	NPN Elite SP	972544276952 Amitbe@matrix.co.il	DGX/HGX
Crusoe Cloud	Neocloud	crusoe.ai/cloud/pricing	HGX B300 / GB300 NVL72, self-serve; Israel (Afula) from 2026 — no Israel-specific rate published

COMPANY	ROLE	CONTACT	DETAIL
Nebius AI Cloud	NCP	docs.nebius.com	**me-west1 carries B200 only — B300 is not in the Israel region.** B300 $9.50/GPU-hr in uk-south1 / eu-west2 / us-north1
NVIDIA Israel	OEM R&D	info@nvidia.com	Beer Sheva, Raanana, Tel Aviv, Tel Hai, Yokneam — **does not sell**

Israel rules. 18% import VAT plus customs duty; the HS 8471 duty rate was not verified. Dual-use export control under the Defense Export Control Law 576-2007 may apply — applicability to AI servers was not confirmed against an authoritative source, so treat it as an open question. Israel is not in Country Group A:5 and has no STA listing, so advanced-compute shipments need BIS licence or exemption screening. The Israeli AI Research Compute Center is the outlier on price: **B200 at $0.70 + VAT per GPU-hour including 5 TB storage**, 80% subsidised — the cheapest compute published anywhere in this study, and a useful benchmark for what a government-subsidised program does to your effective cost.

23 · Seller directory — Turkey

COMPANY	ROLE	CONTACT	DETAIL
GTM Teknoloji	**NPN partner + Supermicro's official Turkey distributor**	info@gtmteknoloji.com +90 212 639 45 43 Yenibosna Mah. 29 Ekim Cad. Vizyon Park 4.Blok 9/B, Bahçelievler/İstanbul	**GB300 NVL72 and HGX B300 JumpStart remote test**; **3–7 day rack delivery; GB300 lead time 4–8 weeks with local stock**; 24-hour quote. The shortest quoted lead time in this study
Creentech · Intron IT · KoçSistem · NSC	Resellers (networking)	Creentech (212) 284 85 45 NSC 0212 461 6300	**Networking only**

Turkey rules. 20% KDV on (CIF + duty); the HS 8471 duty rate was not verified. Strategic-goods licensing by the Ministry of Trade / Ministry of National Defence / TAEK applies to controlled IT, and customs will delay without a permit. Turkey is not in Country Group A:5 — BIS screening is required on paper, though enforcement targets D:1/D:4/D:5.

24 · China, Hong Kong, Macau — no compliant route

> **Do not buy B300 through a China, Hong Kong or Macau channel.** B300 is ECCN 3A090.a; the US requires a licence and none is being issued for these destinations. The mainland B30A SKU is also blocked. The Shenzhen spot market operates in a grey market with no NVIDIA warranty, no support entitlement and direct export liability for the buyer. **Financial Times reported in November 2026 that a DGX B300 in China now sells for more than 8 million yuan (~US$1.1 million), up from 4 million yuan six months earlier** — a doubling in six months, which is a scarcity signal, not a price discovery. Cross-border transfer of a single 8-GPU B300 system is reported at **8–12 weeks plus ~¥7m in channel cost and ~30% risk compensation** — and the transaction itself remains non-compliant. For a company that needs audit-clean provenance, this route is not a cost trade-off; it is a disqualifier.

For completeness: the China grey-market opex model (~14 kW, PUE 1.3) works out to ¥280–450k per year for electricity, colocation, bandwidth, staff and maintenance, giving a static payback around 45 months on a ¥8m system — attractive on paper, and irrelevant if the hardware cannot be lawfully imported. The market data is also internally inconsistent: figures of "¥7.0m per system" (17 Sep 2026), "¥18.9–19.6m" (20 Oct 2026) and "more than 8m yuan" (Nov 2026) all circulate. Treat all China figures as narrative, not as marks; the only figure with a named publication behind it is the FT's "more than 8 million yuan".

25 · GPU cloud price board — what a B300 costs per hour today

Renting is the cheapest way to find out whether you actually want this hardware. These are live published list prices re-checked in November 2026, not indicative ranges.

PROVIDER	MARKET	PRODUCT	ON-DEMAND	COMMITTED / RESERVED	NOTE
Gata Cloud	Singapore	B300 node	**$5.00 / GPU-hr**	—	Cheapest live Asian on-demand price found in this study CONFIRMED
Clore.ai	Japan / US	B300	$5.16 / GPU-hr	—	Hourly spot marketplace; useful as the lower bound
Starseeds	US	B300	$5.33 / GPU-hr	—	Global benchmark, not Asia
KONST	Singapore / Malaysia	B300	**$6.00 / GPU-hr**	5-year term, billed monthly	Also a confirmed B300 deployer (1,024 GPUs at Cyberjaya); one of the few Asian providers selling and operating the same SKU
Computeify	Taiwan	B300	$6.10 / GPU-hr	—	Taiwan's domestic price sits above the global median
FPT AI Factory	Vietnam	B300	$6.99 / GPU-hr	—	Backed by a government-certified AI factory; local power is cheap
TAC	Taiwan	B300	$8.90 / GPU-hr	—	Premium over US marketplace on identical silicon
GPU.ai	Global	B300	$8.99 / GPU-hr	—	Mid-market reseller
MTC	Malaysia	B300	RM120 / GPU-hr	—	Listed under MYR; identity and currency basis to confirm with the vendor UNVERIFIED
SuperX AI	Singapore	B300 270 GB	**$35.84 / spot-hr** (8-GPU instance)	—	Spot tier only; the 8× instance price divided by 8 GPUs is **$4.48/GPU-hr** — a floor, not a walk-up rate MAJORITY
Yotta — IndiaAI portal	India	B300 SXM 1×	**₹351.0 / GPU-hr**	₹342.0 (1 mo) · ₹333.0 (6 mo) · **₹319.5 (12 mo)**	≈$4.03 on-demand. Government-subsidised capacity: lowest credible live Asian price, but with eligibility conditions and allocation risk
Shakti Cloud — bare metal	India	8× HGX B300	**No published price**	—	The bare-metal table prices H100 (₹351) and B200 (₹473) but publishes **no price for 8× HGX B300** — quote required UNVERIFIED

PROVIDER	MARKET	PRODUCT	ON-DEMAND	COMMITTED / RESERVED	NOTE
NTTドコモビジネス	Japan	B300	—	**¥1,450,000 / month** (node)	Japanese entity, domestic billing, enterprise contract
Core42	UAE	GB300 / B300	"On Request"	—	Named in 15 CFR 740 Supp. 8 as an Approved Entity — but that status **expires after 270 days unless Core42 becomes a "U.S. company" or obtains a BIS Advisory Opinion** (~6 Apr 2027)
Samsung SDS	South Korea	B300	Quote only	—	Placed POs for 2,016 B300 under the MSIT AI Highway
Together AI	Global	B300 270 GB	**$9.99 / GPU-hr**	"from $6.79 / hr"	Publishes a 270 GB cluster spec — one of the vendors that quotes the figure matching NVIDIA's own spec rather than the 288 GB listing figure
Nebius	Global	B300	**$9.50 / GPU-hr**	$8.50 tier	Raised from $7.85 on 1 Oct 2026 (**+21%**). The clearest price-direction signal in the market

GLOBAL REFERENCE BAND

- Median on-demand across 38 providers: **$9.50/GPU-hr, +14% over 90 days**. Treat this as the negotiation anchor — a B300 quote above $9.50 list is above the global median.
- B200 on the same marketplace: **$8.50/GPU-hr** (raised from $7.15). B300 is carrying roughly a 12% premium over B200, not a 3× premium. **Correction to my first edition, which quoted $7.85 here — that was B300's own pre-increase price.**
- H100 on-demand fell to **$1.80–3.50/GPU-hr** in Q2 2026 — down 60–77% from the ~$8 peak — then 1-year reserved pricing bottomed at $1.70 (Oct 2025) and **rose to $2.35 by Mar 2026 (+40%)**. The relevant lesson: older accelerators do not go to zero, they re-price twice.

