EVERCORE ISI

Technology | IT Hardware & Networking

July 06, 2026

Amit Daryanani
415-800-0185
amit.daryanani@evercoreisi.com
Irvin Liu
415-800-0183
irvin.liu@evercoreisi.com
Victor Santiago
212-653-9014
Victor.Santiago@evercoreISI.com
Hannah Liu
212-812-2905
hannah.liu@evercoreISI.com
Caden Dahl
212-653-9034
Caden.Dahl@evercoreisi.com

Primer on GPU-as-a-Service: The Infrastructure Layer For The Tokenomics Era

ALL YOU NEED TO KNOW: In this report, we examine the business model, underlying demand drivers, supply chain backdrop, financing requirements, and competitive landscape of the GPU cloud industry. Fundamentally, we view GPU clouds as a critical layer of the broader AI infrastructure stack that is benefitting from robust demand for AI compute (for both model training and inference, measured by token consumption). The unrelenting demand backdrop for AI compute combined with constrained supply (data center powered shell and GPUs/XPUs) should lead to multi-year AI infrastructure demand and pricing tailwinds as well. Given that the GPU cloud industry is somewhat nascent, providers are still in investment mode which means capex as a percentage of total revenue will remain high, operating margins will be depressed at the consolidated level as GPU cloud providers absorb ramp-up costs related to new capacity, and EPS/FCF will likely remain negative. Therefore, we view backlog/RPO, EBITDA, and capacity (measured in MW, both active and contracted/developable) as the more important near term metrics to focus on. Despite limited profitability at the consolidated level near term, it is important to note multi-year take or pay contracts are underwritten at attractive returns/contribution margins and as new capacity represents a smaller percentage of total looking ahead, EBIT margins should improve over time. While the vast majority of revenue for providers is currently toward take or pay contracts is derived from contracted capacity, longer term, we see potential for upside to models from: 1) GPUs that are leased beyond ~6-year accounting useful lives (EBIT margin would approach EBITDA margin in this scenario), 2) Higher GPU Hour pricing resulting from mix shift toward on demand, and 3) Attach of ancillary software/services with higher margins such as software, CPU, networking, and storage. *Net/net:* While we remain in the early days of the great AI infrastructure buildout, we see robust demand creating tailwinds for GPU cloud providers such as CRWV and suppliers across the stack (server OEMs, data center capacity providers, physical infrastructure suppliers and more). Although there have been new entrants in the GPU cloud space (SPCX, META, SoftBank, etc.), we think this is more a reflection of strong demand/pricing in the industry (not a sign of overbuilt capacity).

AI Compute Demand (Tokens): We view token consumption as a driver of demand for GPU cloud compute. In May, we introduced our Token Consumption Model to better frame the underlying AI infrastructure demand. In our base case, we see total annual token demand growing from ~100 quadrillion in 2026 to ~4.0 quintillion in 2030, representing a ~150% 4-year CAGR. Translating this to compute demand requires an assumption for throughput (TPS per MW) and assuming an average install base throughput of ~500,000 TPS/MW, we see demand for data center capacity reaching ~250 GW by 2030.

GPU Business Model: GPU cloud compute is typically sold on a per GPU hour basis via: **1) Reserved Capacity:** Multi-year reserved capacity with take-or-pay contract structure that is typically priced at a significant discount vs. spot/on-demand agreements. **2) On-demand:** Pay-as-you go structure that is priced on a per GPU hour basis typically at a premium vs. reserved capacity. Pricing is more volatile depending on supply/demand backdrop. Agreements could include an initial period of reserved capacity. **3) Spot:** For non-critical workloads, customers bid for unutilized capacity from neoclouds though capacity is not guaranteed and can be terminated by provider.

Please see the analyst certification and important disclosures on page 29 of this report. Evercore ISI and affiliates do and seek to do business with companies covered in its research reports. Investors should be aware that the firm may have a conflict of interest that could affect the objectivity of this report. Investors should consider this report as only a single factor in making their investment decision.

Where GPU Clouds Sit in AI Infrastructure Stack: GPU cloud providers sit between their suppliers (providers of DC capacity, technical infrastructure such as server/storage/networking OEMs and silicon providers) and their AI customers (frontier model labs, hyperscalers, enterprises, etc.). From a per megawatt math/tokenomics perspective, GPU cloud providers procure data center powered shell capacity at ~$2M/MW annually (mid to high $100/kW per month) and add value by combining technical infrastructure with physical infrastructure to create clouds, and then monetize those megawatts at a multiple (>5x or $10-12M/MW assuming mid-$2 type per GPU hour pricing) of what they are paying data center operators. Frontier AI LLM companies then monetize that compute capacity at a multiple of what they pay GPU cloud providers depending on throughput (TPS/MW) and what they charge customers per million tokens. Assuming LLM customers pay an average of $10-12M per MW of GPU cloud capacity annually, and they achieve a throughput of ~500,000 TPS/MW, the cost per million tokens (Mtok) could be $0.60-0.70. If we then assume they then monetize each Mtok at close to ~$3 on average, AI LLM companies are monetizing their LLMs at a ~4x multiple of what they are paying GPU cloud providers (>$40M per MW vs. $10-12M in annual cost for GPU cloud capacity and ~$2M in annual cost per MW for DC capacity). The value-add multiplier going from DC capacity to GPU cloud and GPU cloud to AI LLM creates a compelling case for broader infrastructure demand.

Capex/Supply Chain: GPU clouds are capex intensive businesses as each GW of compute capacity requires $30-50B in capital spending with the majority allocated on GPU servers (>50%), networking, storage, and CPU servers (management servers). This assumes that GPU cloud providers are leasing data center powered shell capacity from third parties though if they elect to self-build their data centers, this could add an additional $8-15B per GW of capex.

Financing: Financing for GPU cloud capex typically comes from customer prepayments, debit/credit, equity issuances, FCF/cash on hand or any combination of these sources. For GPU clouds with more limited operating history, the preferred form of financing is typically customer prepayments and project-tied debt financing that is backed by multi-year take or pay customer contracts. Hyperscalers that have strong operating histories finance their GPUs with cash/FCF but more recently they accessed the credit and equity markets as well.

Competitive Landscape: GPU clouds operate in a rapidly evolving and highly competitive AI infrastructure industry, driven by rising demand for specialized, high-performance computing and diverse customer solutions. As generative AI adoption continues to expand, it has become a major driver of data center investment across the broader cloud ecosystem. Much of today's AI investment has been concentrated among hyperscalers building their own captive capacity and for frontier model builder customers as well. Hyperscalers providing GPU clouds include Amazon (covered by our colleague Mark Mahaney), Azure (MSFT, covered by our colleague Kirk Materne), GCP (GOOGL, covered by our colleague Mark Mahaney), OCI (ORCL, covered by our colleague Kirk Materne), and per Bloomberg reports, META is also considering launching a GPU cloud business. In addition, a new class of "neocloud" providers is emerging – these include CRWV, Crusoe, NBIS, Nscale, among others. Within this segment, offerings range from highly targeted, niche solutions to broad-based platforms designed to address a wide spectrum of AI applications. Competitive differentiation hinges on factors such as technical architecture, hardware capacity, software orchestration, and security.

Potential Upside Drivers: While GPU rental revenue will drive the majority of top line performance for GPU cloud providers, we see potential for upside in business models from: **1) Extending GPU Utilization Beyond Accounting Useful Life:** GPUs that are leased beyond ~6-year accounting useful lives should contribute to profitability at a much higher margin as EBIT margin would approach EBITDA margin once the GPUs are fully depreciated in this scenario, **2) Higher GPU Hour** pricing resulting from mix shift toward on demand – this mix shift could be driven by GPUs rolling off current contracts or from improving cost of capital to allow for more speculative (spec) builds, and **3) Attach of Ancillary Software/services** with higher margins such as software, CPU, networking, and storage.

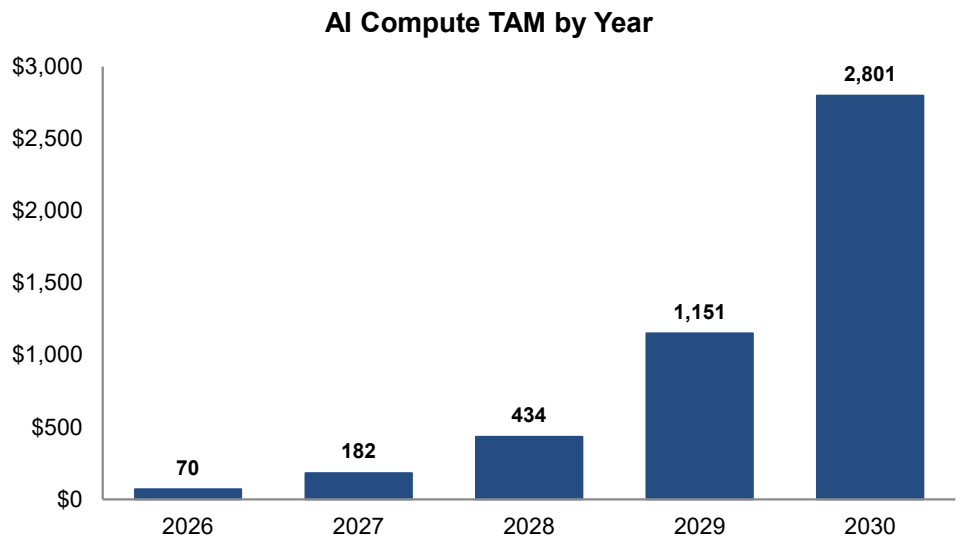
Valuation: There have been multiple frameworks discussed and we believe forward EV/EBITDA and EV/EBIT to be the most appropriate valuation methodologies. For Neocloud providers, we think enterprise value valuations based on forward profitability metrics should be burdened with the future debt/capex necessary to generate that profit. Based on that methodology, the neocloud comp set is trading at a high single digit EBITDA multiple. Investors/creditors also evaluate the "EBITDA power" of providers (similar to book but not billed EBITDA for data centers), particularly those with limited operating history.

Executive Summary

The GPU cloud (or "neocloud") industry has emerged as the connective tissue of the AI build-out or the layer that converts the combination of accelerated silicon (GPUs and/or XPUs), networking, storage and data center powered shell into AI compute infrastructure that frontier LLM labs, hyperscalers and enterprises rely on for AI model training and inference workloads (capacity is typically rented by the per GPU-hour basis). This report provides an overview of the TAM opportunity for GPU clouds, their business model, their role in the broader AI infrastructure stack, supply/demand dynamics, unit economics, financing, and the value add they enable.

The Start of A Demand-led Era: We see the GPU cloud market in the early innings of a multi-year capacity build. Based on our annual token consumption model that estimates annual token demand growing from ~100 quadrillion in 2026 to ~4.0 quintillion in 2030 and assuming an average install base throughput of ~500,000 TPS/MW, we see demand for data center capacity reaching ~250 GW by 2030. If we further assume a cost of ~$11M per MW of compute capacity, this translates to a GPU cloud TAM that will grow from ~$70B in 2026 to ~$2.8T by 2030.

