PART 16 · SECOND EDITION · SEPTEMBER 2026 (FIRST PUBLISHED JULY 2026)

Trends, Roadmaps & the Future

The binding constraint on AI is no longer the chip you can buy — it's the megawatt you can energize, and the date it arrives.

THE QUEUE

>2,060 GW

Generation and storage waiting in US interconnection queues — over 1.5× the nation's installed capacity

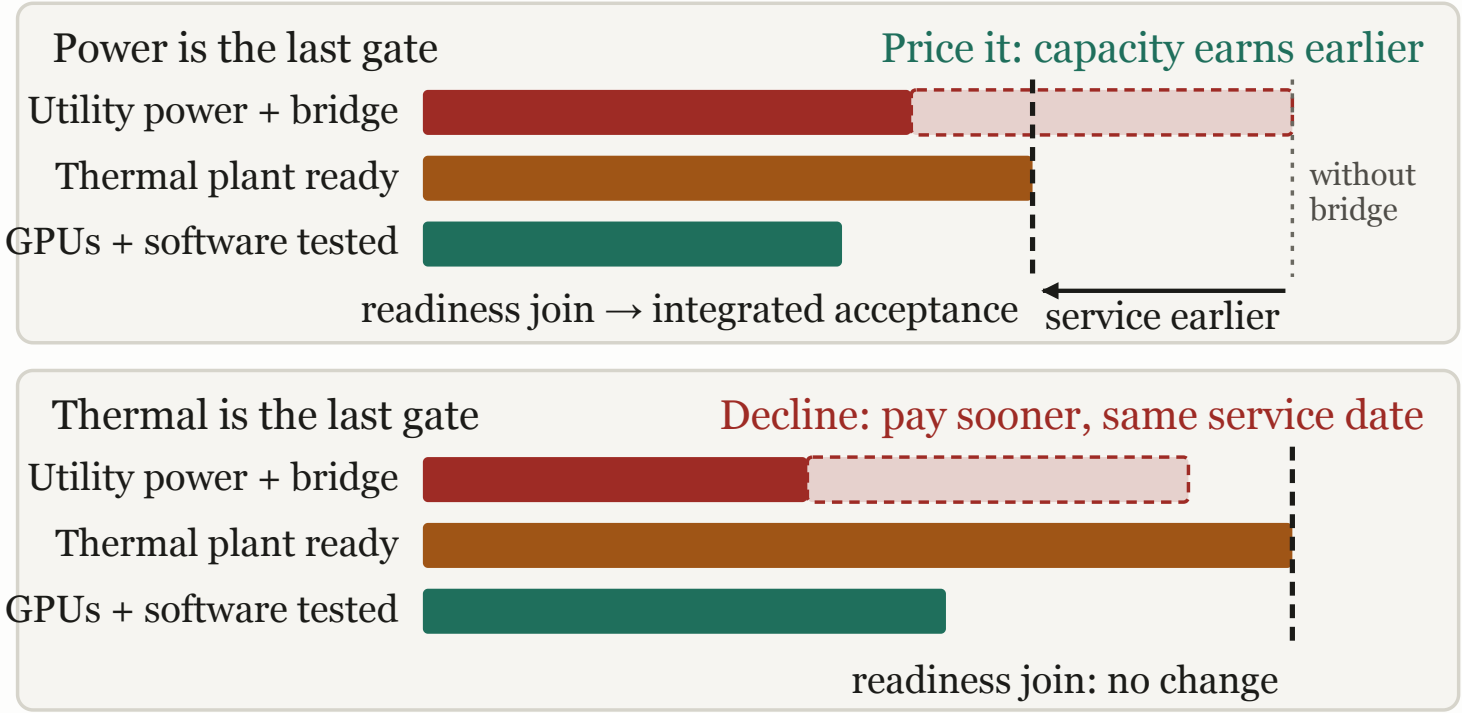
This is a generation-and-storage queue, not a data-center service clock. Read load delivery from the host utility's study and executed milestones.

LBNL, QUEUED UP: 2026 EDITION · AS-OF END-2025

THREE SEPARATE POPULATIONS **Generation/storage queues and data-center pipelines have different denominators**

Buy a bridge only if service moves earlier

An on-site power bridge pays only if power is the last thing ready.



time cut by the bridge · bar length = time to ready →

Schematic · Source: Guide derivation · as of September 2026

The >2,060 GW figure is US generation/storage queued for interconnection. Announced and under-construction data centers are separate populations, not downstream attrition stages.