> **How to use this board.** Published on-demand rates are list prices for walk-up users; committed enterprise terms in Asia typically land 30–60% below list and are never published, so this board is a ceiling, not a forecast of what you will pay on a 24-month contract. Always quote *per GPU-hour for an identical SKU* — the 8-GPU B300 node and the GB300 NVL72 rack are not the same product, and a rack at a lower per-GPU rate is not a cheaper substitute for a node if your workload needs fewer than 72 GPUs.

26 · Hardware price anchors — what the systems actually cost

ITEM	PRICE	PRICE TYPE	NOTE
Supermicro B300 server (EUR listing, two SKUs)	**€559,477.68** €567,074.61	Third-party listing	Two configurations, ~1.4% apart. **Not re-confirmed on Supermicro's own page in this pass** — treat as a market observation UNVERIFIED
Supermicro B300 server (USD listing)	**$570,220.95**	Third-party listing	≈$60.8k/GPU including CPU, memory, storage, fabric and support UNVERIFIED

ITEM	PRICE	PRICE TYPE	NOTE
Supermicro comparable 4U B200-class system	€59,326	OEM EU store	Supermicro's own European storefront shows a comparable 4U Blackwell system at €59,326 — **roughly one tenth of the B300 listings above**, which is why the B300 figures need a written quote, not a listing
Thai reseller — SYS-822GS-NB3RT	**THB 25,500,000** (~US$775k)	Reseller list	A local Thai reseller listing for the 8-GPU B300 system — the most concrete Asian landed-quote reference found MAJORITY
GIGABYTE B300 system (FOB)	**$689,000**	FOB quote	Excludes freight, insurance, duty and on-site install. **Not re-confirmed on GIGABYTE's own page in this pass** UNVERIFIED
NVIDIA DGX B300 (configured)	**$300,000–500,000**	List band	NVIDIA-branded 4-system SuperPOD; the band is wide because configuration and support tier drive it
HGX B300 8-GPU baseboard (bare)	**from $485,000**	Quote floor	The board only — no chassis, CPU, NIC, storage, warranty
GB300 NVL72 rack	**$3.7–4.0M** street ≈$3.99M BOM	Integrator quotes	≈$51–56k per GPU — **cheaper per GPU than the node** and materially better per token
Quantum Bricks (Korea)	₩1.2bn / ₩12bn	Reported	Two configurations reported in Korean press; treat as indicative UNVERIFIED
B300 system lead time	**8–14 weeks**	OEM guidance	From confirmed PO to delivery
GB300 rack lead time	**4–8 weeks**	Local stock	Where stocked in-region; **GTM Teknoloji (Turkey) quotes 3–7 day rack delivery** — the shortest in this study

> **The hardware price anchors are the weakest numbers in this dossier.** The €559k / $570k / $689k figures could not be re-confirmed on the vendors' own pages in this pass, and Supermicro's own European storefront shows a comparable 4U Blackwell system at **€59,326** — an order of magnitude below the B300 listings. Two readings are possible: the B300 listings are genuine at roughly 10× a B200-class system, or they are stale, misconfigured or speculative. **Do not put a hardware budget together from this table.** Get a written, dated quotation with a SKU and a delivery date; use these numbers only to sanity-check what a seller comes back with. One part of the table is nevertheless corroborated independently: the **$3.7–4.0M rack band** implies $51–56k per GPU, and the DataSection × Inventec order in §6 — 10,000 B300 for **US$521M** — works out to **$52.1k per GPU**, inside that band. The **node** system prices are not corroborated by anything comparably firm; only the **$485k bare-baseboard floor** ($60.6k per GPU) is cross-checked, against the $64.0k per GPU implied by the DataSection × Compal order.

> **Read the spread as integration content, not as margin.** The gap from a $485k bare baseboard to a $570–775k delivered system is the chassis, CPU, memory, storage, NICs, rails, wiring and a warranty — roughly 20–40%. A buyer comparing a $485k board quote against a $775k system quote is comparing two different products. And note the direction of travel in the market: the confirmed Asian deployments in §6 are overwhelmingly **racks**, because $3.7–4.0M for 72 GPUs beats $570k for 8 GPUs on every axis except granularity and the ability to fit in an existing room.

27 · Export controls and import permits — the compliance spine

This is where Asian procurement most often breaks. A B300 is ECCN **3A090.a**; the US requires a licence to export to most of the markets in this dossier, and the licence conditions differ by destination and by end-user.

JURISDICTION	INSTRUMENT	STATUS IN NOV 2026	WHAT IT MEANS FOR A B300
United States (origin)	BIS EAR, ECCN 3A090.a	In force	Licence required for most destinations below. Destination grouping changed **10 Jul 2026** (FR doc 2026-14132): the **UAE was removed from Country Groups D:3 and D:4 and added to A:5**, materially easing UAE shipments. **Correction to my first edition, which said D:1** CONFIRMED
United States — named end-users	15 CFR 740 Supp. 8	**Approved Entity status expires after 270 days** (91 FR 43038)	G42, Core42, UAE government agencies and HUMAIN Cloud are named as Approved Entities receiving licence-free access to specified AI hardware. **The status expires after 270 days from the effective date (~6 Apr 2027) unless the entity becomes a "U.S. company" or obtains a BIS Advisory Opinion.** This is the single most important date on this page for anyone buying in the Gulf. A contract signed today on the assumption of continued licence-free access has a hard expiry
Japan	METI import-approval list	**Zero HS 8471 / 8542 entries**	Servers and GPU boards do not require Japanese import approval. Clean import path; still subject to US-side screening and end-use terms
South Korea	MOTIE strategic-goods revision	**Effective 1 Sep 2026**	3A903 → **3A501.a.16** (TPP threshold ≥ 6,000), 4A902 → 4A507, new 3A502.i, 3B901–3B904 → 3B501–3B504. The **Exporter's Written Pledge is abolished**; Article 18-3 due-diligence now applies. Import purpose confirmation (수입목적확인서) takes ~7 days CONFIRMED
Taiwan	MOEA Bureau of Industry and Trade	Import permit, 1–10 working days	Straightforward permit regime. A **Foreign Trade Act revision went to consultation on 2 Nov 2026** — watch for changes to strategic-goods listing MAJORITY