Figure 1: Annual GPU Cloud TAM ($B)

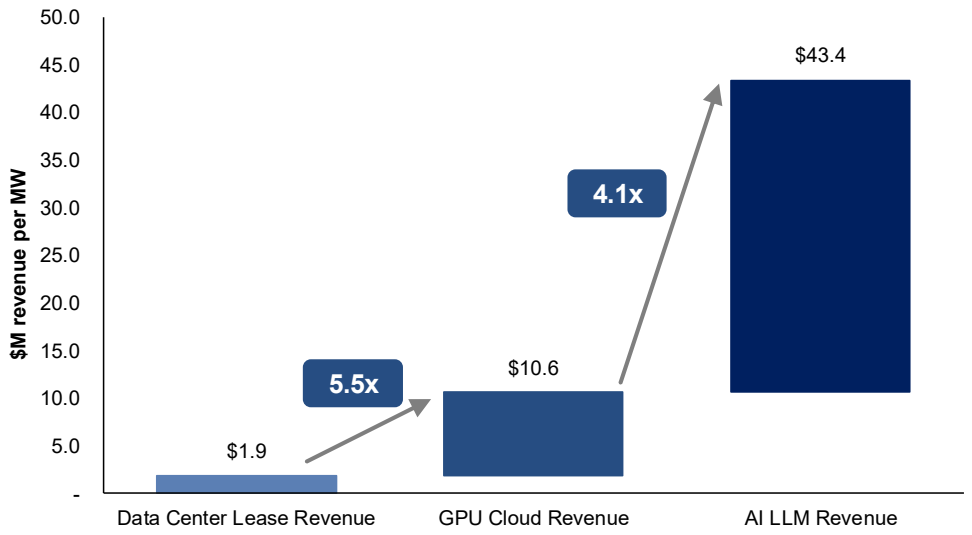


Source: Company Data, Evercore ISI Research

Compelling Value Add Multiplication from DC Capacity ➔ GPU Cloud ➔ AI LLM

Assuming LLM customers pay an average of $10-12M per MW of GPU cloud capacity annually, and they achieve a throughput of ~500,000 TPS/MW, the cost per million tokens (Mtok) could be $0.60-0.70. If we then assume they then monetize each Mtok at close to ~$3 on average, AI LLM companies are monetizing their LLMs at a ~4x multiple of what they are paying GPU cloud providers. Putting this in to per MW terms, GPU cloud providers procure data center powered shell capacity at ~$2M/MW annually (mid to high $100/kW per month) and add value by combining technical infrastructure with physical infrastructure to create clouds, and then monetize those megawatts at a multiple (>5x or $10-12M/MW assuming mid-$2 type per GPU hour pricing) of what they are paying data center operators. Frontier AI LLM companies then monetize that compute capacity at a multiple (>4x or $40M per MW assuming ~$2.75 blended price per Mtok) of what they pay GPU cloud providers.

Figure 2: Value Add Multiplier from DC Capacity to GPU Cloud to AI LLM ($M per MW)



Source: Company Data, Evercore ISI Research

The illustrative example above (base case) where AI LLM firms can realize a ~4x multiplier on cloud spending is based on a GPU cloud cost per MW of $10-11M, a throughput of ~500k TPS/MW, and blended price per Mtok of $2.75. However, it's worth noting that even at premium pricing for GPU cloud capacity ($12 per GPU hour or ~$49M per MW), AI frontier model lab firms can still generate revenue at multiples (~9x) of their cloud spend for inference workloads if they are able to achieve high throughput (2.8M TPS per MW) and raise blended price per Mtok to ~$5.00 as shown in the High Cost, High Throughput, High LLM Price case shown below. Notably, Anthropic is charging $10 per Mtok for input tokens and $50 per Mtok for output tokens on their latest Claude Fable 5 flagship tier model – assuming a 3:1 input/output ratio, the blended price per Mtok could be closer to ~$20. OpenAI's gpt 5.5 pro model charges $60 per Mtok for input tokens (long context) and $270 per Mtok for output which means blended price could be >$100 per Mtok.

Figure 3: Value Add Multiplication Through The AI Infrastructure Stack

AI Infrastructure Value Stack

How Each Megawatt Is Monetized Across the AI Stack: Data Center → GPU Cloud → AI LLM

($ per year per MW of data center IT load, unless noted otherwise)

Value Driver (per MW)	Base Case	High Cost, High Throughput, High LLM Price
Data Center Layer		
Powered-shell pricing ($/kW/month)	$160	$200
Annual DC lease revenue per MW	**$1.9M**	**$2.4M**
GPU Cloud Layer		
GPU-server power share (of DC IT load)	84.3%	84.3%
GPU-server power per MW	843 kW	843 kW
kW per NVL72 rack	130 kW	130 kW
NVL72 racks per MW	6.5	6.5
GPUs per NVL72 rack	72	72
Total GPUs per MW	467	467
Price per GPU-hour	$2.60	$12.00
GPU cloud revenue per MW	**$10.6M**	**$49.1M**
GPU cloud value-add multiplier vs. DC capacity	**5.5x**	**20.4x**
AI LLM Layer		
Throughput (tokens/sec per MW)	500,000	2,800,000
Seconds per year	31,536,000	31,536,000
Tokens per MW per year (Mtok)	15,768,000	88,300,800
Blended price per Mtok	$2.75	$5.00
AI LLM revenue per MW	**$43.4M**	**$441.5M**
AI LLM value-add multiplier vs. GPU cloud	**4.1x**	**9.0x**
AI LLM value-add multiplier vs. DC capacity	**22.6x**	**184.0x**

Source: Company Data, Evercore ISI Research

Risks to the Model

While the demand backdrop is well understood, we note that GPU cloud providers do face a higher level of execution risk given multiple moving pieces to the model, high capex requirements (and related to that, financing requirements as well), reliance on third parties for data center powered shell capacity (supply constrained) and technical infrastructure (GPU servers, CPUs, storage, networking, memory, etc.). Some of the key risks are summarized below:

- **Demand air-pocket:** a pause in frontier training, token spend optimization efforts, or slower-than-expected enterprise/agentic adoption would leave capacity stranded against fixed cost.
- **GPU depreciation & obsolescence:** six-year accounting lives may overstate economic life if newer silicon compresses the residual value of installed fleets.
- **Capital intensity & financing:** each GW requires $40-60B; a higher-rate or risk-off environment raises the cost of the debt that funds the build.
- **Power & supply constraints:** grid interconnection and powered-shell delivery are now the binding constraint on growth in key markets.
- **Policy & concentration:** export controls, customer concentration and hyperscaler insourcing each threaten the revenue base.
- **Competition:** particularly from better-resourced hyperscalers looking to monetize and underutilized capacity.

Ten Key Takeaways

1. **AI compute demand is the engine.** Token consumption — our proxy for demand — grows from ~100 quadrillion (2026) to ~4.0 quintillion (2030), a ~150% CAGR, implying ~250 GW of data center capacity by 2030.
2. **Take-or-pay contracts create visibility.** For CRWV in particular, multi-year reserved capacity supports multi-year revenue visibility and allows for project-type financing rarely available to early-stage businesses.
3. **The model is intensely capital-intense.** Each GW of capacity requires ~$40-60B of capex (>50% GPUs); self-building shell adds $8-15B/GW.
4. **Unit economics work at mid-$2 pricing.** Even at a mid-$2 per GPU hour pricing (multi-year reserved capacity), deals are being underwritten at mid-20% contribution margin in years 1-5, rising toward >60% once hardware is fully depreciated assuming similar renewal pricing beyond year 6.
5. **Financing is a differentiator.** For CRWV, Investment-grade-rated, non-recourse DDTL structures have reduced cost of debt capital by ~750 bps from DDTL 1.0 to 4.0.
6. **Pricing durability a swing factor.** H100 spot pricing has trended up YTD in CY26 despite newer silicon, which we view as a positive demand signal and potential upside driver once take or pay contracts expire and are renewed (or converted to on-demand capacity).
7. **Software & services move operators up-stack.** Ancillary CPU, storage, networking and software carry higher margins and improve stickiness.
8. **Power is the binding constraint.** Powered-shell constraints in key markets is gating overall growth.
9. **Competition is intensifying.** The attractive return profile and revenue generation profile is attracting competition from larger, better-resourced providers.
10. **Valuations have room to shift upward.** GPU clouds trade at a discount to hyperscalers and DC REITs; we see re-rating potential as business-model confidence builds.

