AI Infrastructure Valuations: Q2 2026

Q2 2026 finds AI infrastructure, the compute and data substrate beneath artificial intelligence, in a state Windsor Drake characterises as a capital supercycle meeting a valuation reckoning. Hyperscaler capital spending is climbing to roughly **$725B** for 2026, up about **77%** year on year (Goldman Sachs), underwriting multi-year demand for everything from accelerated cloud to the data platforms that feed it. Yet the market has grown sharply more discriminating about how that demand converts into value.

The defining feature of the quarter is bifurcation. Capital-light software, the data platforms, inference engines and retrieval layers, commands software-plus multiples, while capital-intensive compute re-rates on forward revenue. Windsor Drake's working benchmark for public data-infrastructure software has settled near **11x EV/ Revenue**, but the spread around it is extreme: private AI-native data platforms such as Databricks trade near **25x** annualised revenue, while neoclouds such as CoreWeave sit near **6.5x** on forward revenue despite triple-digit growth. The discriminator is gross-margin quality and capital intensity, not the AI label.

The macro backdrop is constructive. The Federal Reserve held its funds range at **3.50% to 3.75%** in April 2026 and added an easing bias, with the March dot plot pointing to no more than one further cut in 2026. Consumption economics are doing the heavy lifting on the software side: Snowflake net retention of **125%** and Databricks net retention above **140%** turn AI adoption into compounding revenue. Capital markets have reopened in parallel, with Cerebras pricing at **$185** and surging on its May 2026 debut, and strategic consolidation has begun in earnest, exemplified by IBM's **$11B** acquisition of Confluent, completed in March 2026. Private capital is pressing in behind: PE invested about **$45.7B** in US data centers in 2025, a five-year high, and the **$40B** Aligned Data Centers sale set a record for the asset class.

This report sets out institutional-grade analysis for navigating that split market, one in which durable, consumption-led data and software infrastructure is valued like premium software while capital-heavy compute is priced on contracted backlog, margin and the cost of capital.

What multiples are AI infrastructure companies trading at?

The Q2 2026 valuation picture turns on a single divide: software and consumption economics on one side, capital intensity and resold compute on the other. The broad public data-infrastructure cohort clusters near Windsor Drake's 11x benchmark, but the gap between the top and bottom of the table is the widest the category has seen. Investors are paying up for recurring, consumption-led revenue, software gross margins and a defensible position in the AI stack.

Capital-light platforms and SaaS-style infrastructure are valued on light capital requirements and net revenue retention that compounds with AI workloads. Compute-intensive neoclouds, by contrast, are priced like contracted infrastructure: backlog quality and the cost of capital, not topline growth, drive the multiple, and depreciation and debt loads keep forward multiples in single digits.

Table 1. AI Infrastructure Valuation Multiples by Subsegment

Subsegment	EV/Revenue Range	YoY Trend	Primary Driver
AI-Native Data Platforms	18.0x - 25.0x	Rising	Consumption retention, data gravity
Inference & Model Serving	15.0x - 19.0x	Rising	Token-volume growth, managed demand
Vector & Retrieval (private)	12.0x - 20.0x	Volatile	RAG as a default AI layer
Observability & Data Ops	9.0x - 20.0x	Strengthening	AI telemetry, Rule of 40
Cloud Data Platforms	8.5x - 12.0x	Stable	Consumption, AI workload gravity
Streaming & Pipelines	8.0x - 11.0x	Consolidating	Real-time data for enterprise AI
GPU & Accelerated Cloud (forward)	6.0x - 10.0x	Compressing	Backlog versus capital intensity

Source: Windsor Drake analysis of PitchBook, CB Insights and S&P Global Market Intelligence data.

Subsegment dynamics driving the dispersion

Data-pipeline and lakehouse platforms sit at the top of the table because they monetise every downstream AI workload through consumption, the closest thing in AI infrastructure to a compounding software annuity. Inference platforms re-rate fastest of all, with Fireworks reaching roughly $800M and Baseten roughly $600M in annualised revenue within a few quarters, though GPU costs in cost of goods hold their gross margins near 50%. Neoclouds move the other way: hypergrowth is real, but capital intensity and debt compress the forward multiple, and improving chip access is beginning to erode the scarcity that once justified premium pricing.

Table 2. Subsegment Valuation Drivers and Principal Risks

Subsegment	Premium Driver	Principal Risk
AI-Native Data Platforms	Net retention above 140% and data gravity	Private-mark reset on an AI-spend pause
Inference & Model Serving	Explosive consumption growth	GPU-cost gross margins near 50%
Vector & Retrieval	Retrieval as a default AI layer	Feature-versus-product squeeze by incumbents
GPU & Accelerated Cloud	Contracted multi-year backlog	Capital intensity, debt and depreciation
Cloud Data Platforms	Consumption that compounds with AI	Slower growth at scale
Data Centers & Power	Contracted, bond-like cash flows	Power and interconnection constraints

Source: Windsor Drake analysis of McKinsey, Goldman Sachs and S&P Global Market Intelligence research.