JURISDICTION	INSTRUMENT	STATUS IN NOV 2026	WHAT IT MEANS FOR A B300
Malaysia	MITI Directive 1/2025	Active; KLIA seizures reported 2026	**Correction to my earlier advice:** the directive governs **export, transit, transshipment and brokering** — it is not a blanket domestic import licence for HS 8471. Domestic import of servers is not itself MITI-licensed. Where a Malaysia-based party will re-export or broker, an AICA form to admin.sts@miti.gov.my is required with **≥30 days** lead. Customs seizures at KLIA in 2026 confirm real enforcement **CONFIRMED**
India	DGFT Import Management System	HS 8471 — Restricted	Import permitted against the IMS with a restricted-category licence. The current **policy window runs 22.12.2025 → 15.12.2026** — after that date the authorisation route must be re-verified before committing **MAJORITY**
Thailand	Customs Department / BOI	Permit + registration	Standard import licensing; BOI-granted projects can obtain additional incentives. No B300-specific restriction identified
Indonesia	TIU / TDP, customs import permit (Lartas)	Standard regime	Import via the Trade Import Unit with a customs import permit; AI compute items have attracted recent government attention on data-centre siting
Vietnam	Customs + MIC ICT permit	Standard	ICT import permits plus the data-centre licensing regime that underpins the FPT / VNG / Viettel deployments
Philippines	BPS / DTI	Standard	ATIGA zero duty for HS 8471 within the ASEAN cumulation area; no B300-specific control identified
Saudi Arabia	MOFCOM / CITC, SASO conformity	Standard + conformity	No B300-specific export control identified. The binding constraint in Saudi is the end-user and the MASEC / licence structure, not import licensing
UAE	Country Group **A:5** from 10 Jul 2026	Eased but conditional	No licence requirement for the named Approved Entities — but that status **expires after 270 days (~6 Apr 2027)** unless the entity becomes a "U.S. company" or obtains a BIS Advisory Opinion, after which the routing must be re-established
Qatar, Kuwait, Oman, Bahrain	Standard import licensing	Open	No additional B300-specific control identified. Demand-led by sovereign programmes; verify the end-user's own downstream licence position
Israel	Ministry of Economy dual-use licensing	Applies to controlled IT	Israel's dual-use regime does capture advanced compute; obtain a MOD import confirmation before quoting a delivery date
Turkey	Ministry of Trade / MND / TAEK strategic-goods licensing	Applies	Licensing applies to controlled IT; customs will hold a shipment without a permit. Turkey is not in Country Group A:5
China, Hong Kong, Macau	ECCN 3A090.a; mainland B30A SKU also blocked	**No licence issued**	**No compliant route exists.** Any offer into these markets is non-compliant by construction and the buyer carries the liability

> **The 6 April 2027 cliff.** The UAE's licence-free access for G42, Core42, UAE government agencies and HUMAIN Cloud is time-boxed. If your Gulf strategy depends on shipping into the UAE beyond that date, you need a fallback now — either a licence-backed route (Core42 offers one commercially) or a different end-user structure. This is a live planning constraint, not a theoretical one, and it will surface as a contract-renewal negotiation in roughly 18 months.

28 · Tax, duty and landed cost

Hardware duty is almost never the deciding factor for B300. Import VAT/GST is, because it is a cash-flow event on a very large invoice and it is recoverable only if you are registered and actually taxable.

MARKET	CUSTOMS DUTY	VAT / GST	EFFECTIVE ON IMPORT	NOTE
Singapore	**0%**	9%	9%	Zero-rated GST rate; full amount is recoverable against output GST
Malaysia	0% (HS 8471)	SST applies to imported taxable goods and services	Rate depends on the taxable-service/goods classification	**The exact SST rate applicable to the HS 8471 line was not verified in this pass** — confirm with the customs broker, not from a web table
Thailand	0%	7%	7%	VAT relief extended to 30 Sep 2027
Indonesia	**5%** (HS 8471.50)	11%	11% on (CIF + duty)	**Correction to my earlier advice:** I previously recorded Indonesia as duty-free under the ITA. HS 8471.50 carries **5% duty**. On a $570k node that is $28.5k, not $0
Vietnam	0%	10%	10%	Land-use and rental exemptions specific to AI/data-centre projects apply
Philippines	0% (ATIGA)	12%	12%	Recoverable against output tax
Japan	0%	10%	10%	Fully recoverable
South Korea	0%	10%	10%	Recoverable; MOTIE import-purpose confirmation (~7 days) runs in parallel
Taiwan	ITA duty-free concession	5%	5% business tax on importation	Lowest effective import burden in Greater China, subject to ITA permit 1–10 working days
India	BCD **0%**	IGST 18%	18% + SWS 10% of BCD	With BCD at 0%, SWS computes to nil. Input tax credit available; the real cost is the **cash-flow drag on a 22.12.2025–15.12.2026 policy window**
UAE	5%	5%	5% on (CIF + duty)	Recoverable for registered businesses
Saudi Arabia	5% (GCC CET)	15%	15%	Highest VAT in the GCC by a wide margin
Qatar, Kuwait	5%	none	5%	No VAT — the lowest tax-loaded import in the GCC

MARKET	CUSTOMS DUTY	VAT / GST	EFFECTIVE ON IMPORT	NOTE
Israel	0%	18%	18%	Recoverable; often netted out within the same month
Turkey	Not verified for HS 8471	20% KDV	20% on (CIF + duty)	Highest VAT rate among the markets surveyed
Bangladesh	Exempt (HS 8471.50.90)	—	**≈9.5%** all-in	Lowest effective import burden anywhere in this dossier — but no confirmed B300 channel

> **Two practical points that dominate the tax table.** First, every VAT/GST/IGST above is **recoverable** for a registered, taxable business — so the true landed-cost impact of Indonesia's 5% duty is roughly $28.5k on a $570k node, and the true impact of India's 18% IGST is *timing* (a $102k cash outflow recovered later), not cost. Second, the honest cost is not the tax but the **working capital, the broker and the demurrage**: a $689k FOB system into a non-ATIGA country lands with freight, insurance, duty, VAT, 3–7 days of port and customs handling and a broker fee, and the document set (ETA, AICA, MOTIE/ITA/IMS permits, end-user statement) is what actually determines the delivery date.

29 · Data-centre siting and facility economics

A ~14 kW B300 node and a 142 kW GB300 rack are not interchangeable facilities decisions. Air-cooled 8-GPU nodes drop into an existing hall; a liquid-cooled rack needs CDU loops, high-voltage distribution, and a power envelope that most colocation contracts will not give you at a per-kW rate designed for a 20 kW cabinet.

PARAMETER	8-GPU B300 NODE	GB300 NVL72 RACK	NOTE
IT power draw	~14 kW (NVIDIA DGX B300 spec)	~142 kW	NVIDIA's facility reference gives 820 kVA avg / 1,228 kVA peak for a 16-DGX-B300 configuration; design against that
Cooling	Air	Direct-to-chip liquid	The rack cannot run in a conventional air-cooled hall without retrofit
Effective PUE	1.2–1.4 (air)	1.1–1.2 (liquid)	Applying a single 1.3 PUE to both understates the rack's efficiency
Colocation billing	~14 kW	~142 kW + CDU loop	**Check whether your colo rate is power-inclusive** — many Asian contracts bill a committed kW including power, which removes the electricity line entirely and changes the arithmetic

PARAMETER	8-GPU B300 NODE	GB300 NVL72 RACK	NOTE
Annual facility cost (model)	≈**$48k** /yr	≈**$445k** /yr	At $150/kW-month, $0.13/kWh, PUE as above, power billed separately. Assume both are excluded from the facility cost to see the gross facility burden
Indicative colo $/kW-month	Tier-1 (Singapore, Tokyo, Seoul, Hong Kong): $130–200 · Taiwan: $100–140 · Malaysia, Thailand, Indonesia, Vietnam, Philippines: $70–110 · India: $60–95 · UAE, Saudi: $90–140 · Israel: $100–130 · Turkey: $60–90		Indicative market bands, **not quotes** UNVERIFIED — confirm with a colo RFQ per city
Indicative industrial power $/kWh	SG $0.11–0.19 · TW $0.11–0.14 · JP $0.13–0.17 · KR $0.09–0.11 · VN $0.08–0.10 · TH $0.09–0.11 · MY $0.09–0.11 · ID $0.10–0.14 · IN $0.07–0.09 · AE $0.08–0.10 · SA $0.05–0.08 · QA $0.03–0.05 · IL $0.08–0.10 · TR $0.05–0.09		Indicative bands UNVERIFIED; regulatory and FX moves in 2026 have been large
Capacity per rack / per hall	Fits existing densest-in-a-hall halls	~142 kW ≈ one 140 kW cabinet class — a full cabinet	Many Asian colo halls are being re-roofed from 20 kW to 40–60 kW per rack. A 142 kW rack is still ahead of most Asian halls