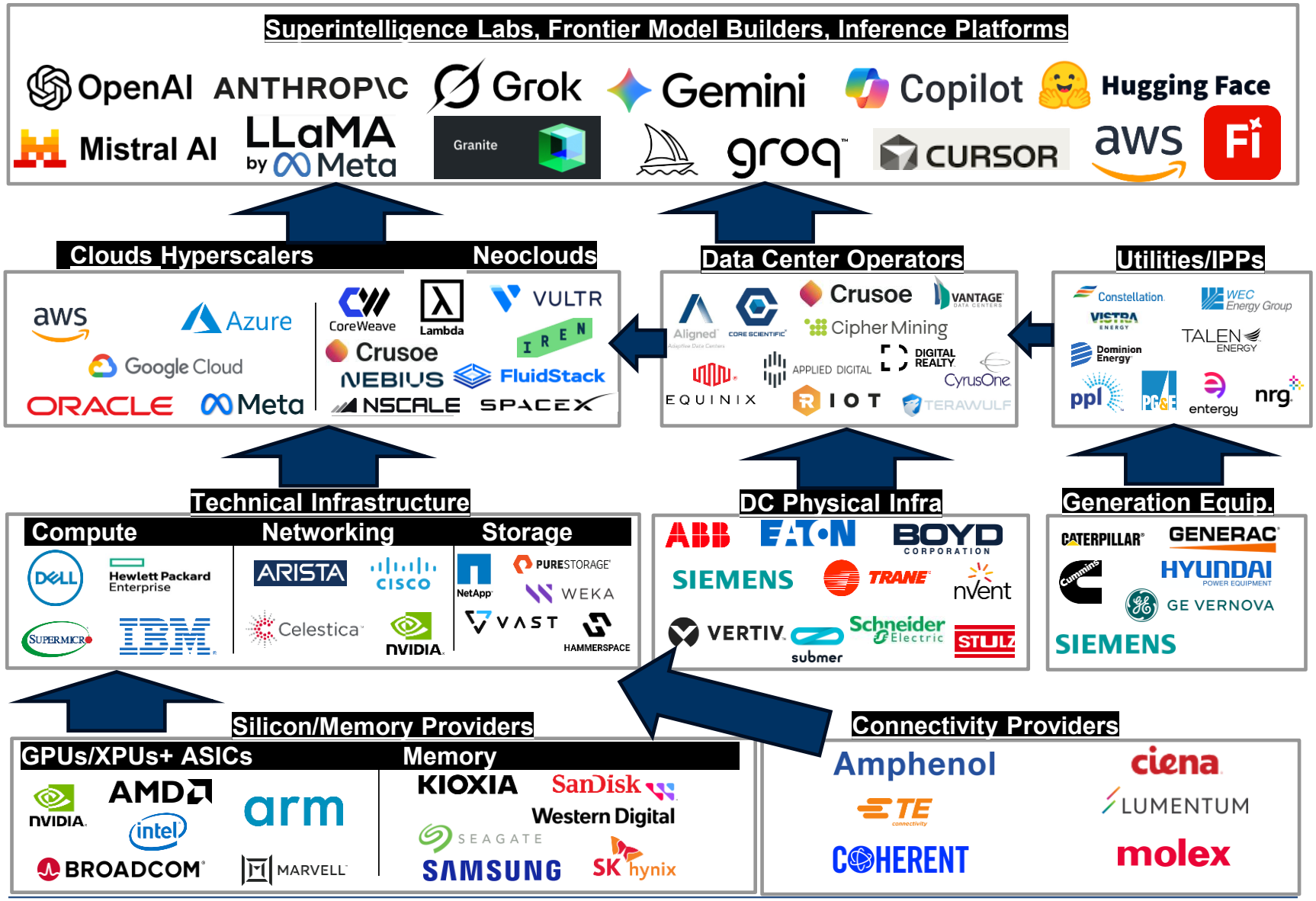
The GPU Cloud Business Model

A GPU cloud delivers GPU compute capacity to frontier AI model builders, hyperscalers, and enterprises, and other organizations looking to train AI models and run inference workloads without the hassle of owning/operating complex AI infrastructure that is becoming increasingly more difficult to manage. GPU clouds are much more complex vs. general purpose compute infrastructure given increased scale, power + power density (requires much larger utility agreements and more robust physical infrastructure with liquid cooling capabilities) and increased networking needs. GPU clouds are responsible for procuring GPUs/servers, other ancillary technical infrastructure, and data center powered shell capacity and converting those components into a cloud service for customers.

GPU cloud compute is typically sold on a per GPU hour basis via: **1) Reserved Capacity:** Multi-year reserved capacity with take-or-pay contract structure that is typically priced at a significant discount vs. spot/on-demand agreements. Deals are underwritten with a target contribution margin/return in mind. **2) On-demand:** Pay-as-you go structure that is priced on a per GPU hour basis typically at a significant premium vs. reserved capacity. Pricing is more volatile depending on supply/demand backdrop. Agreements could include an initial period of reserved capacity. **3) Spot:** For non-critical workloads, customers bid for unutilized capacity from neoclouds though capacity is not guaranteed and can be terminated by the provider at any time. We have seen pricing per GPU-hour range from $0.75 to $12+ depending on the silicon, time to availability, and the contract.

Where GPU Clouds Sit in AI Infrastructure Stack: GPU cloud providers sit between their suppliers (providers of technical infrastructure such as server/storage/networking OEMs and silicon providers) and their AI customers (frontier model labs, hyperscalers, enterprises, etc.) as seen on the following page on Figure 4.

Figure 4: The AI Infrastructure Stack



Source: Company Data, Evercore ISI Research

Unit Economics of a 5-Year Take-or-Pay Contract

Figure 5 on page 10 provides an illustrative example of the economics of a multi-year take or pay contract. In our example, a ~$6B TCV 5-year GB200 customer contract could require ~$2.8B of capex assuming a mid ~$2 type per GPU pricing. Initially, contribution margins will be muted/negative as GPU cloud providers such as CRWV incur data center lease/power/physical infra depreciation expenses once they accept data center capacity from a third-party provider without recognizing revenue from the customer associated with the contract as they install GPUs into the facility; this takes anywhere from 10-90 days (1 to 3 months). Post Month-3 of accepting data center capacity, they are able to generate revenue at fully stabilized levels and drive project contribution margin at a mid 20% level. Finally, we see further margin upside potential from monetization of GPUs beyond the 6-year accounting useful life (depreciation no longer an expense), software upsell, storage attach, and monetization of software stack on other cloud platforms. Beyond year-6 once the hardware is fully depreciated, we see potential for >60% type contribution margins assuming pricing remains at similar levels.

- **Total GPUs:** $2.8B of capex will provide ~732 Nvidia NVL72 Blackwell servers with a unit cost of ~$3.2M each (including management storage/networking/compute infrastructure). With ~732 NVL72 units, total GPUs in the deployment will be ~53k.
- **Revenue:** The revenue associated with this hypothetical contract will be calculated by multiplying total number of GPUs by per hour pricing. We are assuming mid-$2 per GPU hour in our analysis. At ~53k GPUs, (24 hours per day and 365 days per year), we see annual revenue associated with this deployment at ~$1.2B.
- **Depreciation Expense:** $2.8B in capex across a 6-year accounting useful life (assuming straight line depreciation), annual depreciation expense should be ~$400-500M in years 1-6. Beyond year 6, depreciation expense should approach 0 (this will be the largest driver of EBIT margin expansion beyond year-6).
- **Data Center Capacity Costs:** Assuming CRWV relies on third party data center operators for capacity in our hypothetical deployment (vs. using captively developed data center capacity), we think data center lease costs could be $216M annually, which is based on ~120 MW of IT load power demand and $150/kW per month pricing.
- **Power Costs:** We estimate ~$128M in annual power costs associated with a 732-unit NVL72 deployment based on ~120 MW of IT load power use, 1.15 PUE, and ~$0.10/kWh type pricing for power.
- **Installation/Management/Admin Costs:** We assume CRWV needs one full time resource for every ~100 servers. At a cost of $250,000 per resource, we think annual costs related to management/admin of the servers in our analysis could be ~$2M.

Figure 5: CRWV Take or Pay Contract Economics

	M1	M2	M3	M4	M5	M6	M7-M12	Year 1	Years 2-5	Years 7+
Revenue Build										
NVL72 Units in Deployment	732	732	732	732	732	732	732	732	732	732
GPUs per NVL72 Units	72	72	72	72	72	72	72	72	72	72
Total GPUs in Deployment	52,685	52,685	52,685	52,685	52,685	52,685	52,685	52,685	52,685	52,685
Pricing per GPU	$2.6	$2.6	$2.6	$2.6	$2.6	$2.6	$2.6	$2.6	$2.6	$2.6
Utilization	0%	50%	100%	100%	100%	100%	100%	100%	100%	100%
Cash Revenue ($M)	$0.0	$50.0	$100.0	$100.0	$100.0	$100.0	$600.0	$1,050.0	$1,200.0	$1,200.0
Expense Build										
Data Center Capacity Costs										
Data Center IT Load Capacity (MW)	120.0	120.0	120.0	120.0	120.0	120.0	120.0	120.0	120.0	120.0
DC Capacity Lease Rate ($/kW per month)	$150.0	$150.0	$150.0	$150.0	$150.0	$150.0	$150.0	$150.0	$150.0	$150.0
DC Capacity Costs	$18.0	$18.0	$18.0	$18.0	$18.0	$18.0	$108.0	$216.0	$216.0	$216.0
									18%	18%
Power Costs										
Total IT Load Power Required (kW)	120.0	120	120	120	120	120	120	120	120	120
PUE	1.15	1.15	1.15	1.15	1.15	1.15	1.15	1.15	1.15	1.15
Power Price (kWh)	$0.10	$0.10	$0.10	$0.10	$0.10	$0.10	$0.10	$0.10	$0.10	$0.10
Power Costs ($M)	$10.1	$10.1	$10.1	$10.1	$10.1	$10.1	$60.3	$120.6	$120.7	$120.7
Personnel										
Annual Cost per Systems Admin Resource	$250,000	$250,000	$250,000	$250,000	$250,000	$250,000	$250,000	$250,000	$250,000	$250,000
Servers per Systems Admin	100	100	100	100	100	100	100	100	100	100
Admin Expenses ($M)	$0.2	$0.2	$0.2	$0.2	$0.2	$0.2	$0.9	$1.8	$1.8	$1.8
Total Cost of Revenue ($M)	$29.2	$29.2	$29.2	$29.2	$29.2	$29.2	$175.2	$350.4	$350.5	$338.5
T&I										
Tech Infra Gross PP&E	$2,809.9	$2,809.9	$2,809.9	$2,809.9	$2,809.9	$2,809.9	$2,809.9	$2,809.9	$2,809.9	$2,809.9
Tech Infra D&A	$9.8	$39.0	$39.0	$39.0	$39.0	$39.0	$234.2	$234.2	$468.3	$0.0
Other T&I as % of Stabilized Revenue	9%	9%	9%	9%	9%	9%	9%	9%	9%	9%
Total T&I	$12.0	$48.0	$48.0	$48.0	$48.0	$48.0	$288.2	$540.3	$576.3	$108.0
Total Expenses ($M)	$41.2	$77.2	$77.2	$77.2	$77.2	$77.2	$463.4	$890.7	$914.8	$446.5
Total Operating Profit (Expense)	-$41.2	-$27.2	$22.8	$22.8	$22.8	$22.8	$136.6	$159.2	$285.1	$753.4
Contribution Margin	NA	-54%	23%	23%	23%	23%	23%	15%	24%	63%

Source: Company Data, Evercore ISI Research

Robust Demand

We view token consumption as a representation of demand for GPU cloud compute. In May, we introduced our Token Usage Model to better frame the underlying AI infrastructure demand. In our base case, we see total annual token demand growing from ~100 quadrillion in 2026 to ~4.0 quintillion in 2030, representing a ~150% 4-year CAGR. Translating this to compute demand requires an assumption for throughput (Tokens per watt per MW or TPS per MW which also depends on Interactivity or TPS/user) and assuming an average install base throughput of ~500,000 TPS/MW, we see demand for data center capacity reaching ~250 GW by 2030.

Tokens are fundamental units of data processed in AI model training and inference workloads. They are broken down from larger chunks of information and as AI models process tokens, they learn relationships (between tokens) which leads to prediction, generation, and reasoning capabilities. Tokens are used in both LLM training and inferencing workloads. At the unit level, a token can be thought of as four characters of English text (each token represents ¾ of a word assuming average word is five letters) – a prompt to a LLM platform that generates a ~300 word response might use ~400 tokens. Beyond text, tokens are also used for image (100-1,000 tokens per still image) and video creation (>250 tokens per second of video). As AI frontier model companies continue to train new models, consumers adopt LLM usage, and enterprises use AI for both inference and agentic AI workflows, token usage should grow exponentially over the next few years creating demand for AI data center infrastructure.

Token Consumption Model (BASE CASE)

In our base case, we see total annual token demand growing from ~100 quadrillion in 2026 to ~4.0 quintillion in 2030, representing a ~150% 4-year CAGR. Key assumptions here include ~1.5B AI consumer users growing to ~3.7B by 2030 and ~50M AI agents growing to ~800M by 2030.