THE CREDIBILITY TEST

≥16 GW vs ~5 GW

One tracker's announced and under-construction capacity sample

Announcement, financing, construction, energization and productive service are distinct states. The contested sample does not establish that two-thirds is fictitious.

SIGHTLINE CLIMATE · AS-OF 2026-02

ACT I

The power-bound era

Time-to-megawatt is the master variable. It subsumes the queue, the transformer, the bridge plant, and the price of skipping all three.

SUPPLY-SIDE CONTEXT — NOT A LOAD-SERVICE CLOCK

>2,060 GW

Active generation and storage in US interconnection queues at end-2025

This is generation and storage, not data-center load service. Build energization from the host utility tariff, studies, agreement, upgrades, equipment path, and dated milestones.

LBNL, QUEUED UP: 2026 EDITION · AS-OF END-2025

WHAT THE WAIT COSTS

~$12–13 B/GW-yr

A single-source revenue-per-GW estimate, not a phase-release rule

Gross revenue is not recovered margin. Price accepted useful service, ramp, timing and costs before treating earlier power as an economic win.

SEMIANALYSIS · AS-OF 2026-06

THE FOUR LEVERS

Four ways to compress time-to-power

LEVER	TIME-TO-POWER	WHAT IT COSTS YOU
Host-utility firm service	Project-specific tariff, studies, agreements, upgrades + equipment path	Often the lowest-cost supply; no national portable load-service duration
Behind-the-meter gas	Named turbine, fuel, permit and commissioning schedule	Combustion/fuel emissions: compare actual grid. Cleanview: ~2 GW operating, 1.2% building
Flexible / curtailable load	Only as studied and contracted by the host utility	Duke's modeled national headroom is not a site offer; price lost work, recovery and controls
Buy an energized site	Fastest — months	Premium for live power; scarce, bid-up, often the wrong shell

ACT II

Every subsystem climbs at once

Power, cooling, silicon, fabric, and storage are all replatforming on overlapping three-year clocks. The roadmap question is which irreversible substrate to over-build today.

THE DENSITY RAMP

The density ramp: named equipment bases, then a roadmap figure

132 kW nominal TDP: 115 kW liquid + 17 kW air

HPE GB200 NVL72: 132 kW nominal rack TDP

HPE, NVIDIA GB200 NVL72 BY HPE DATA SHEET, OCTOBER 2026 EDITION. · AS-OF 2026-10

188 kW (Max Q) / 228 kW (Max P)

Pegatron Vera Rubin NVL72: 188 kW Max Q / 228 kW Max P; 330 kW DSX facility basis

PEGATRON RA4803-72N3 DATASHEET · AS-OF 2026-08

~600 kW

Kyber-class rack (NVIDIA: H2 2027; SemiAnalysis: 2028) — per-rack planning figure

NVIDIA GTC · AS-OF H2 2027 ROADMAP (GTC 2025); 2028 REPORTED JULY 2026

THE LOCK-IN DECISION

Which scale-up fabric do you underwrite?

NVLink / Fusion

— Shipping at scale; deepest software stack

— 1.8 → 3.0 TB/s per GPU across generations

— Highest vendor lock-in — NVIDIA-centric

UALink / UALoE

— UALink Consortium: 1.0 public April 2025; current catalogue separates later component revisions

— Multi-vendor intent; interoperability needs proof

— Qualify scale, fault behavior and support on the offered system

ESUN / SUE-T (OCP)

— Ethernet-based scale-up; merchant silicon

— Rides the broadest supply chain

— Emerging; smallest proven domains

SO WHAT

A multi-year bet that sets accelerator optionality, your copper-vs-optics reach budget, and whether co-packaged optics lands on your 2027 critical path.