The facility asymmetry is the real reason Asian buyers order racks in a hurry and nodes when they can. A ~14 kW node can be dropped into almost any AI-ready hall in Kuala Lumpur, Jakarta, Bangkok or Ho Chi Minh City. A 142 kW liquid-cooled rack needs a purpose-built hall — which is exactly why the confirmed deployments in §6 cluster in places that can build one (Riyadh, Noida, Chennai, Jakarta's Zankore campus, Taiwan, Seoul) and why the only vendor quoting 3–7 day rack delivery in the entire study is the one holding local stock in Turkey. **Also check your colocation contract for a minimum-committed-power clause**: partial utilisation of a 142 kW rack typically still bills the full envelope, which is the single biggest driver of the rack-versus-node decision in §30.

30 · Verdict — buy, rent, or wait

The short answer. If you need B300 capacity in the next 2–4 months, **rent first and buy the node, not the rack**, unless you are an anchor tenant building a purpose-built liquid-cooled hall. If your workload can wait 12–18 months, **wait on Rubin and rent B300 in the meantime** — the generation handover is priced and dated, and B300 supply will not loosen before FY2028. If you are a sovereign or enterprise buyer who can absorb a 3-year commitment, **buy the rack**, but only with a price-review clause and a delivery-date remedy.

Buy, rent, or wait — the decision table

YOUR SITUATION	DO THIS	WHY	WATCH OUT FOR
Need capacity in < 8 weeks	**Rent**	Cloud B300 is live at $5.00–$9.50/GPU-hr across Singapore, Taiwan, Vietnam, India, Japan and Korea	Annual contracts at committed rates are 30–60% below list — never sign a 3-year deal in month one
Steady demand, 20–200 GPUs, no liquid-cooled hall	**Buy 8-GPU nodes**	Drop into an existing densest-in-a-hall facility; ~8–14 week lead time; modular scaling	Idle GPUs in a node are fully billable — the node is only efficient if your fleet runs near-full
500+ GPUs, purpose-built liquid-cooled hall, anchor tenant	**Buy GB300 NVL72 racks**	$51–56k per GPU in a rack vs **~$71k** for a delivered 8-GPU node (or ~$61k for the bare baseboard), and rack-tuned $/M-token is ~$0.0044 vs $0.0065 untuned B300	142 kW per rack is ahead of most Asian halls; the only 3–7 day rack delivery in this study was Turkish local stock
Need capacity but workload is experimental	**Rent 12 months, then decide**	Rubin production shipments began Aug 2026 and Rubin is quoted at 35× lower token cost than GB300	Do not sign 3-year flat rate on Blackwell. NVIDIA's own roadmap devalues it
Gulf buyer (UAE / Saudi)	**Buy via the licence-free route now; renegotiate before Apr 2027**	UAE is Country Group A:5 from 10 Jul 2026 and G42/Core42/HUMAIN Cloud are named in 15 CFR 740 Supp. 8	**That Approved Entity status expires after 270 days (~6 Apr 2027)** unless the entity becomes a "U.S. company" or obtains a BIS Advisory Opinion. Build the fallback into the contract, not the footnote
Buyer in China / HK / Macau	**Do not buy**	No licence is issued; the mainland B30A SKU is blocked	There is no compliant route. Any price is irrelevant

Break-even arithmetic — node versus rack

Model, stated so you can re-run it: **4-year amortisation**; node at **$570k** for 8 GPUs; rack at **$3.99M** for 72 GPUs; facility cost **$48k/yr** (node) and **$445k/yr** (rack) from §29; reference rental at the **38-provider median $9.50/GPU-hr** and at the **Asian on-demand floor $5.00**. **Caveat: the $570k node input is a third-party listing, not a vendor-confirmed price** (§26) — re-run the model with your own quotation before acting on the percentages.

RENT REFERENCE	NODE ANNUAL COST	NODE MAX REVENUE	NODE BREAK-EVEN	RACK ANNUAL COST	RACK MAX REVENUE	RACK BREAK-EVEN
$9.50 (global median)	$190.5k	$666k	**29%**	$1,442k	$5,993k	**24%**

RENT REFERENCE	NODE ANNUAL COST	NODE MAX REVENUE	NODE BREAK-EVEN	RACK ANNUAL COST	RACK MAX REVENUE	RACK BREAK-EVEN
$6.00 (committed band, e.g. KONST)	$190.5k	$420k	**45%**	$1,442k	$3,783k	**38%**
$5.00 (Asia on-demand floor)	$190.5k	$350k	**54%**	$1,442k	$3,154k	**46%**

> **The cash break-even is low; the practical threshold is much higher.** A 29% cash break-even is not a reason to buy a node — it assumes you can sell the idle hours and it ignores that an idle GPU inside a ~14 kW node still draws power and still bills. The number that matters for the **rack** is operational, not financial: you cannot shed one GPU from a 72-GPU rack, and the colocation contract bills the full 142 kW envelope regardless. **A rack is a safe commitment only when you can genuinely keep roughly 84% of it sold or internally utilised** — below that, you are paying rack-rate facility cost for node-rate demand, and nine 8-GPU nodes ($5.13M) will beat one rack ($3.99M) on capital while letting you right-size. The node's practical threshold is correspondingly low: you can start one and add more, so **~35% sustained fleet utilisation** is the point where owning nodes clearly beats renting them.

- **Rent-vs-buy at the node level:** one node at $570k, amortised over 4 years with facility, costs $190.5k/yr. Rented at $9.50/GPU-hr it must bill 29% of 8 × 8,760 = 70,080 GPU-hours to break even — roughly 20,100 hours a year of saleable time. At $5.00 it must sell 38,100 hours, i.e. 54% utilisation. **Under $5/GPU-hr the maths turns decisively against owning anything.**
- **Rent-vs-buy at the rack level:** one rack at $3.99M costs $1,442k/yr all-in. At $9.50/GPU-hr it needs 24% of 630,720 GPU-hours; at $5.00 it needs 46%. But a partially used rack still bills 142 kW, so treat 46% as a floor, not a target.
- **Time is the bigger risk than price.** NVIDIA says supply stays a bottleneck **through the end of FY2028**, and the observed market behaviour is *price rises* — Nebius raised B300 21% on 1 Oct 2026 and the 90-day median is +14%. Waiting does not get you a cheaper B300. What waiting gets you is **Rubin**, and you cannot buy Rubin in volume either — which is precisely why the sensible path for most buyers is *rent now, migrate later*.

31 · RFQ checklist — what to send, and what to make them answer

Send this list with the first request for quotation to every vendor in §7–§24. A vendor that cannot answer items 1–5 in writing is not quoting you the product you think they are.