Components/assumptions of our token consumption model (BASE CASE):

- **Consumer Inference:** Consumer inference represents AI LLM usage among consumer users for applications such as text-based AI queries and image/video creation (among other use cases). Key inputs/assumptions here include the number of global AI users, queries per user per day, average output tokens per query, and an input/output token ratio (3x the volume of information sent to generate inputted per response). For 2026, we are assuming 1.5B global AI users, 12 queries per user per day, 1,000 output tokens used per query, and a 3:1 input/output token ratio. This will drive ~26 quadrillion token usage per year. Looking ahead into 2030, we see global AI users rising to ~3.7B, an increase to 25 queries per user per day, 5,000 tokens used per output, a 3:1 input output token ratio leading to token consumption increasing to ~674 quadrillion.
- **Enterprise Inference (Non-agentic):** As enterprises adopt AI, we see enterprise token usage increasing exponentially. Assuming ~214M AI-enabled global white collar workers, 12 output queries per day, 2,000 tokens per query, and an input/output token ratio of 3:1, we think total enterprise inference token use could amount to ~7.5 quadrillion in 2026. As global enterprise AI adoption grows with ~1.3B AI-enabled white collar workers by 2030, 25 output queries per day, the same 3:1 input/output ratio, and average of 10k tokens per query, we think total annual token usage for enterprise inference could reach >460 quadrillion.
- **Agentic AI (Enterprise):** For 2026, our token usage estimate from agentic AI is based on 50M AI agents deployed, each consuming 400,000 tokens per day. This represents ~7.3 quadrillion tokens used annually. By 2030, we see AI agents increasing to 800M, with each agent using 8.5M tokens per day. This represents ~2.5 quintillion tokens used annually by 2030.
- **Frontier Model LLM Training:** For frontier model LLM training, we are assuming 12 frontier model companies that on average release four models per year. Each model is trained on 600B parameters and each parameter requires 500 tokens to train. We are also assigning a 3x other training token multiplier to account for other token usage related to model training (pre-training, fine tuning, etc.). For 2026, we see ~59 quadrillion tokens used by LLM training. By 2030, we see token usage from model training increasing to 373 quadrillion.
- **Enterprise SLM Training:** We see enterprise AI adoption leading to deployments of firm specific SLMs that require fewer parameters (<10B). Assuming ~86,000 enterprises train 2 SLMs per year, each with 4.5B parameters, and 132 tokens/parameter we see enterprise SLM training using ~100 trillion tokens in 2026, increasing to 24 quadrillion by 2030.

Figure 6: BASE CASE Token Consumption Model (CY24-CY30E)

Token Consumption Model

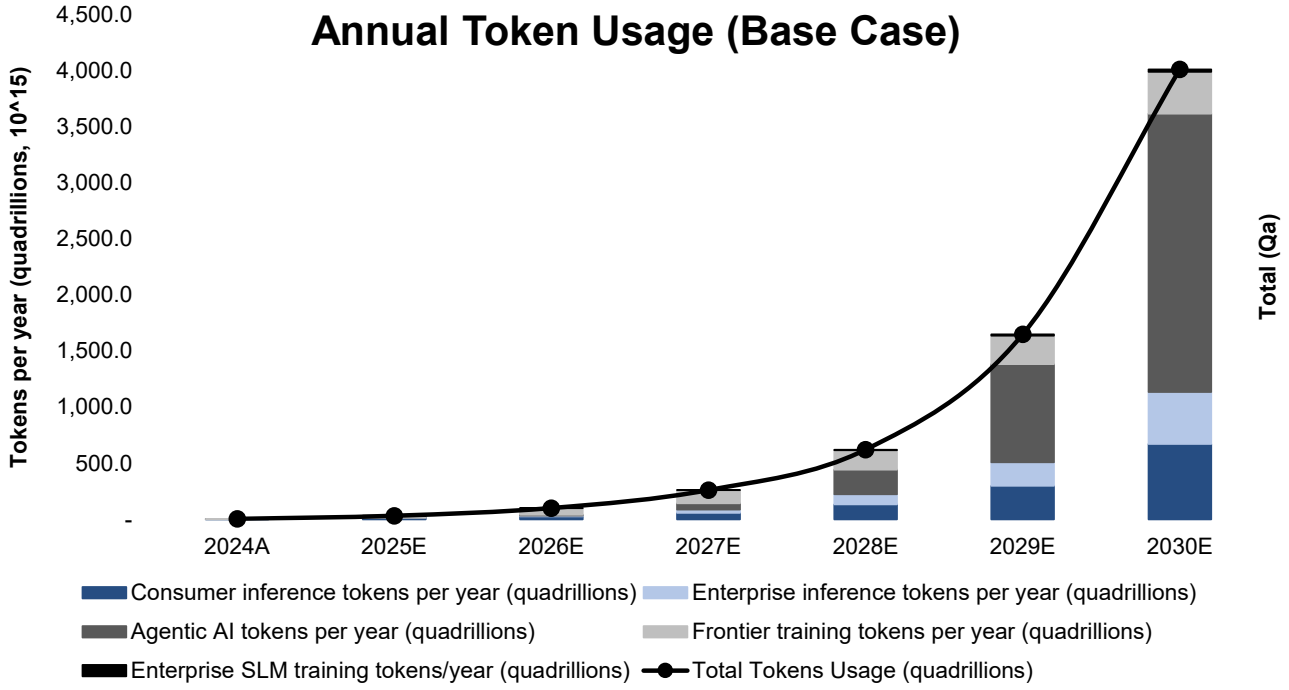
Bottoms-up Build

(units as noted per row)

	2026E	2027E	2028E	2029E	2030E
CONSUMER Inference					
Global AI users (M)	1,500	1,875	2,344	2,930	3,662
Queries / user / day	12	14	17	21	25
Output tokens / query	1,000	1,500	2,250	3,375	5,063
Input : Output token ratio	3.0x	3.0x	3.0x	3.0x	3.0x
Total output tokens per day (trillions)	18	41	91	205	461
Total input tokens per day (trillions)	54	122	273	615	1,384
Total tokens per day (input + output, trillions)	72	162	365	820	1,845
Total tokens per year (input + output, quadrillions)	**26.3**	**59.1**	**133.0**	**299.3**	**673.5**
ENTERPRISE Inference (Non-agentic)					
Global AI-enabled white collar population (millions)	214	455	804	1,021	1,257
Queries / user / day	12	14	17	21	25
Output tokens / query	2,000	3,000	4,500	6,750	10,125
Input : Output token ratio	3.0x	3.0x	3.0x	3.0x	3.0x
Total output tokens per day (trillions)	5.1	19.7	62.5	142.9	316.7
Total input tokens per day (trillions)	15.4	59.0	187.6	428.8	950.2
Total tokens per day (input + output, trillions)	20.5	78.6	250.2	571.7	1,266.9
Total tokens per year (input + output, quadrillions)	**7.5**	**28.7**	**91.3**	**208.7**	**462.4**
AGENTIC AI (Enterprise)					
Agents deployed (millions)	50	100	200	400	800
Avg tokens / agent / day (thousands)	400	1,500	3,000	6,000	8,500
Agentic AI tokens per day (trillions)	**20.0**	**150.0**	**600.0**	**2,400.0**	**6,800.0**
Days in year	365	365	365	365	365
Total tokens per year (input + output, quadrillions)	**7.3**	**54.8**	**219.0**	**876.0**	**2,482.0**
FRONTIER-CLASS MODEL TRAINING Tokens					
Frontier-class model firms (count)	12	13	14	15	16
Frontier training runs per year per firm	4	4	4	4	4
Total frontier model training runs	48	52	56	60	64
Parameters per frontier model (B)	600	1,000	1,250	1,563	1,953
Average Tokens per Parameter	510	561	617	679	747
Other training token multiplier	3.0x	3.0x	3.0x	3.0x	3.0x
Frontier training tokens per year (quadrillions)	**15**	**29**	**43**	**64**	**93**
Other training tokens per year (quadrillions)	44	88	130	191	280
Total frontier training tokens per year (quadrillions)	**59**	**117**	**173**	**255**	**373**
ENTERPRISE SLM Training					
Enterprises Using AI SLMs (thousands)	86	364	684	868	1,069
Enterprise SLM training runs per firm	2	3	5	8	11
Avg parameters per SLM training run (B)	4.5	6.8	10.0	10.0	10.0
Avg training tokens per SLM parameter	132.0	145.2	159.7	175.7	193.3
Enterprise SLM training tokens/year (quadrillions)	**0**	**1**	**6**	**12**	**24**
GRAND TOTAL — Inference + Training Tokens (Annual)					
Consumer inference tokens per year (quadrillions)	26.3	59.1	133.0	299.3	673.5
Enterprise inference tokens per year (quadrillions)	7.5	28.7	91.3	208.7	462.4
Agentic AI tokens per year (quadrillions)	7.3	54.8	219.0	876.0	2,482.0
Total Inference token demand per year (quadrillions)	**41.1**	**142.6**	**443.4**	**1,384.0**	**3,618.0**
Frontier training tokens per year (quadrillions)	58.8	116.7	172.8	254.6	373.3
Enterprise SLM training tokens/year (quadrillions)	0.1	1.2	5.5	11.6	23.5
Total Training tokens per year (quadrillions)	58.9	117.9	178.3	266.1	396.9
Grand total token demand per year (quadrillions)	**99.9**	**260.5**	**621.7**	**1,650.2**	**4,014.8**
Y/Y Growth	*220.4%*	*160.6%*	*138.7%*	*165.4%*	*143.3%*
Training share of total tokens	*58.9%*	*45.3%*	*28.7%*	*16.1%*	*9.9%*

Source: Company Data, Evercore ISI Research

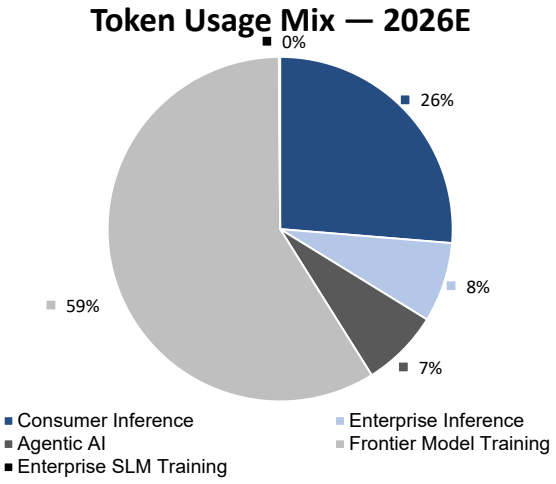
Figure 7: BASE CASE Token Consumption (CY24-CY30E)



Source: Company Data, Evercore ISI Research

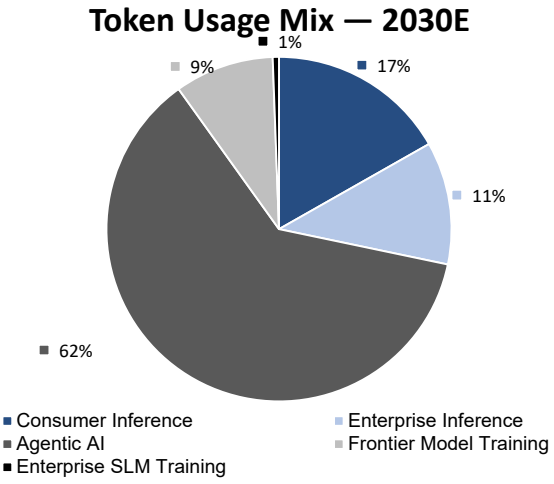
In 2026, we think frontier model training will represent the majority of token usage (~60%) followed by consumer user inference (26%), enterprise inference (8%), agentic AI (~7%), and enterprise SLM training (<1%). By 2030, we think Agentic AI will be the largest user of tokens, followed by consumer inference, and frontier model training.