THE SUPPORTING CAST

Three roadmap numbers that reshape the building

~30W → ~9W

NVIDIA's 1.6 Tb/s interface: 30/9 ≈ 3.3×; NVIDIA now claims 5× lower power per CPO port

NVIDIA TECHNICAL BLOG, SCALING AI FACTORIES WITH CO-PACKAGED OPTICS FOR BETTER POWER EFFICIENCY, AUGUST 18, 2025, FIGURE 3: 20 W DSP + 10 W LASER VERSUS 7 W OPTICAL ENGINE + 2 W LASER, BOTH 1.6 TB/S. ROUNDED ENDPOINTS IMPLY ABOUT 3.3× · AS-OF AUGUST 18, 2025 VENDOR ILLUSTRATION

HBM3E → GB300; HBM4 → Vera Rubin

2026 HBM/CoWoS allocation map — packaging, not fabs, is the gate; no gap figure

SK HYNIX GTC 2026 ACCOUNT · AS-OF 2026

>55%

Reported cold-plate cooling forecast share; not a single-phase deployment census

PERSISTENCE MARKET RESEARCH, DATA CENTER LIQUID COOLING MARKET · AS-OF 2026

THE DISCIPLINE

Price hard-to-retrofit floor, water and riser options now; defer refresh-cycle equipment until the named system is accepted.

ACT III

Software eats the hardware gains

Software can release useful capacity between hardware generations. Demand absorbs the gain only when the workload and elasticity support it.

HISTORICAL PRE-TRAINING FIT

~8 months

Estimated halving time for compute at fixed pre-training capability

The eight-month estimate is not an installed-serving upgrade. Price validation, software work and quality/SLO acceptance before claiming a capacity gain.

EPOCH AI, ALGORITHMIC PROGRESS IN LANGUAGE MODELS · AS-OF 2024-03

THE DEMAND MULTIPLIER

Over 100× compute

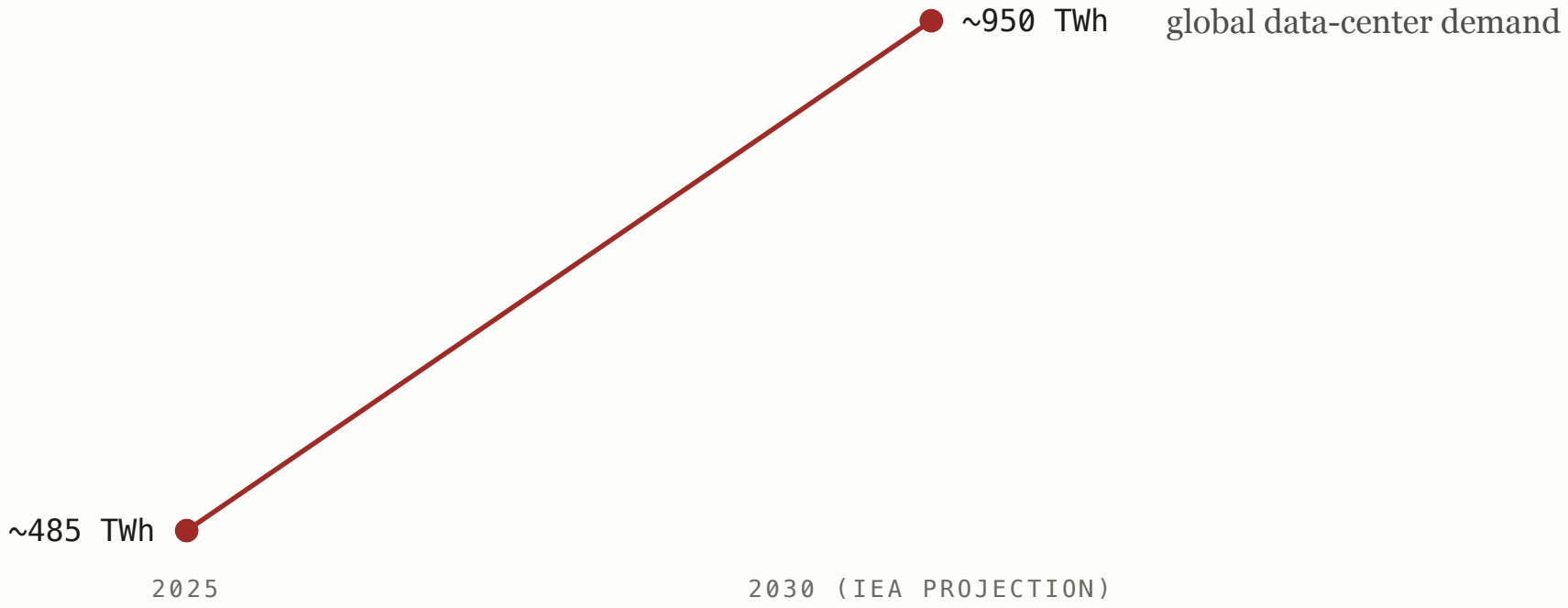
NVIDIA's February 2025 challenging-query compute example

Over 100× compute in the named challenging-query example. Long reasoning can favor decode, but context, batch and model change the bottleneck; this is not a fleet tokens-per-query factor.