1. **Exact SKU and bin.** "HGX B300" is ambiguous in the listings: 288 GB or 270 GB HBM3e; 1,400 W or 1,100 W per GPU; 108 or 144 PFLOPS dense NVFP4. NVIDIA's own spec table says **2.1 TB total (270 GB × 8)**, so demand the part number, the NVIDIA bin, and the HBM stack part number in writing. *Three different "B300s" are being sold today.*
2. **Numeric path of your workload.** Ask which precision your workload runs on and confirm the vendor has measured it on this SKU. **B300's INT8 is 3 POPS (B200: 72) and FP64 is 10 TFLOPS (B200: 296)** — if your model is INT8-quantised or FP64, this is a downgrade and the vendor must show you the measurement.
3. **Baseboard or system?** Ask whether the quote is the bare HGX B300 baseboard, an OEM server built around it, a DGX B300, or a GB300 NVL72 rack. Then ask for chassis, CPU (count, model, PCIe generation), RAM, local storage, NIC count and topology, and warranty start date — itemised, not as a lump sum.
4. **Allocation quarter and delivery date, in writing.** Ask for the allocation vintage, the ship date and the date hardware lands at destination. Quote 8–14 weeks for B300 systems, 4–8 weeks for GB300 where regionally stocked — anything materially faster is either grey-market or not real.
5. **Fabric and scale-up topology.** NVLink 5 within the node (14.4 TB/s), ConnectX-8 at 800 Gb/s per GPU east–west, and for racks the NVLink switch fabric and the 130 TB/s scale-up domain. Ask which switch generation and whether the fabric is included or customer-supplied.
6. **Support and NVIDIA entitlement.** 24/7 4-hour response, on-site SLA, and confirmation that NVIDIA support entitlement follows the hardware (ask them to state it explicitly — in the grey market it does not). Spare-parts lead time and depot location.
7. **Export and import compliance pack.** Ask them to confirm, in writing: ECCN 3A090.a; the destination country group and whether the sale is licence-exempt or licence-backed; the end-user screening they will perform; and whether they will supply the end-user statement, ScreenSafe / certificate of conformance, and commercial invoice with the correct valuation. For the Gulf, ask them to state **which licence pathway they are using**, whether the buyer is relying on

15 CFR 740 Supp. 8 Approved Entity status, and **what happens after that status expires (~6 Apr 2027) unless the entity becomes a "U.S. company" or obtains a BIS Advisory Opinion**.

8. **Local import pathway.** Name the exact permit you are applying for and who files it: Malaysia AICA (≥30 days, admin.sts@miti.gov.my) where brokering or re-export is involved; Korea MOTIE import-purpose confirmation (~7 days) plus the 1 Sep 2026 strategic-goods revision; Taiwan ITA import permit (1–10 working days); India DGFT Import Management System restricted-category licence (window closes 15 Dec 2026); Thailand customs/BOI; Indonesia TIU/TDP import permit; Vietnam MIC ICT permit; Japan — confirm no import approval is required for HS 8471/8542.
9. **DDP landed cost, itemised.** Ask for separate lines: ex-works unit price; freight and insurance; customs duty at the stated HS rate; import VAT/GST/IGST; broker and handling; port/demurrage; on-site install and commissioning; and the total. This is the only way to catch an Indonesian 5% duty hiding inside a "landed" price.
10. **Power and facility fit.** Confirm the vendor's stated IT draw (~14 kW node; ~142 kW rack), cooling type, and the minimum committed power and rack-space requirement they will contract for. Ask the colo separately whether your rate is power-inclusive, and get the minimum-committed-power clause in writing.
11. **Term, price review and exit.** Ask for 12 / 24 / 36-month pricing separately, whether there is a price-review or benchmark clause, whether the rate is tied to a published index, and what the exit/termination terms are. The H100 precedent (list fell 60–77%, then reserved pricing rose 40%) means a 3-year flat rate is a directional bet.
12. **Performance commitment.** Ask for the measured figure on *your* workload: tokens/s per GPU at your target context length and batch size, on what serving stack and version (vLLM, Dynamo, SGLang), and whether the numbers are per-node or per-rack. Reject rack-level headline numbers quoted against a node SKU.
13. **Reference and payment terms.** Ask for one customer reference in your own market, and for payment terms against a documentary credit, escrow through a neutral third party, or staged payment against delivery and acceptance.

32 · What in this dossier is unreliable

Stated plainly, because the most expensive errors in this market come from treating a circulating number as a quote.

FIGURE	STATUS	WHY, AND WHAT TO USE INSTEAD
"B300 costs $550k" / "~$350k per system"	REJECT	No attributable source. **Use $485k (bare baseboard floor) and the $775k Thai reseller system reference** from §26 — but note the $570k / $689k figures in that table are themselves third-party listings, not vendor-confirmed UNVERIFIED
Supermicro €559k / $570k and GIGABYTE $689k hardware listings	UNVERIFIED	Could not be re-confirmed on the vendors' own pages in this pass, and Supermicro's own EU store shows a comparable 4U Blackwell system at €59,326. **Do not budget from these numbers** — get a written, dated quotation
"288 GB per GPU"	LISTING FIGURE	NVIDIA's own HGX table gives **2.1 TB total (270 GB × 8)**. The 288 GB figure appears in Supermicro, KONST and several cloud price cards. Treat 2.1 TB as the design figure and demand the memory part number
"No MLPerf submission for an 8-GPU B300 node"	SOFTENED	Restated in this edition as "no published node-level submission was located". MLPerf v6.0 does list Blackwell Ultra systems; the absence of a located node-scale result is not proof that none exists
"INT8 and FP64 carry over from B200"	REJECT	They do not. NVIDIA's own table: **INT8 72 POPS → 3 POPS (24× cut); FP64 296 TFLOPS → 10 TFLOPS (30× cut)**. Confirm the numeric path of your workload in the RFQ
Lenovo / Alibaba marketplace listing	REJECT	Marketplace listings are not quotes; often stale, sometimes a deposit or a single component. Use the OEM's own product and price page, or a written quotation

FIGURE	STATUS	WHY, AND WHAT TO USE INSTEAD
China system price "¥7.0m" vs "¥18.9–19.6m"	REJECT	A four-week move larger than any credible price change, on an **inherently non-compliant** channel. Narrative, not marks. Do not use to benchmark an Asian price
"~¥7m channel cost" for cross-border transfer	UNATTRIBUTED	No source. Treat as an order-of-magnitude warning about grey-market friction, not a number
38-provider median $9.50, +14% in 90 days	MAJORITY	Scraped from public price cards: mixes SKUs, regions, term lengths and on-demand vs reserved. **Directionally reliable, not a quotable price**. Use as a negotiating anchor only
Floor / spot prices on cloud marketplaces	MAJORITY	Clearance pricing from a thin market. Useful as a lower bound for walk-up; irrelevant to a 12-month contract
PUE 1.3 applied to a liquid-cooled GB300 rack	REJECT	Liquid racks run 1.1–1.2; applying an air-cooled PUE misstates both the facility cost and the break-even. If colo bills power-inclusive, the electricity line is not yours at all
"HUMAIN 18,000 GB300"	MAJORITY	Superseded. Current primary disclosure: **84 MW / ~27,000 GPUs — Phase 1 ≈ 6,000 HGX B300 on Quantum-X800, 70 MW Vera Rubin from Jun 2027**
"14.5 kW per node"	RESTATED	NVIDIA's DGX B300 specification says **~14 kW**; the 14.5 kW figure used in the first edition was an unattributed rounding. **Design electrical infrastructure against NVIDIA's facility reference: 820 kVA average / 1,228 kVA peak for a 16-DGX-B300 configuration**
NVIDIA "$0.12 per M tokens"	MISLEADING	Rack-level GB300 compared against H200 — not an 8-GPU node figure, and no **published node-level** MLPerf submission for an 8-GPU B300 was located. Use the per-SKU table in §4
"15 PFLOPS dense NVFP4 per GPU"	MAJORITY	NVIDIA's blog number. The HGX product page implies 13.5 (108 ÷ 8). **Plan on 13.5 for HGX**
Colocation $/kW-month and electricity $/kWh bands	UNVERIFIED	Indicative market ranges, not quotes, and 2026 tariff and FX moves have been large. **Replace with a live colo RFQ before signing**
Lead times and "in stock"	MAJORITY	"In stock" means regional stock, not your configuration, and a lead time is a guide until it is a contractual delivery date with a remedy