Figure 8: Token Usage Mix (BASE CASE) – CY26E



Source: Company Data, Evercore ISI Research

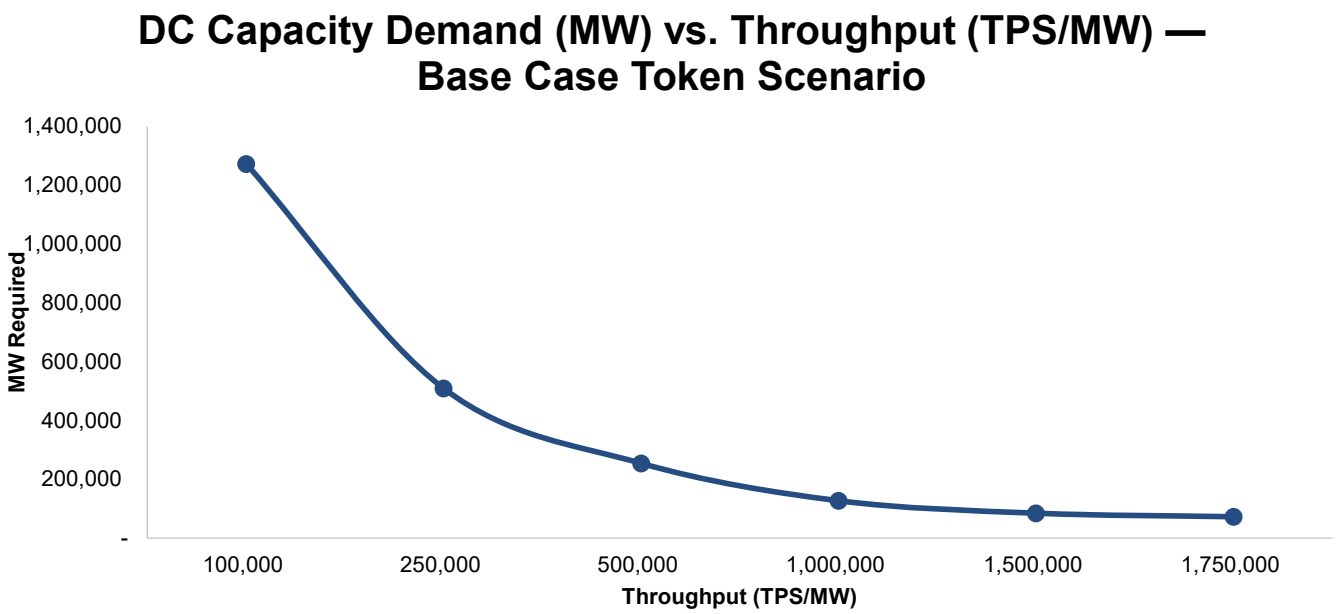
Figure 9: Token Usage Mix (BASE CASE) – CY30E



Source: Company Data, Evercore ISI Research

The logical next question is what future token consumption will represent for data center demand – this will vary based on average install base throughput (Tokens per watt per MW or TPS per MW which also depends on Interactivity or TPS/user). Based on our base case ~4.0 quintillion token usage by 2030 and assuming an average install base throughput of ~500,000 TPS/MW, we see demand for data center capacity reaching ~250 GW by 2030.

Figure 10: Base Case 2030 Data Center Capacity Demand (Assuming 4.0 Quintillion token consumption in 2030)



Source: Company Data, Evercore ISI Research

For our analysis, total AI token volume is the key driver of physical infrastructure demand. Our 2030 base case projects ~4.0 quintillion tokens generated annually predominantly from agentic inference workloads. Translating tokens into data center capacity demand requires assumptions made for blended average throughput (TPS/MW, how many tokens a megawatt of data center compute can produce per second). Throughput will depend on GPU/XPU productivity but also interactivity (TPS/User, inverse relationship between throughput and interactivity). Fixing our throughput assumption at ~500,000 TPS/MW for the blended average install base, we estimate serving the 2030 base case (~4.0 quintillion) requires ~250 GW of installed data center capacity.

Figure 11: Data Center Capacity Demand Based on Average Install Base Throughput vs. Token Usage Sensitivity

2030 Sensitivity Matrix — Data Center Capacity (MW) Required

TPS/MW ↓ | Token Usage (Quadrillions) →

	1,500	2,000	2,500	3,000	3,500	4,000	4,500	5,000	5,500	6,000
1,750,000	27,180	36,240	45,300	54,360	63,420	72,480	81,539	90,599	99,659	108,719
1,600,000	29,728	39,637	49,547	59,456	69,365	79,274	89,184	99,093	109,002	118,912
1,450,000	32,803	43,738	54,672	65,606	76,541	87,475	98,410	109,344	120,279	131,213
1,300,000	36,588	48,784	60,980	73,176	85,373	97,569	109,765	121,961	134,157	146,353
1,150,000	41,361	55,147	68,934	82,721	96,508	110,295	124,082	137,869	151,656	165,442
1,000,000	47,565	63,420	79,274	95,129	110,984	126,839	142,694	158,549	174,404	190,259
850,000	55,958	74,611	93,264	111,917	130,570	149,223	167,875	186,528	205,181	223,834
700,000	67,950	90,599	113,249	135,899	158,549	181,199	203,849	226,499	249,148	271,798
550,000	86,481	115,308	144,135	172,963	201,790	230,617	259,444	288,271	317,098	345,925
400,000	118,912	158,549	198,186	237,823	277,461	317,098	356,735	396,372	436,010	475,647
250,000	190,259	253,678	317,098	380,518	443,937	507,357	570,776	634,196	697,615	761,035
100,000	475,647	634,196	792,745	951,294	1,109,843	1,268,392	1,426,941	1,585,490	1,744,039	1,902,588

Source: Company Data, Evercore ISI Research

To better assess the demand backdrop, we think it is helpful understand tokenomics/compute capacity as an input cost for their LLM customers. Figure 2 on Page 4 provides an illustration on how frontier AI model labs are monetizing compute capacity at multiples of GPU clouds on a per

MW basis. Our analysis makes key assumptions surrounding throughput and compute cost per MW however we note that there is quite a bit of variability in both of these metrics. The table below provides wide range of per million token costs based on different throughputs and annual compute cost per MW.

Figure 12: Cost per Mtok for LLM Companies Based on Throughput vs. Annual Compute Cost per MW Sensitivity

Cost per Million Tokens ($) — Throughput × Annual Compute Cost

TPS/MW ↓ | Annual compute cost ($M / MW / yr) →

TPS/MW \ $/MW	8	10	12	14	16	18	20	22	24	26	28	30
3,000,000	$0.08	$0.11	$0.13	$0.15	$0.17	$0.19	$0.21	$0.23	$0.25	$0.27	$0.30	$0.32
2,750,000	$0.09	$0.12	$0.14	$0.16	$0.18	$0.21	$0.23	$0.25	$0.28	$0.30	$0.32	$0.35
2,500,000	$0.10	$0.13	$0.15	$0.18	$0.20	$0.23	$0.25	$0.28	$0.30	$0.33	$0.36	$0.38
2,250,000	$0.11	$0.14	$0.17	$0.20	$0.23	$0.25	$0.28	$0.31	$0.34	$0.37	$0.39	$0.42
2,000,000	$0.13	$0.16	$0.19	$0.22	$0.25	$0.29	$0.32	$0.35	$0.38	$0.41	$0.44	$0.48
1,750,000	$0.14	$0.18	$0.22	$0.25	$0.29	$0.33	$0.36	$0.40	$0.43	$0.47	$0.51	$0.54
1,500,000	$0.17	$0.21	$0.25	$0.30	$0.34	$0.38	$0.42	$0.47	$0.51	$0.55	$0.59	$0.63
1,250,000	$0.20	$0.25	$0.30	$0.36	$0.41	$0.46	$0.51	$0.56	$0.61	$0.66	$0.71	$0.76
1,000,000	$0.25	$0.32	$0.38	$0.44	$0.51	$0.57	$0.63	$0.70	$0.76	$0.82	$0.89	$0.95
750,000	$0.34	$0.42	$0.51	$0.59	$0.68	$0.76	$0.85	$0.93	$1.01	$1.10	$1.18	$1.27
500,000	$0.51	$0.63	$0.76	$0.89	$1.01	$1.14	$1.27	$1.40	$1.52	$1.65	$1.78	$1.90
250,000	$1.01	$1.27	$1.52	$1.78	$2.03	$2.28	$2.54	$2.79	$3.04	$3.30	$3.55	$3.81
100,000	$2.54	$3.17	$3.81	$4.44	$5.07	$5.71	$6.34	$6.98	$7.61	$8.24	$8.88	$9.51

Source: Company Data, Evercore ISI Research

While there is a wide range for cost per million tokens, we note that OpenAI and Anthropic's flagship models are monetizing tokens at several multiples of what they are paying per Mtok.

Figure 13: Select LLM Model Pricing

OpenAI

Flagship models

Standard | Batch | Flex | Priority

Our latest models

Prices per 1M tokens.