How are AI infrastructure companies valued in 2026?

Valuation in 2026 has coalesced around a disciplined framework built on consumption durability, gross-margin quality and capital efficiency. The growth-at-all-costs playbook is gone. In its place is a multi-factor model in which net revenue retention and the Rule of 40 are table stakes, and capital intensity is an explicit discount.

The Rule of 40 mandate

The Rule of 40, where revenue growth plus EBITDA margin reaches at least 40%, remains the primary filter for a premium multiple across data and infrastructure software. In 2026, balance matters as much as the headline: data and infrastructure software clearing the threshold trades at roughly **2 to 3 times** the revenue multiple of those below it. AI-native platforms that pair high growth with consumption margins command the richest marks of all.

Table 3. Rule of 40 Performance Tiers

Performance Tier	Rule of 40 Score	Avg EV/Revenue	Premium vs Median
Top Quartile	Above 50	18x and above	+2x to +3x multiple
Rule of 40 Met	40 to 50	11x - 18x	Healthy premium
Near Miss	30 to 39	8x - 11x	Modest discount
Below Threshold	Below 30	4x - 8x	Deep discount

Source: Windsor Drake analysis of McKinsey and Bain & Company software value-creation research.

Consumption economics under scrutiny

Net revenue retention has become the defining quality signal for AI-infrastructure software. Snowflake reported **125%** net retention and Databricks net dollar retention above **140%**, meaning existing customers expand spend automatically as AI workloads grow. Usage-based pricing aligns vendor revenue directly with customer value, but investors now look past topline to the gross-margin spread: software at roughly 75% or better is valued far above resold compute near 50%, because only the former turns consumption into durable profit.

A credible path to margin

For any asset valued above ten times revenue, the market now expects a believable path to durable, software-grade gross margins and operating leverage. Inference platforms are the clearest test: their consumption is compounding, but GPU costs hold gross margins near 50%, so the winners must climb the software-margin curve to defend their marks. For capital-intensive compute, the equivalent test is the cost of capital: with more

than $20B of debt on the largest neocloud balance sheets, utilisation and financing cost, not growth, determine equity value.

What is driving AI infrastructure valuations this quarter?

Valuations in Q2 2026 reflect an interplay of powerful expansionary forces and real compressive pressures. Reading those drivers correctly is what separates a defensible valuation from a mispriced one, and in AI infrastructure the same force, AI demand, expands software multiples while exposing the capital intensity of compute.

Table 4. Valuation Drivers, Expansion versus Compression

Factor	Driver	Effect on Multiples	Notable Examples
Expansion	AI demand and capex supercycle	Premium for substrate exposure	$725B hyperscaler 2026 capex
Expansion	Consumption durability	Retention compounds revenue	Snowflake 125%, Databricks >140%
Expansion	Rate normalisation	Lower discount rates lift growth assets	Long-duration software and compute
Compression	Capital intensity	Forward multiples compress	Neoclouds, >$20B debt loads
Compression	GPU commoditisation	Resale margins erode	Commodity GPU cloud
Compression	Power and supply	Buildout and siting constraints	Data-center capacity

Source: Windsor Drake analysis of Goldman Sachs, McKinsey and Federal Reserve data.

Geographic variation

Location still matters for valuation, and AI infrastructure is the most geographically concentrated category in technology. North America dominates investment on the strength of hyperscaler capex, deep capital markets and the leading neoclouds. The fastest-growing pools of new capital, however, are sovereign: Gulf, European and APAC state-backed vehicles such as MGX are deploying record sums into compute and data centers, creating both buyers and partners for cross-border assets.

Table 5. Geographic Valuation Variation

Region	Investment Share	Posture	Key Drivers
North America	~60%	Premium	Hyperscaler capex, deep capital markets, leading neoclouds
APAC	~22%	Growth	Sovereign-AI buildouts and mobile-first scale
Europe	~13%	Sovereign	Data-sovereignty initiatives and state-backed capital
Rest of World	~5%	Selective	Gulf sovereign vehicles against higher political risk

Source: Windsor Drake analysis of Goldman Sachs and S&P Global Market Intelligence data.

Public and private markets diverge

One of the defining features of the quarter is that, unlike mature software, the AI-infrastructure private premium is widening rather than converging. Public data-infrastructure software has re-rated steadily to about 11x, while private AI-native leaders have run to 20-25x, widening the spread from roughly 4x in 2023 to about 10x today. The cause is scarcity: a thin supply of scaled, independent AI-native assets, led by Databricks at a $134B valuation, keeps private marks running ahead. Non-AI laggards, by contrast, are seeing flat marks and becoming prime candidates for strategic M&A.