Confidence statement. Deployments (§6), regulatory changes (§27) and the price board (§25) rest on named, dated, attributable sources and carry the highest confidence. **The hardware anchors (§26) do not** — the Supermicro and GIGABYTE figures are third-party listings that could not be re-confirmed on the vendors' own pages, and are flagged accordingly. Seller contacts (§7–§24) were read from each vendor's own published page and are accurate as of the date shown, but contact ownership and stock move fast — re-verify before the first call. The facility-economics bands (§29) and the break-even model (§30) are my own construction on stated assumptions: they are decision tools, not quotes. **What would change these conclusions:** the 270-day expiry of the 15 CFR 740 Supp. 8 Approved Entity status (~6 Apr 2027); the India's DGFT policy window closing 15 Dec 2026; the outcome of Taiwan's 2 Nov 2026 Foreign Trade Act consultation; any change to NVIDIA's allocation policy; and the arrival of Rubin at a price that re-prices B300 down faster than its utilisation recovers.

33 · Used, refurbished and secondary-market channels

Short answer: **there is no genuine used B300 market yet.** The part only reached volume shipping in 2026, is allocation-gated, and is still being absorbed by new-build orders. Every "used B300" listing located in this pass resolved to one of three things — a **cancelled-order allocation** that is effectively new, a **Chinese grey-market unit**, or an **outright fraudulent listing**. What does exist is the channel structure that will create the used market, and three of those channels can move real hardware today.

> **The listings that look like a bargain are bait.** Alibaba carries several "used HGX B300 8-GPU 288 GB" listings at **€946–€1,990** — for a part that lists at $485k–$775k. The same listings mix model names that cannot coexist (CMP 170HX is a mining card with no HBM and no NVLink; B300 has no PCIe variant), quote "32 GB" memory, and ship from unverified trading companies. At a genuine used discount of 5–15%, a real used node would still be $450k+. **Any complete 8-GPU B300 offered below roughly $400k is not a B300.** eBay carries the same pattern in smaller volume.

Who is in the channel

SELLER / CHANNEL	BASE	WHAT THEY ACTUALLY HAVE, OR WHAT ROLE THEY PLAY	CONTACT	STATUS
server-parts.eu	EU — ships worldwide	**Up to 128 × Lenovo ThinkSystem SR680a V4 with 8× HGX B300 NVL8 (1,100 W bin)**, air-cooled, 3-year Lenovo warranty, sourced from a **reported cancelled enterprise order**, offered for immediate shipment. **Eligibility: existing NVIDIA customer with a valid opportunity number.** Pricing on request	info@server-parts.eu server-parts.eu	**MAJORITY** offer page 22 Oct 2026
SK TES	Singapore HQ, global	ITAD and remarketing of decommissioned data-centre assets — acquires, tests and resells retiring infrastructure. The most likely first *legitimate* source of used B300 in Asia once the 2027–28 refresh starts	sktes.com	**CHANNEL** no B300 stock yet
Sims Lifecycle Services (ValidCore)	Global, Asia facilities	Recertified-GPU programme: tests, remanufactures and resells retiring GPUs with a warranty. The most productised route to a warrantied used accelerator	simslifecycle.com	**CHANNEL** no B300 stock yet
Maxicom	Singapore	GPU server **buyback** — buys retired NVIDIA and AMD servers from Singapore enterprises, labs and data centres. A seller to them, not a buyer from them, today	maxicom.sg	**BUYBACK**
HKCHL	Hong Kong	New and refurbished servers from an HK free-port warehouse; 1-year warranty, per-unit burn-in report, export documentation, 3–7 day air freight. Currently lists H100 / A100 / L4 — **no B300**	hkchl.com/contact	**DEALER** no B300

SELLER / CHANNEL	BASE	WHAT THEY ACTUALLY HAVE, OR WHAT ROLE THEY PLAY	CONTACT	STATUS
ANTTEC	Hong Kong	IT asset remarketing across Asia, Europe and the Americas — brokerage rather than stock	hkanttec.com	BROKER
GYS recyclers / 酷别深图	China	Advertise *回收B300服务器* — they **buy** decommissioned B300, they do not sell them. Useful only as a price signal on Chinese残值	gys.cn listings	DISQUALIFIED
Gigant	Moscow	120+ ASUS B300 9U systems, supply from Feb 2026, **from ₽59,192,401 excl. 22% VAT** — a parallel-import route around the licensed channel	gpu-server.gigant.ru	PARALLEL

Why used will not save you much

The one dealer publishing tracked secondary pricing — HKCHL — puts used H100 PCIe 80 GB at **≈$28,000 against a new band of $25,000–$32,200**, and states plainly that "used H100 pricing barely undercuts new — demand still outruns secondary supply". The mechanism is structural: HBM cost sets a floor, and every retiring current-generation GPU is bid for by someone who cannot get allocation. **Plan on a 5–15% discount, not the 40–50% that used enterprise hardware used to carry.** A used B300 also forfeits the support entitlement unless the OEM confirms transfer in writing.

> **Used does not mean exempt.** The EAR has no second-hand carve-out: ECCN 3A090.a attaches to the *item*, not to the invoice. Re-exporting a used B300 from Hong Kong, Singapore or Malaysia to China is exactly as unlawful as exporting a new one, and a private sale does not move the liability to the seller. Grey-market and cancelled-order units frequently carry no NVIDIA opportunity number — which means no warranty, no RMA, no driver validation and no support path when a GPU drops out of the NVLink domain.

Checking a used unit before money moves

1. **Serial and part number against NVIDIA.** Confirm the GPU bin too: 1,100 W vs 1,400 W, and whether the listing says 270 GB or 288 GB per GPU (see §1).
2. **Provenance and licence trail.** Ask for the original end-user, the original export authorisation, and the country of first placement. A seller who cannot produce these is selling a compliance liability.
3. **Warranty transfer in writing** from the OEM (Lenovo, Dell, HPE, Supermicro) — not from the broker.
4. **Full burn-in on the actual workload**: HBM ECC telemetry over a sustained run, FP4 and FP8 stress at rated TDP, and an NVLink lane validation against a reference baseboard.
5. **Escrow or letter of credit** until the burn-in report is delivered — not a deposit to a trading company's personal account.
6. **Assume no support.** Price the deal as if the hardware is unserviced, because in most cases it is.

> **Bottom line.** If you need B300 capacity in the next two quarters, used is not a strategy — the only real secondary-shaped supply found anywhere is a cancelled EU allocation that still requires an NVIDIA opportunity number. If you are planning for 2027–28, register interest now with SK TES and Sims Lifecycle Services, because they will be the ones holding the first genuine used units, and they will price them against new, not against scrap.

34 · Where to put it — the data-centre side

The GPU is the easy half. The binding constraint is a hall that will take the power and the heat. A **~14 kW air-cooled node** drops into almost any AI-ready hall in Asia. A **142 kW liquid-cooled GB300 rack** drops into almost none — which is precisely why the rack buyers in §6 are the ones building their own campuses rather than leasing.