	Short context			Long context		
Model	Input	Cached input	Output	Input	Cached input	Output
gpt-5.5	$5.00	$0.50	$30.00	$10.00	$1.00	$45.00
gpt-5.5-pro	$30.00	-	$180.00	$60.00	-	$270.00
gpt-5.4	$2.50	$0.25	$15.00	$5.00	$0.50	$22.50
gpt-5.4-mini	$0.75	$0.075	$4.50	-	-	-
gpt-5.4-nano	$0.20	$0.02	$1.25	-	-	-
gpt-5.4-pro	$30.00	-	$180.00	$60.00	-	$270.00

Anthropic

Model pricing

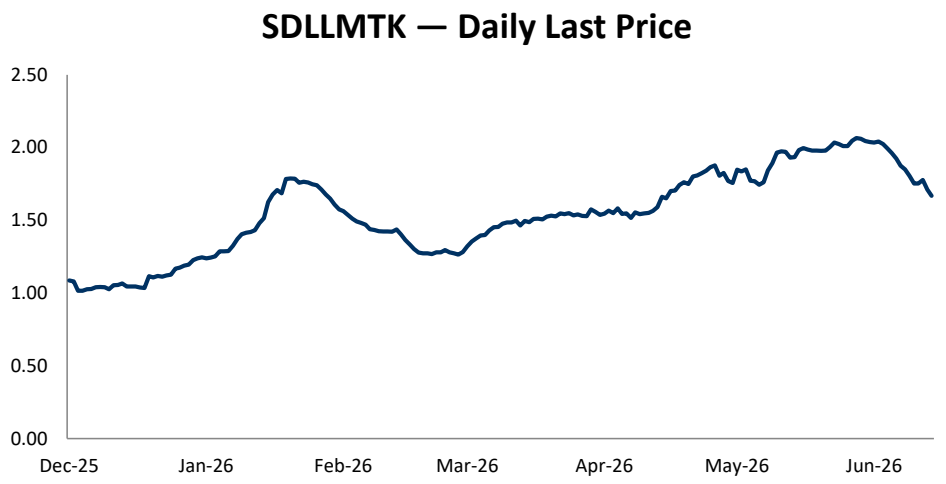
The following table shows pricing for all Claude models:

Model	Base Input Tokens	5m Cache Writes	1h Cache Writes	Cache Hits & Refreshes	Output Tokens
Claude Fable 5	$10 / MTok	$12.50 / MTok	$20 / MTok	$1 / MTok	$50 / MTok
Claude Mythos 5 (limited availability)	$10 / MTok	$12.50 / MTok	$20 / MTok	$1 / MTok	$50 / MTok
Claude Opus 4.8	$5 / MTok	$6.25 / MTok	$10 / MTok	$0.50 / MTok	$25 / MTok
Claude Opus 4.7	$5 / MTok	$6.25 / MTok	$10 / MTok	$0.50 / MTok	$25 / MTok
Claude Opus 4.6	$5 / MTok	$6.25 / MTok	$10 / MTok	$0.50 / MTok	$25 / MTok
Claude Opus 4.5	$5 / MTok	$6.25 / MTok	$10 / MTok	$0.50 / MTok	$25 / MTok
Claude Opus 4.1 (deprecated)	$15 / MTok	$18.75 / MTok	$30 / MTok	$1.50 / MTok	$75 / MTok
Claude Opus 4 (retired, except on Google Cloud)	$15 / MTok	$18.75 / MTok	$30 / MTok	$1.50 / MTok	$75 / MTok
Claude Sonnet 5 through August 31, 2026	$2 / MTok	$2.50 / MTok	$4 / MTok	$0.20 / MTok	$10 / MTok
Claude Sonnet 5 starting September 1, 2026	$3 / MTok	$3.75 / MTok	$6 / MTok	$0.30 / MTok	$15 / MTok
Claude Sonnet 4.6	$3 / MTok	$3.75 / MTok	$6 / MTok	$0.30 / MTok	$15 / MTok
Claude Sonnet 4.5	$3 / MTok	$3.75 / MTok	$6 / MTok	$0.30 / MTok	$15 / MTok
Claude Sonnet 4 (retired, except on Bedrock and Google Cloud)	$3 / MTok	$3.75 / MTok	$6 / MTok	$0.30 / MTok	$15 / MTok
Claude Haiku 4.5	$1 / MTok	$1.25 / MTok	$2 / MTok	$0.10 / MTok	$5 / MTok
Claude Haiku 3.5 (retired, except on Bedrock and Google Cloud)	$0.80 / MTok	$1 / MTok	$1.60 / MTok	$0.08 / MTok	$4 / MTok

Source: Company Websites

The SDLLMTK Token index provides another data point to track token pricing trends for how LLM companies are monetizing tokens. As long as there is a meaningful spread between what they pay for tokens and what they charge, we think demand for AI compute should remain strong.

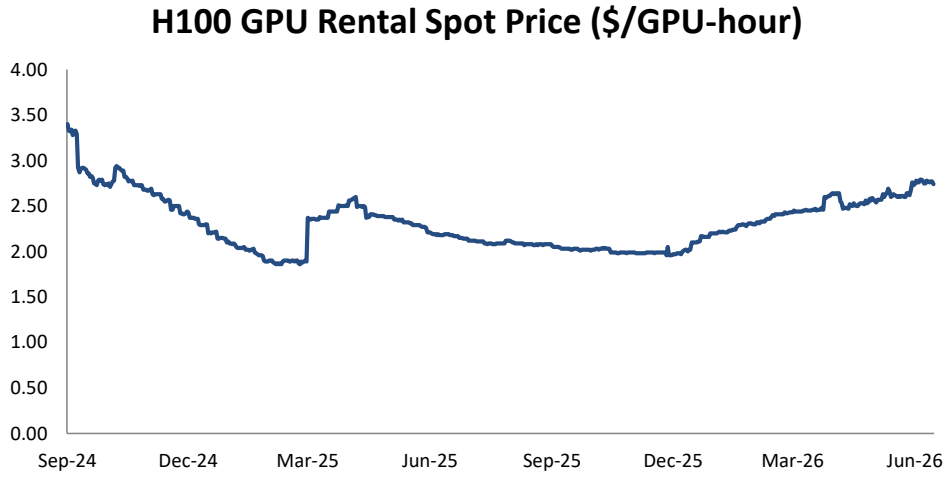
Figure 14: SDLLMTK Index (Price per Mtok)



Source: Bloomberg

We also view per GPU spot pricing for Nvidia H100 silicon as an indication of demand. Notably the platform was developed more than three years ago and year to date pricing has trended positively so far in CY26 despite newer GPU generations which we think is due to strong demand.

Figure 15: Spot H100 Pricing Index (per GPU Hour)

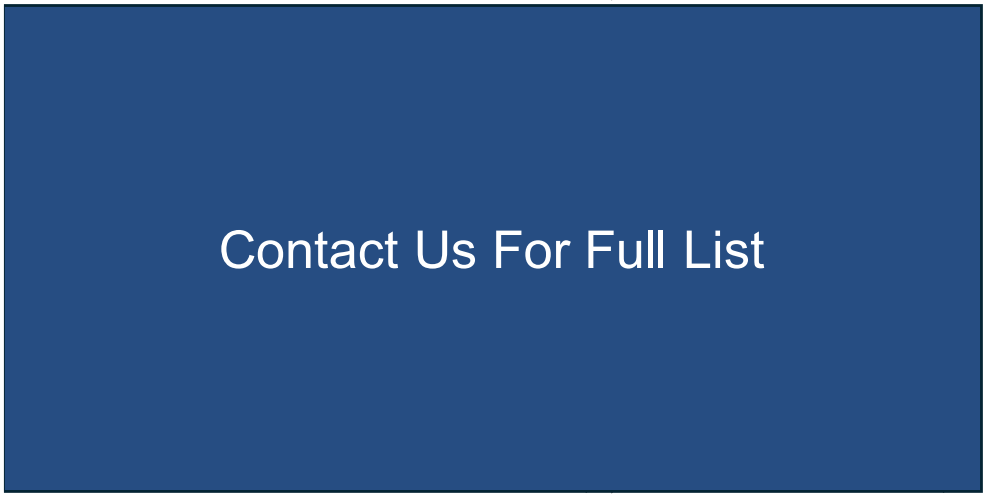


Source: Bloomberg

We have tracked cloud deals with AI frontier labs to get a better sense of end demand. The per GPU pricing range is somewhat wide and depends on the nature of the contracts.

Figure 16: Select GPU Cloud Deals Between Providers and AI Customers

Date Announced	Cloud Provider	Customer	TCV ($M)	Total Capacity (MW, IT Load)	Estimated per GPU Hour Pricing
6/5/2026	SpaceX	Google	NA	199	$11.46
6/1/2026	QumulusAI	Hyperbolic	$124.0	2	$3.69
5/7/2026	Iris Energy	Nvidia	$3,400.0	60	$2.92
5/6/2026	SpaceX	Anthropic	NA	587	$5.27
4/10/2026	CoreWeave	Anthropic	NA	NA	NA
4/9/2026	CoreWeave	Meta	$21,000.0	311	$2.60



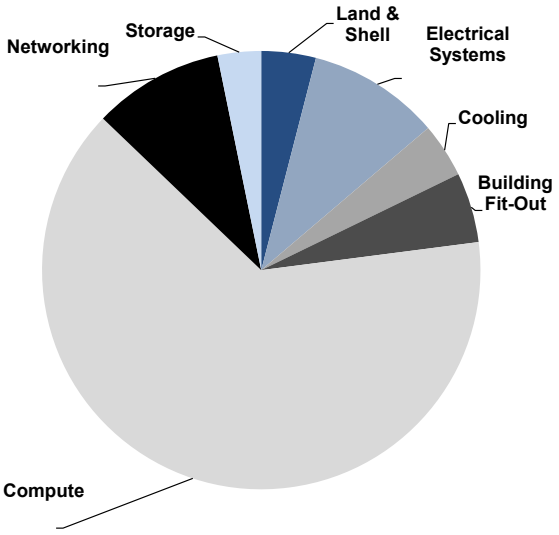
Source: Company Data, Evercore ISI Research

Capex/Supply

GPU clouds are capex intensive businesses as each GW of compute capacity requires $30-50B in capital spending with the majority allocated on GPU servers (>50%), networking, storage, and CPU servers (management servers). This assumes that GPU cloud providers are leasing data center powered shell capacity from third parties though if they elect to self-build their data centers, this could add an additional $8-15B per GW of capex.

- **1 GW of DC capacity equates to $40-60B in capex** across physical infrastructure (typically $10-15B per GW) and technical infrastructure ($25-45B).
- **Technical Infrastructure** includes compute, storage, and networking equipment such as servers (including xPUs), switches, routers, and optical interconnects that power data processing and communication Though we estimate ~2/3 of total spend has gone to compute over last few years, we expect spending to diversify toward storage/networking.
- **Physical Infrastructure** includes land/building shell, fit out, power/thermal management and facility systems (back up, generators, chillers, racks, cabling, that supply, protect, and physically support the IT hardware within a data center.

Figure 17: GPU Compute + Data Center Capacity Capex Cost Breakdown



Source: Company Data, Evercore ISI Research

Figure 18: Each GW of DC Capacity Represents $40-60B of Capex Potential

Source: Company Data, Evercore ISI Research

From a data center capacity perspective, we think a combination of data center operators and cryptominers that are converting their cryptoming power capacity to HPC/AI powered shell are a key source of DC capacity. We have tracked deals between clouds and DC capacity providers (helps tracks DC lease expense trends). Recently announced capacity deals between GPU/xPU providers and DC capacity providers suggest pricing for scale capacity for AI workloads is ~$140-180 per kW/month while powered shell type delas are trending toward the low $100 range.