NVIDIA, HOW SCALING LAWS DRIVE SMARTER, MORE POWERFUL AI, 12 FEBRUARY 2025. · AS-OF 2025-02-12

THE DEMAND CURVE

Data-center electricity roughly doubles by 2030



IEA's April 2026 scenario roughly doubles annual energy demand. Where local firm supply misses the ramp, phase capacity waits.

IEA, KEY QUESTIONS ON ENERGY AND AI · AS-OF 2026-04

ACT IV

The money, and the three futures

A multi-trillion-dollar bet, levered against the shortest-lived collateral ever financed at this scale.

THE CAPITAL STACK

The build-out, in three numbers

~$2.9T / ~$1.5T

Morgan Stanley's July 2025 capex and funding-gap forecast through 2028

MORGAN STANLEY RESEARCH, BRIDGING THE DATA CENTER GAP, 16 JULY 2025. · AS-OF 2025-07

$11.3M/MW

JLL's January 2026 shell-and-core forecast; technology is separate

JLL, 2026 GLOBAL DATA CENTER MARKET OUTLOOK · AS-OF 2026-01

~$1.5T lifetime per capex-year

Cahn's lifetime end-customer revenue requirement for one annual investment cohort

DAVID CAHN, AI'S $1.5T QUESTION, 8 JULY 2026 · AS-OF 2026-07

SCENARIOS FOR 2030

Three futures — and what each one strands

SCENARIO	CORE MECHANISM	WHAT STRANDS
Supercycle	Reasoning and agentic demand absorb every megawatt built	Latecomers stuck in the queue; under-provisioned power and cooling
Digestion	The extraordinary base gets absorbed; growth decelerates without a crash	Speculative greenfield; over-levered merchant capacity
Bubble	Revenue never arrives at scale; spend behaves like a utility	Long-lived debt on short-lived silicon; under-depreciated fleets

ILLUSTRATIVE SECTOR CASH SCREEN

$200B/year (assumed sector CFO)

Cash capex release depends on funding and the cash floor

Guide assumptions: the funding gap rises from $40B to $90B under CFO/capex stress. Lower utilization or price cuts project cash for debt: use Chapter 1.8's cash flows and Chapter 2.5's DSCR case.

THE AI DATA CENTER GUIDE, ILLUSTRATIVE TEACHING FIXTURE · AS-OF 2026-09-08

THE HONEST CAVEAT

Every headline number on the 2027–2030 horizon is a dated, contested forecast — not a fact.

The discipline is to keep the date and the caveat attached, and never let a forecast harden into the assumption behind an irreversible commitment.

TAKE IT HOME

What to do Monday

01 Reorganize the program around time-to-megawatt: the energization date is the root constraint, and the GPU fits the power envelope.

02 Separate announced, financed, energized and productive capacity; release phases against power, funding and accepted demand.

03 Price floor, water, voltage and riser options; defer refresh equipment until interfaces and software pass acceptance.

04 Stress the historical deflation fit, quality, SLOs and demand elasticity together; a serving gain requires acceptance.

05 Choose supercycle, digestion or bubble; price hedges for the other two, then run joint power, demand and residual shocks.

06 Keep the date and the caveat attached to every roadmap number you carry into a financial model.

THE DEFINITIVE GUIDE TO

AI Data *Centers*

This deck is the map. The guide is the territory — every number here is dated, sourced, and kept current in the full chapters.