Three routes — pick one before you pick a vendor

ROUTE	WHAT YOU ARE BUYING	COST BASIS	WHO IT SUITS
Colocation	Floor space, power, cooling, cross-connects. You own the hardware and the depreciation	$130–200/kW-month in tier-1 cities, $60–110 in emerging markets (§29)	You have bought the node or rack and want it in-country. The default for enterprise buyers
Neocloud / GPU cloud	Capacity by the GPU-hour, no capex, no facility risk	$5.00–9.99/GPU-hr list (§25); 30–60% below list on committed terms	Anything under a 12-month horizon, or a workload you have not yet proven
Build or JV	The hall itself — shell, power, CDU plant	Greenfield capex, typically **24–36 months** from land to first rack	Sovereign, hyperscale, or anyone who needs more than ~5 racks of liquid cooling

> **The lead-time inversion is the thing most buyers get wrong.** A GB300 rack can be delivered in **4–8 weeks** from regional stock (§26). A purpose-built liquid-cooled hall takes **24–36 months**. Buyers who order the GPUs first and then look for a home end up paying for silicon that has nowhere to run, or accepting a retrofitted air-cooled hall at a derated clock. **In an Asian rack deal, the data centre is the long-lead item, not the GPU.** Secure the hall — or the neocloud contract — before you sign the hardware PO.

Operators that can actually host a B300 today

OPERATOR	MARKETS	WHAT IS VERIFIED, AND WHAT IT MEANS FOR A B300	CONTACT	STATUS
ST Telemedia Global Data Centres (STT GDC)	Singapore HQ; 12 markets incl. SG, MY, IN, TH, JP, KR, ID, VN, PH, UK	**2.3 GW design capacity across 12 markets**; one of four winners of Singapore's DC-CFA2 round; Johor campus planned to **120 MW** with **STT Johor 1 at 16 MW by end-2026**, built for **immersion and direct-to-chip** cooling; reported deployment of up to **9,000 GPUs** at its Noida campus	sttgdc.com	CONFIRMED rack-capable
Digital Realty	Singapore, Tokyo, Osaka, Seoul, HK, Sydney	One of four winners of Singapore's DC-CFA2 round (50 MW). Platform-scale liquid-cooled designs; the most likely route to a multi-rack Tokyo or Singapore deployment inside an existing operator's estate	digitalrealty.com	CONFIRMED SG award

OPERATOR	MARKETS	WHAT IS VERIFIED, AND WHAT IT MEANS FOR A B300	CONTACT	STATUS
Equinix	20+ Asian metros	The deepest interconnection footprint in Asia and one of the four Singapore DC-CFA2 winners. Best fit for **node-scale** deployments that need carrier density more than raw megawatts	equinix.com	CONFIRMED SG award
Keppel Data Centres	Singapore, Johor, Batam	The fourth Singapore DC-CFA2 winner. Long-standing AI/high-density operator with its own design-and-build arm	keppel.com	CONFIRMED SG award
Bridge Data Centres	Malaysia (Johor, Cyberjaya), India, Thailand	**1.8 GW** of Malaysian capacity across Johor and Cyberjaya; its **MY07** hall was built with Vertiv specifically for liquid-cooled high-density AI (announced May 2026)	bridgedatacentres.com	MAJORITY liquid build
KONST	Jakarta, Cyberjaya, Singapore	Leases **15 MW of liquid-cooled Tier III colocation at SMX01**, Jakarta (Sinar Mas / LG Sinar Mas JV) — and separately operates **1,024 B300 GPUs at Cyberjaya** (§6). The only operator in this study that has bought, operated *and* housed B300	konsttech.ai	CONFIRMED both sides
NTT	Japan, Singapore, India, HK, and global	**250 MW+ of liquid-cooled capacity in service globally**; **Osaka 7** is one of Japan's first NVIDIA DGX-Ready facilities. Containerised DLC campus at Ishikari, Hokkaido (JPDC AI Container Village) targeting early 2027	group.ntt	CONFIRMED rack-capable
Cyfuture	India (multiple sites)	**240 kW per rack direct-to-chip**, 10 MW liquid-cooled hall live, with 20 MW and 100 MW campuses planned inside an SEZ enclave for duty-free hardware import. Explicitly engineered for Blackwell and Vera Rubin densities	cyfuture.com	MAJORITY vendor claim
Yotta	India (Navi Mumbai, Greater Noida)	**40,000 GB300 planned at NM2 Navi Mumbai** plus 40,000 Vera Rubin at D4 Greater Noida — and the operator behind the IndiaAI portal pricing in §25. The single largest B300-adjacent facility commitment in India	yotta.com	MAJORITY announced
Samsung SDS	South Korea	Placed POs for **2,016 B300** under the MSIT AI Highway programme; operates the host halls as well as the cloud layer	samsungsds.com	CONFIRMED 2,016 GPUs
Global Switch · Khazna · Princeton Digital Group · Fujitsu · Cyxtera	Asia-wide; UAE; SG/JP/IN/ID	Named on NVIDIA's own **DGX-Ready Colocation** partner list for Asia — the vendor's certified set for air-cooled and liquid-cooled DGX hosting. **Start here if you want a hall NVIDIA has already validated**	nvidia.com/.../colocation-partners	CONFIRMED

The 200 MW that decides Singapore

Singapore stopped handing out new data-centre megawatts for years; the second allocation round (DC-CFA2) released **200 MW and split it equally — 50 MW each — between Digital Realty, Equinix, Keppel Data Centres and ST Telemedia Global Data Centres**, chosen from more than 20 proposals. Practically: **those four are the only route to new Singapore capacity until the next window opens.** Everyone else selling "Singapore" capacity is reselling existing stock, or is actually in Johor or Batam. If your latency budget genuinely requires the island, the queue is four names long.

What to ask a colo that claims to be B300-ready

1. **Committed kW per cabinet, not average.** Ask for the number in the contract, not the brochure. A "40–60 kW hall" cannot take a 142 kW rack.
2. **Is the rate power-inclusive?** Many Asian contracts bundle a committed kW *including* electricity — which deletes the $/kWh line and changes the whole model in §29. Ask which one you are signing.
3. **CDU loop terms.** Is the coolant distribution unit inside the base rent or a separate capital line? Supply/return temperature, flow rate per cabinet, and who owns the manifold.
4. **Redundancy at the CDU, not just the UPS.** On CDU failure, does the rack throttle or trip? N+1 on the loop is the question, and most halls answer "no".
5. **Peak envelope, not just continuous.** 142 kW is average; ask what the facility can absorb on a 1,228 kVA-style transient (§29).
6. **Ramp and minimum commitment.** Can you take 2 racks in year one and 8 in year two without renegotiating the whole contract?
7. **A reference site you can visit.** Not a rendering — a hall with a live liquid-cooled GPU deployment. If there isn't one, you are the pilot.
8. **Temperature SLA.** Uptime SLAs are standard; a supply-temperature SLA is what actually protects your clocks.

> **Bottom line on the data-centre question.** For an **8-GPU node**, the hall is not your constraint — almost any AI-ready facility in Kuala Lumpur, Jakarta, Bangkok or Ho Chi Minh City will take 14 kW at $70–110/kW-month, and the scarce item remains the GPU allocation. For a **GB300 NVL72 rack**, the hall *is* the constraint: in Asia today that means STT (Johor, Inzai), Bridge (Johor MY07), KONST (Jakarta SMX01), Cyfuture or Yotta (India), NTT (Osaka) — or building your own, which costs two years you may not have. Sequence it correctly: **hall or neocloud contract first, hardware PO second.**

SOURCES

Regulatory — United States. Bureau of Industry and Security, Export Administration Regulations, ECCN 3A090.a; 15 CFR 740 Supplement No. 8 (Approved Entity authorisation for G42, Core42, UAE government agencies and HUMAIN Cloud, **expiring after 270 days** — published at 91 FR 43038); Federal Register document 2026-14132 and the BIS press release (UAE removed from **Country Groups D:3 and D:4** and added to A:5, effective 10 Jul 2026).