Figure 19: Announced Deals Between DC Capacity Providers and Cloud Providers

Date Announced	Developer	Cloud Provider Customer	Location	TCV ($M)	Total Capacity (MW, IT Load)	Implied Monthly Rent ($/kW per Month, GAAP, IT Load)	Capex Required
6/8/2026	Applied Digital	Hyperscaler (IG, undisclosed)	Southern US	5,200.0	210	$138	NA
5/20/2026	Applied Digital	Hyperscaler (IG, undisclosed)	Northern US	7,500.0	300	$139	NA
4/23/2026	Applied Digital	Hyperscaler (IG, undisclosed)	Southern US	7,500.0	300	$139	NA
1/16/2026	RIOT	AMD	Rockdale, TX	311.0	25	$100	~$3.6M/MW
12/18/2025	WhiteFiber	Nscale	Madison, NC	865.0	40	$180	NA

Contact Us For Full List

Source: Company Data, Evercore ISI Research

Competitive Landscape

GPU clouds operate in a rapidly evolving and highly competitive AI infrastructure industry, driven by rising demand for specialized, high-performance computing and diverse customer solutions. As generative AI adoption continues to expand, it has become a major driver of data center investment across the broader cloud ecosystem. Much of today's AI investment has been concentrated among hyperscalers building their own captive capacity and for frontier model builder customers as well. Hyperscalers providing GPU clouds include AWS, Azure (MSFT), GCP (GOOGL), and OCI (ORCL). In addition, a new class of "neocloud" providers is emerging – these include CRWV, Crusoe,NBIS, Nscale, and Fluidstack among others. Within this segment, offerings range from highly targeted, niche solutions to broad-based platforms designed to address a wide spectrum of AI applications. Competitive differentiation hinges on factors such as technical architecture, hardware capacity, software orchestration, and security.

Traditional Cloud Competitors

GPU cloud competitors are widely recognized to include Amazon Web Services (AWS), Microsoft Azure, Google Cloud Platform (GCP), and Oracle Cloud. These providers represent the dominant hyperscale players in the market. Worth noting that these hyperscalers are also customers of neoclouds such as CRWV. A large portion of AWS and GCP AI infrastructure investments are tied to their proprietary silicon platform (Trainium/Inferentia for AWS, TPU for GOOG); companies such as GOOG still rely on CRWV to provide GPU compute infrastructure at scale.

Amazon Web Services offers a comprehensive, integrated cloud platform that enables customers to manage their infrastructure needs within a unified environment. It provides NVIDIA-accelerated instances, including the latest Blackwell GPUs, alongside SageMaker, a managed service for large-scale AI training and inference. AWS primarily targets enterprise customers with long-term cloud commitments and also supports AI-driven workloads in areas such as simulation and digital twins across multiple industries.

Microsoft Azure delivers general-purpose cloud computing as part of a broad, diversified product portfolio, contrasting with CoreWeave's specialized AI focus. Azure offers NVIDIA-accelerated instances, including the latest Blackwell GPUs, and provides enterprise-grade AI infrastructure through its strategic partnership with OpenAI. Similar to AWS, Azure primarily targets large enterprises with long-term cloud commitments, and appeals to organizations seeking cost-effective, compliant solutions for managing AI and broader cloud workloads.

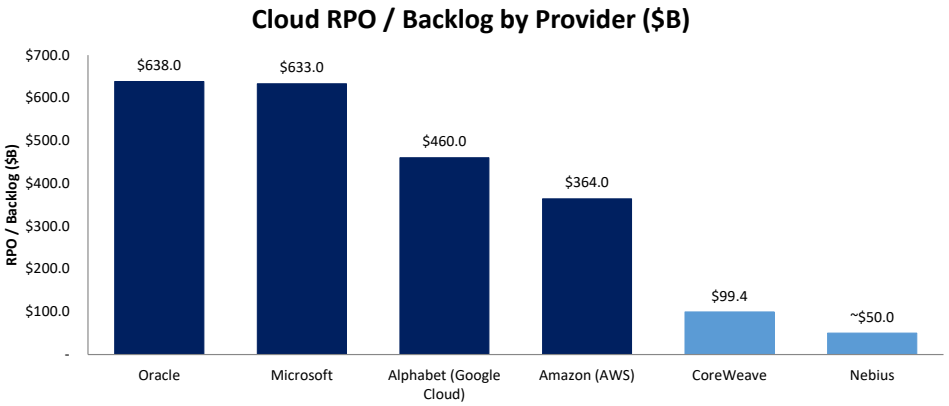
Google Cloud Platform delivers AI infrastructure as part of its broader innovation-led cloud ecosystem, setting itself apart from CoreWeave's AI-specialized approach. GCP supports large-scale model development through NVIDIA-accelerated instances, including the latest Blackwell

GPUs, and its proprietary Tensor Processing Units (TPUs). Its end-to-end machine learning platform, Vertex AI, offers integrated tools for building, training, and deploying models at scale.

Oracle Cloud Infrastructure (OCI) provides high-performance cloud services with a strong presence in government and defense contracts, alongside general-purpose computing. It offers advanced AI infrastructure, including AMD Instinct MI300X GPUs for multi-node training and large-scale LLM deployment, as well as NVIDIA-accelerated instances, including support for the latest Blackwell GPUs. OCI delivers both IaaS and PaaS capabilities tailored for AI workloads and targets customers seeking competitive pricing and the ability to fit large LLMs within a single node.

CoreWeave offers a compelling alternative to traditional hyperscalers by delivering AI-native infrastructure designed specifically for training and inference workloads. Its vertically integrated platform enables higher performance, faster deployment, and more efficient resource utilization than general-purpose cloud providers.

Figure 20: RPO/Backlog by Provider (Includes non-GPU services for hyperscalers)



Source: Company Data, Evercore ISI Research

Other Neocloud Competitors

While CRWV is widely recognized as the leading neocloud operator in the AI infrastructure market, it operates in a competitive environment with several emerging players.

- **Nebius** operates as a specialized AI infrastructure provider focused on delivering high-performance computing solutions tailored specifically for AI and machine learning workloads. Unlike traditional cloud providers, Nebius has built its platform from the ground up with proprietary hardware and software architecture specifically designed for intensive AI workloads, rather than adapting general-purpose cloud infrastructure. Nebius's primary product is its AI Cloud platform, which delivers a comprehensive suite of services designed specifically for AI development and deployment including full-stack AI infrastructure featuring large-scale GPU clusters. NBIS previously announced a $17.4B TCV deal (MSFT has the option to expand deal to $19.4B) that will be deployed over several tranches starting in 2025 through 2026. NBIS will provide MSFT with dedicated GPU infrastructure capacity at their Vineland, NJ data center. Earlier in March, NBIS also announced a ~$27B expansion with META.
- **Crusoe** is a vertically integrated AI infrastructure provider focused on delivering energy-optimized computing solutions. Its distinctive “energy-first” approach leverages stranded and underutilized energy resources to power AI workloads. Crusoe offers a suite of AI-optimized infrastructure products tailored for enterprise customers, anchored by Crusoe Cloud, the company’s flagship platform. Crusoe Cloud is a scalable, intuitive environment built on state-of-the-art GPU infrastructure, purpose-built to support next-generation AI workloads.

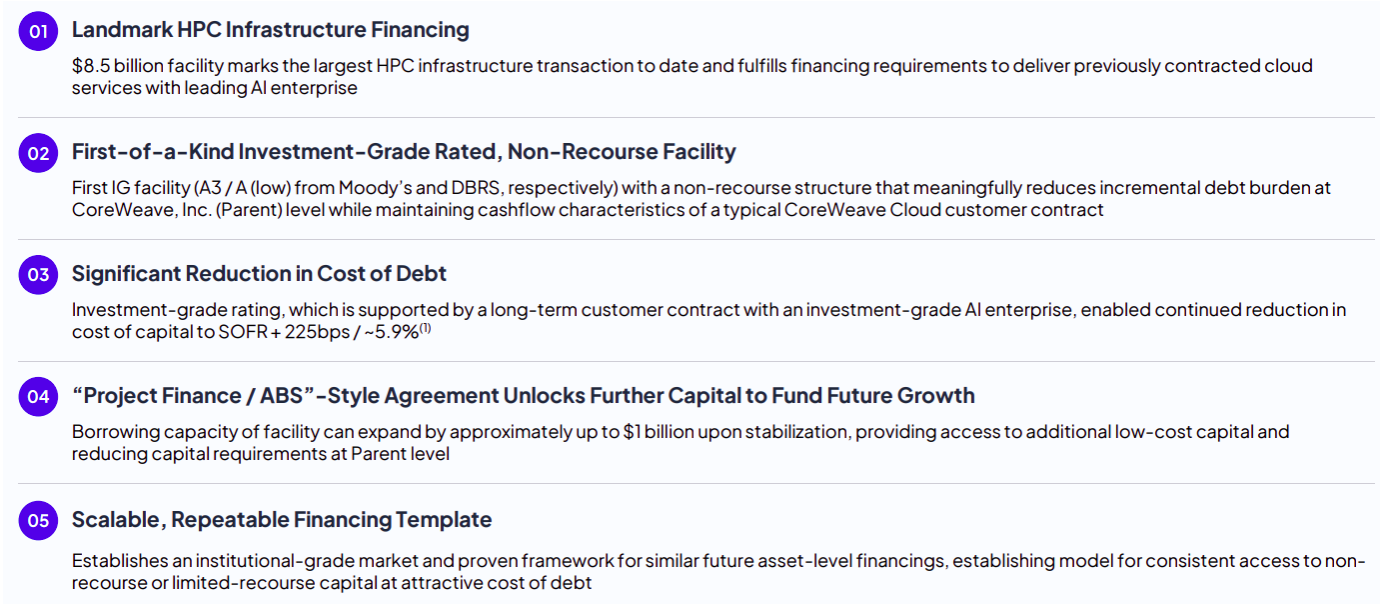
- **White Fiber** is a specialized AI infrastructure provider operating as a subsidiary of Bit Digital, Inc., with a focused presence in the GPU-as-a-Service market. The company's core offering is its GPUaaS platform, which delivers high-performance compute resources tailored for AI development and deployment. By providing flexible and scalable access to GPU infrastructure, White Fiber aims to support the foundational compute needs of teams building and deploying modern AI applications.
- **DigitalOcean (via Paperspace)** - DigitalOcean acquired Paperspace, an AI-focused cloud startup, to expand its offerings beyond traditional web hosting. Paperspace brings GPU instances featuring NVIDIA A100 and H100 GPUs, as well as managed AI tooling like notebooks, model training workflows, and inference endpoints. With DigitalOcean's developer-friendly approach, Paperspace is well-positioned to serve startups, SMBs, and individual developers looking for affordable, user-friendly GPU cloud services without enterprise-level complexity.
- **Vultr -** Vultr is a global independent cloud provider with data centers in over 30 regions. It recently launched NVIDIA GPU instances to target AI developers and enterprises. Vultr emphasizes cost efficiency and global reach, marketing itself as a cloud for AI "everywhere." The company has attracted attention with competitive pricing, simplicity, and rapid scaling, positioning itself as an alternative to hyperscalers while serving customers who need localized compute options.
- **OVHcloud** - OVHcloud is a European cloud giant offering GPU cloud services as part of its broader IaaS portfolio. Its AI cloud features NVIDIA GPUs, paired with networking and AI development tooling. OVHcloud appeals especially to European enterprises and governments due to its EU-based infrastructure, GDPR compliance, and sovereignty guarantees. It positions itself as a lower-cost, sovereign alternative to the US-based hyperscalers while still delivering advanced GPU-as-a-Service capabilities.
- **Vast.ai** - Vast.ai operates as a GPU marketplace rather than a centralized provider. It aggregates GPU capacity from data centers and individual operators worldwide, making GPUs and even consumer cards like the 4090 available at competitive rates. Vast.ai differentiates itself through flexibility and pricing transparency—users can shop for GPU power much like they would on Airbnb. This model is particularly popular among researchers, startups, and cost-sensitive developers who want direct access to raw GPU power without hyperscaler overhead.
- **Fluidstack -** Fluidstack is a London-based AI cloud provider that has rapidly evolved from a GPU aggregator into a gigawatt-scale data center developer, offering large-scale GPU (and Google TPU) capacity for both training and inference. Fluidstack sources power and shells largely by leasing capacity from bitcoin miners, including ~360MW contracted at TeraWulf's New York campus
- **IREN (formerly Iris Energy)** - is a bitcoin miner that has pivoted into AI/HPC cloud and data center development. The company operates six locations across roughly 4,900 acres with approximately 5 GW of secured, largely renewable power, and now runs an "AI Cloud" GPU business alongside its legacy mining and data center capacity, including an AI cloud partnership with Microsoft. IREN is a leading example of the miner-conversion cohort, one that operates its own neocloud rather than purely hosting third-party capacity.
- **TensorWave** - TensorWave is a U.S. (Las Vegas)-based neocloud built entirely on AMD Instinct accelerators, positioning itself as the AMD-native alternative to Nvidia-centric GPU clouds for both training and inference workloads. The company raised a $350M Series B (announced June 2026) to scale out its AMD-based cloud.