Ch 16.1 The power-bound era and time-to-megawatt

Ch 16.2 Subsystem roadmaps, 2026 → 2030

Ch 16.4 The economics of the build-out

Ch 16.5 Scenarios for 2030 and what strands

READ THE FULL PART, WITH EVERY SOURCE AND DERIVATION, AT AIDATACENTERGUIDE.COM

SOURCES

Every figure in this deck traces to the guide's dated register (1/3)

LBNL, Queued Up: 2026 Edition (data through end-2025) · as-of end-2025 · https://eta-publications.lbl.gov/sites/default/files/2026-06/queued_up_2026_edition.pdf

Sightline Climate (now Currence), Data Center Outlook Q1 2026 · as-of 2026-02 · https://www.currence.ai/blog/data-center-outlook

SemiAnalysis ('To Boldly Go', Jun 2026) · as-of 2026-06 · https://newsletter.semianalysis.com/p/to-boldly-go-the-case-for-space-datacenters

Cleanview, Michael Thomas, Bypassing the Grid, public report; explicitly describes mid-2026 operating capacity and gives no publication day. Verification on September 8, 2026 does not create a source date. · as-of mid-2026 (publisher's observation window) · https://cleanview.co/reports/behind-the-meter-data-centers

HPE, NVIDIA GB200 NVL72 by HPE data sheet, October 2026 edition. · as-of 2026-10 · https://www.hpe.com/psnow/generateDDS/NVIDIA%20GB200%20NVL72%20by%20HPE%20data%20sheet-PSN1014890104MYEN.pdf

Pegatron RA4803-72N3 datasheet · as-of 2026-08 · https://svr.pegatroncorp.com/SpecFile/PEGATRON_RA4803-72N3_DS_2026v1.pdf

NVIDIA GTC (Jensen Huang); DCD, Tom's Hardware · as-of H2 2027 roadmap (GTC 2025); 2028 reported July 2026 · https://www.tomshardware.com/pc-components/gpus/nvidia-shows-off-rubin-ultra-with-600-000-watt-kyber-racks-and-infrastructure-coming-in-2027

SOURCES

Every figure in this deck traces to the guide's dated register (2/3)

NVIDIA Technical Blog, Scaling AI Factories with Co-Packaged Optics for Better Power Efficiency, August 18, 2025, Figure 3: 20 W DSP + 10 W laser versus 7 W optical engine + 2 W laser, both 1.6 Tb/s. Rounded endpoints imply about 3.3×; the article's separate 3.5× marketing claim is not that quotient. · as-of August 18, 2025 vendor illustration · https://developer.nvidia.com/blog/scaling-ai-factories-with-co-packaged-optics-for-better-power-efficiency/

SK hynix GTC 2026 account; Micron HBM4 production release · as-of 2026 · https://newsletter.semianalysis.com/p/ai-capacity-constraints-cowos-and

Persistence Market Research, Data Center Liquid Cooling Market (PMRREP13330, March 2026); repeated by Schneider Electric, 5 May 2026 · as-of 2026 · https://www.persistencemarketresearch.com/market-research/data-center-liquid-cooling-market.asp

Epoch AI, Algorithmic progress in language models · as-of 2024-03 · https://epoch.ai/publications/algorithmic-progress-in-language-models

NVIDIA, How Scaling Laws Drive Smarter, More Powerful AI, 12 February 2025. · as-of 2025-02-12 · https://blogs.nvidia.com/blog/ai-scaling-laws/

IEA, Key Questions on Energy and AI (Apr 2026) · as-of 2026-04 · https://www.iea.org/reports/key-questions-on-energy-and-ai

SOURCES

Every figure in this deck traces to the guide's dated register (3/3)

Morgan Stanley Research, Bridging the Data Center Gap, 16 July 2025. · as-of 2025-07 · https://www.morganstanley.com/content/dam/msdotcom/en/assets/pdfs/Research_Bridging-Data-Center-Gap.pdf

JLL, 2026 Global Data Center Market Outlook · as-of 2026-01 · https://www.jll.com/en-us/insights/market-outlook/data-center-outlook

David Cahn, AI's $1.5T Question, 8 July 2026 · as-of 2026-07 · https://dcahn.substack.com/p/ais-15t-question

The AI Data Center Guide, illustrative teaching fixture; unsupported assumptions, not a forecast for named companies. · as-of 2026-09-08 · https://aidatacenterguide.com/part-16-trends-roadmaps-and-the-future/16-4-the-economics-of-the-build-out

SOURCES

About these figures

Figures are resolved by reference from the guide's claim register at build time (Second edition · September 2026 (first published July 2026)) — a number appears here only as it appears, dated and sourced, at aidatacenterguide.com. Newer data may exist: check the live register.