Regulatory — Asia. Malaysia MITI Directive 1/2025 (export, transit, transshipment and brokering; AICA form to admin.sts@miti.gov.my, ≥30 days; KLIA enforcement reported 2026). South Korea MOTIE strategic-goods revision effective 1 Sep 2026 (3A903 → 3A501.a.16 TPP ≥ 6,000; 4A902 → 4A507; new 3A502.i; 3B901–3B904 → 3B501–3B504; Exporter's Written Pledge abolished; Article 18-3 due-diligence; import-purpose confirmation ~7 days). Taiwan MOEA Bureau of Industry and Trade import permit (1–10 working days) and Foreign Trade Act consultation opened 2 Nov 2026. India DGFT Import Management System, HSN 8471 Restricted, policy window 22.12.2025 – 15.12.2026. Japan METI import-approval list (no HS 8471 / 8542 entries). Indonesia HS 8471.50 duty 5% + 11% VAT. Thailand 7% VAT to 30 Sep 2027. Vietnam 0% + 10% VAT with AI-chip land-use and rent exemptions. Philippines ATIGA 0% + 12% VAT. Bangladesh HS 8471.50.90 exemption (~9.5% all-in).

Regulatory — West Asia. UAE Country Group A:5 and Supp. 8 authorisation above; GCC Common External Tariff 5%. Saudi Arabia 15% VAT. Israel 18% import VAT and Ministry of Economy dual-use licensing. Turkey 20% KDV with Ministry of Trade / Ministry of National Defence / TAEK strategic-goods licensing.

Product and vendor. NVIDIA HGX B300 product page and DGX B300 / GB300 NVL72 product and facility documentation; NVIDIA Q2 FY2027 results (revenue $96.221bn, Data Center $89.0bn) and Q3 FY2027 outlook (Rubin ≈20% of Data Center revenue, supply bottleneck through FY2028); NVIDIA Partner Network Elite Solution Provider listings and NVIDIA Marketplace distributor/Cloud Partner directories; NVIDIA Asia-Pacific and EMEA networking-reseller pages; OEM product and price pages for Supermicro, GIGABYTE, Dell, HPE, Lenovo, ASUS, Supermicro, Foxconn, Quanta, Pegatron and Wiwynn; supplier and integrator pages for the ~90 sellers listed in §7–§24.

Price and performance. NVIDIA Marketplace cloud price cards and published on-demand pricing for the 38-provider reference set; Nebius pricing page and pricing notice effective 1 Oct 2026; published price cards for Gata Cloud, Clore.ai, Starseeds, KONST, Computeify, FPT AI Factory, TAC, GPU.ai, MTC, SuperX AI, Together AI, NTTドコモビジネス, Core42 and Samsung SDS; the **live IndiaAI Compute portal price list** (B300 SXM ₹351.0 / ₹342.0 / ₹333.0 / ₹319.5) and Shakti Cloud's bare-metal pricing page; MLCommons MLPerf Inference v6.0 results; vLLM, NVIDIA Dynamo and SGLang benchmark publications; AMD MI355X vendor comparisons; InferenceX deployment write-ups.

Deployments and filings. DataSection investor-relations disclosures (11 Dec 2025 order with Inventec, US$521M / 10,000 B300 / Sydney; 7 May 2026 order with Compal, US$325M / 5,080 B300 / Chiba); HKEX listing documents for Unity Group and XMax; MSIT AI Highway award announcements; company announcements for Yotta, HUMAIN, G42/Stargate UAE, Foxconn/Big Innovation, YCO Cloud × Aolani, Zankore, Bitdeer AI, KONST, GMO Internet, Giga Computing, SuperX, Unity Group, Together AI / L&T Vyoma.AI, and the Financial Times reporting on China grey-market DGX B300 pricing.

Secondary market. server-parts.eu Lenovo ThinkSystem SR680a V4 B300 allocation notice (22 Oct 2026); SK TES and Sims Lifecycle Services (ValidCore) service pages; Maxicom Singapore GPU buyback page; HKCHL refurbished-server catalogue and its tracked H100 asking-price data (Sep 2026); ANTTEC company page; STCN 算力租赁市场 report (Nov 2026) for Chinese B300 pricing; Gigant (Moscow) ASUS B300 stock page; Alibaba "used HGX B300 8-GPU" listings sampled Nov 2026 and rejected as non-credible.

Data centres. NVIDIA DGX-Ready Colocation Data Centers partner directory (Asia listings for Global Switch, Khazna, Princeton Digital Group, Fujitsu and Cyxtera); Data Center POST reporting on Singapore's DC-CFA2 award (200 MW provisionally allocated at 50 MW each to Digital Realty, Equinix, Keppel Data Centres and ST Telemedia Global Data Centres); STT GDC newsroom releases on the 50 MW Singapore AI-ready development, the KKR-led acquisition (2.3 GW design capacity across 12 markets) and the Johor campus; MIDA project brief for STT Johor 1 (16 MW by end-2026 within a 120 MW campus, immersion and direct-to-chip cooling); Bridge Data Centres Malaysia location and liquid-cooling (MY07 with Vertiv, 22 May 2026) pages; KONST SMX01 Jakarta project page (15 MW Tier III liquid-cooled colocation lease); NTT newsroom and investor material on AIOWN and liquid-cooled capacity (250 MW+ globally), plus the Osaka 7 DGX-Ready reference and the JPDC AI Container Village at Ishikari; Cyfuture liquid-cooled AI data-centre pages (240 kW/rack D2C, 10/20/100 MW campuses); Yotta announcement coverage for 40,000 GB300 at NM2 Navi Mumbai and 40,000 Vera Rubin at D4 Greater Noida; Samsung SDS filings under the MSIT AI Highway programme.

Economics and modelling. Third-party cost-per-token analysis at userecurrence.ai (H200 8×, B300 8×, GB300 standard, GB300 + Dynamo, GB300 rack-tuned); CBRE Asia-Pacific data-centre and colocation cost indicators; national grid tariff publications and customs authority schedules for the jurisdictions in §28–§29. All break-even figures in §30 are this dossier's own model on the stated assumptions.

NVIDIA HGX B300 in Asia — sellers, stock, pricing, export controls and tax. Second edition, full re-verification, 7 November 2026. ~90 named sellers across 23 Asian and GCC markets · 18 confirmed B300/GB300 deployments · consolidated price board · per-market regulatory and tax matrices.

Prepared as a sourcing dossier for procurement and infrastructure teams. Product specifications are as published by NVIDIA and OEMs and are subject to change without notice; NVIDIA does not disclose B300 revenue separately. Deployment figures are drawn from company announcements, exchange filings and regulatory publications. Seller contact details were verified against vendors' own published pages and change frequently. Facility-cost bands and all break-even calculations are modelled on stated assumptions and are not quotations. Compliance guidance is commercial analysis, not legal advice — confirm every export-control and import position with your own counsel and the relevant authority before contracting. **Hardware procurement across every jurisdiction covered here requires an end-user statement and, in most cases, an export licence or exemption that must be confirmed in writing by the seller before you commit funds.**