Which valuation metric should apply?

Selecting the right metric is what separates a professional valuation from a careless one. AI infrastructure spans pure software, capital-intensive compute and physical real assets, and leaning on a single EV/Revenue multiple can badly misprice a neocloud or a data center.

EV/Revenue: the software and inference metric

EV/Revenue suits high-growth software with recurring, consumption-led revenue that is reinvesting ahead of profit, including data platforms, inference and retrieval. The essential adjustments are for net retention and gross margin: a dollar of software revenue at a 75%-plus margin is not comparable to a dollar of resold compute earned near a 50% margin.

EV/EBITDA and contracted backlog: the compute metrics

EV/EBITDA fits mature, cash-generative platforms and physical data centers, where contract quality and utilisation drive value. For neoclouds whose story is dominated by take-or-pay backlog, an EV/contracted-backlog lens, paired with the cost of capital, values the equity more honestly than a trailing revenue multiple, because depreciation and debt sit between revenue and cash flow.

Capacity and power: the physical-asset metric

Physical data centers are valued on contracted EBITDA, megawatts and power access rather than revenue. Long-term hyperscaler leases convert them into bond-like assets, and secured energy and interconnection, increasingly the binding constraint of the cycle, are now the primary source of scarcity value.

Table 6. Valuation Methodology Matrix

Subsegment	Primary Metric	Typical 2026 Range	Key Adjustment
AI-Native Data Platforms	EV/Revenue	18x - 25x revenue	Net retention, gross margin
Inference & Model Serving	EV/Revenue	15x - 19x revenue	GPU-cost gross margin
Vector & Retrieval	EV/Revenue	12x - 20x revenue	Category leadership, scale
Cloud Data Platforms	EV/Revenue	8.5x - 12x revenue	Consumption retention
GPU & Accelerated Cloud	EV/Revenue or backlog	6x - 10x forward	Backlog, debt, cost of capital
Custom Silicon	EV/Revenue	10x - 18x revenue	Concentration, scarcity
Data Centers & Power	EV/EBITDA	12x - 30x EBITDA	Contract quality, power access

Source: Windsor Drake valuation methodology, calibrated to PitchBook and CB Insights comparables.

Key takeaways for founders

Translating the market picture into strategy means concentrating on six areas that consistently move valuation in the current environment.

1. Make consumption retention the headline metric

Net revenue retention is the single strongest driver of premium AI-infrastructure multiples. Snowflake at 125% and Databricks above 140% set the bar. Instrument expansion revenue, separate committed from on-demand, and report net retention to the board monthly.

2. Protect gross margin

The market rewards software economics over resold compute, paying 2 to 3 times the multiple for the former. Where the model allows, move from raw GPU resale toward higher-margin managed services, optimise utilisation and reserved-capacity costs, and evidence margin durability with clean cohort data.

3. Contract the backlog

Multi-year, take-or-pay backlog is what underwrites premium compute valuations and makes an asset financeable; CoreWeave's roughly $100B backlog anchors its equity story. Convert usage into committed revenue, lengthen contract duration and reduce customer concentration before a process.

4. Clear the Rule of 40

Revenue growth plus margin remains the best predictor of a premium multiple, and clearing the threshold earns 2 to 3 times the multiple of peers. Balance consumption-led growth with a credible path to margin rather than pursuing either in isolation.

5. Weigh capital intensity and geography

Debt and depreciation are explicitly penalised: favour asset-light or contracted structures to protect equity value. Geography compounds the effect, with North America commanding a premium while export controls and sovereign-AI policy shape where cross-border value, and risk, accrue.

6. Prepare for a buyer market while it holds

With about $3.7T of dry powder and incumbents buying the substrate, the prepared asset captures the scarcity premium. A full process runs 12 to 18 months end to end, so a founder who intends to meet the market while the capex supercycle, reopened listings and abundant capital still align is, in practice, preparing in the current cycle.

Sources

- Goldman Sachs, AI capital-expenditure outlook and the AI build-out
- Morgan Stanley, financing the AI infrastructure build-out
- McKinsey & Company, Global Private Markets Report 2026
- McKinsey & Company, the cost of compute and data-center capital needs
- Bain & Company, Global Private Equity Report 2026
- S&P Global Market Intelligence, data-center M&A and private-capital analyses
- PitchBook, AI and data-infrastructure comparable-company data
- CB Insights, State of AI
- PwC, US Technology Deals 2026 Outlook
- EY-Parthenon, M&A Outlook 2026
- Federal Reserve, FOMC statement and Summary of Economic Projections
- Bloomberg, AI infrastructure funding and valuations