Financing

Financing for GPU cloud capex typically comes from customer prepayments, debt/credit, equity issuances, FCF/cash on hand or any combination of these sources. For neoclouds with more limited operating history, the preferred form of financing are typically customer prepayments and project-tied debt financing that is backed by multi-year take or pay customer contracts. Hyperscalers that have strong operating histories finance their GPUs with cash/FCF but more recently they accessed the credit and equity markets as well.

Earlier this year, CRWV achieved a key financing milestone closing on a delayed draw term loan (DDTL 4.0) facility which received an investment grade rating. Key hallmarks of the financing: **1) Investment Grade Rating:** DDTL 4.0 has received an A3 rating from Moody's and an A (low) rating by DBRS; the financing is secured by HPC infrastructure (GPUs) and the associated IG-rated customer's contract. **2) Lower Cost of Capital:** DDTL 4.0 will include a floating rate tranche financed at SOFR +2.25% and a fixed rate tranche at 5.9% - as a point of reference, this is a ~750bps improvement vs. DDTL 1.0 and a ~175bps improvement vs. DDTL 3.0. **3) Higher Loan to Cost:** CRWV's new facility expands the company's borrowing capacity to 102% loan-to-cost vs. 90% previously (helps fund construction phase, reduces need for parent level financing, customer pre-payments, etc.). **4) ABS Style Structure:** The new debt is non-recourse to the parent – CRWV parentco is not burdened from a ratings agency perspective by debt associated with this contract. **5) Expect Similar Types of Financing in the Future:** We think DDTLs with lower interest rates will be CRWV's preferred method of financing to fund infra needed to deliver on customer contracts looking ahead.

Figure 21: CRWV DDTL 4.0 Transaction Overview & Highlights



Source: Company Reports

Upside Drivers

While GPU rental revenue will drive the majority of top line performance for GPU cloud providers, we see potential for upside in business models from: **1) Extending GPU Utilization Beyond Accounting Useful Life:** GPUs that are leased beyond ~6-year accounting useful lives should contribute to profitability at a much higher margin as EBIT margin would approach EBITDA margin once the GPUs are fully depreciated in this scenario, **2) Higher GPU Hour** pricing resulting from mix shift toward on demand – this mix shift could be driven by GPUs rolling off current contracts or from improving cost of capital to allow for more speculative (spec) builds, and **3) Attach of Ancillary Software/services** with higher margins such as software, CPU, networking, and storage.

CRWV has a meaningful amount of GPU compute capacity that will be up for renewal next year and we have analyzed the potential upside from benefitting from higher GPU pricing (can be possible if CRWV shifts capacity to on-demand).

Upcoming Contract Renewals: A high single/low double digit percentage (of current revenue) of CRWV's contracts are set to expire next year – we think CRWV has several options including renewing/backfilling these contracts at similar rates assuming multi-year contracted capacity (given constrained supply, we expect minimal ASP degradation), leasing available capacity on demand (and enjoying premium spot rates), or a mix of both – renting a portion of available capacity at spot pricing could represent a source of upside for CY27. Assuming ~10% of CRWV's current annual revenue is expiring next year and that they are able to rent the GPUs at a spot rate that is ~100% of reserved capacity pricing, we think this could represent ~7% of revenue upside vs. current CY27 revenue estimates.

Figure 22: % Revenue Upside From Shifting Expiring Capacity To On-Demand in CY27

Revenue Upside from Shifting Renewing Capacity to On-Demand (% of CY27 Revenue)

% Curr Rev Avail. for Renewal ↓ | Spot Premium vs. Take or Pay →

% Renewing \ Spot Prem	25%	50%	75%	100%	125%	150%	175%	200%
6%	2.6%	3.1%	3.6%	4.1%	4.6%	5.1%	5.6%	6.1%
7%	3.0%	3.6%	4.2%	4.8%	5.4%	6.0%	6.6%	7.2%
8%	3.4%	4.1%	4.8%	5.5%	6.1%	6.8%	7.5%	8.2%
9%	3.8%	4.6%	5.4%	6.1%	6.9%	7.7%	8.5%	9.2%
10%	4.3%	5.1%	6.0%	6.8%	7.7%	8.5%	9.4%	10.2%
11%	4.7%	5.6%	6.6%	7.5%	8.5%	9.4%	10.3%	11.3%
12%	5.1%	6.1%	7.2%	8.2%	9.2%	10.2%	11.3%	12.3%

Source: Company Data, Evercore ISI Research

Furthermore, assuming the same ~10% of CRWV's current annual revenue is expiring next year and that they are able to rent the GPUs at a spot rate that is ~100% of reserved capacity pricing, we think this could represent ~33% of EBIT upside vs. current CY27 EBIT estimates (assumes a ~75% contribution margin).

Figure 23: % EBIT Upside From Shifting Expiring Capacity To On-Demand in CY27

EBIT Upside from Shifting Renewing Capacity to On-Demand (75% Contribution Margin, % of CY27 EBIT)

% Curr Rev Avail. for Renewal ↓ | Spot Premium vs. Take or Pay →

% Renewing \ Spot Prem	25%	50%	75%	100%	125%	150%	175%	200%
6%	12.5%	15.0%	17.6%	20.1%	22.6%	25.1%	27.6%	30.1%
7%	14.6%	17.6%	20.5%	23.4%	26.3%	29.3%	32.2%	35.1%
8%	16.7%	20.1%	23.4%	26.8%	30.1%	33.4%	36.8%	40.1%
9%	18.8%	22.6%	26.3%	30.1%	33.9%	37.6%	41.4%	45.1%
10%	20.9%	25.1%	29.3%	33.4%	37.6%	41.8%	46.0%	50.2%
11%	23.0%	27.6%	32.2%	36.8%	41.4%	46.0%	50.6%	55.2%
12%	25.1%	30.1%	35.1%	40.1%	45.1%	50.2%	55.2%	60.2%

Source: Company Data, Evercore ISI Research

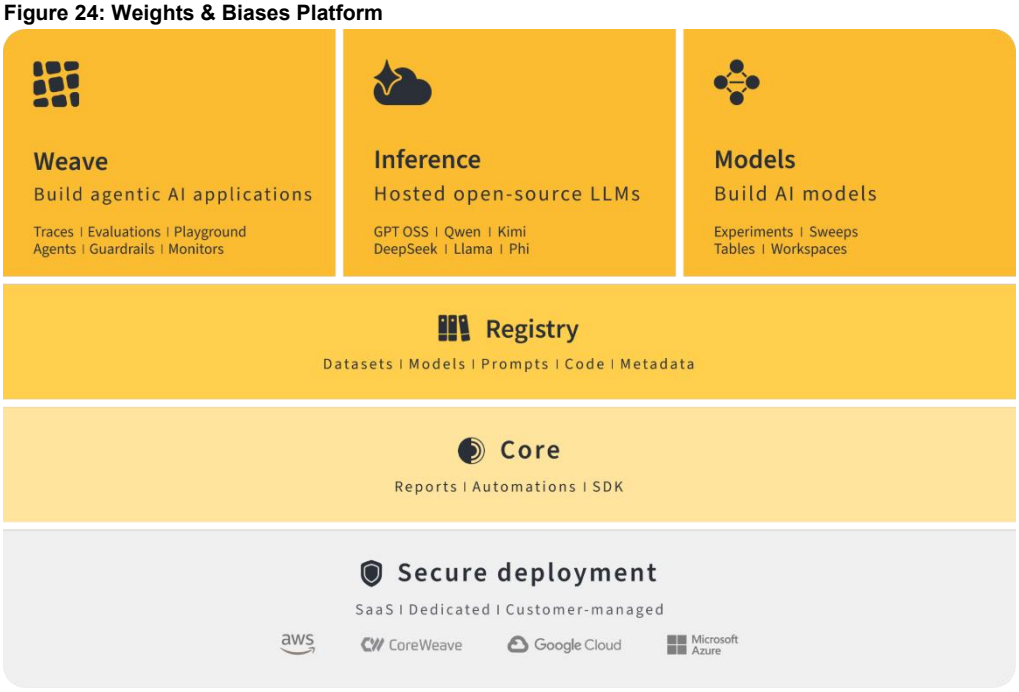
Beyond GPUs, CRWV indicated that they can deliver CPUs, storage, networking, software solutions, and developer tools – each of these ancillary products are expected to reach >$100M of ARR by year end CY26 (implies >$500M in ARR from higher margin solutions). A mix shift toward these solutions should drive revenue and EBIT upside (ancillary solutions are margin accretive). Some of CRWV's ancillary software offerings are highlighted below.

Weights & Biases Acquisition – Moving Up The Value Stack.

CRWV completed its ~$1.7B acquisition of Weights & Biases on May 5, 2025, strengthening its position as an AI Hyperscaler by integrating a leading AI developer platform into its cloud infrastructure. Weights & Biases is a developer platform for building, tracking, and monitoring machine learning and generative AI models. It is widely used by AI teams to reduce overhead,

improve traceability, and deploy models faster—integrating seamlessly with popular ML frameworks and tools. The acquisition of Weights & Biases builds on the effort to diversify CoreWeave's customer base. Weights & Biases serves over 1,400 organizations, across AI labs and enterprises, as well as over 1M AI researchers. Though smaller in scale, CoreWeave has an opportunity to cross sell across their respective customer bases. This move aligns with CoreWeave's strategy to "verticalize the ownership of data center infrastructure," enabling deeper control over power and compute resources while expanding its enterprise customer reach. The acquisition also provides CRWV with new customer relationships for CRWV to sell its core GPU cloud service offering to.

Figure 24: Weights & Biases Platform



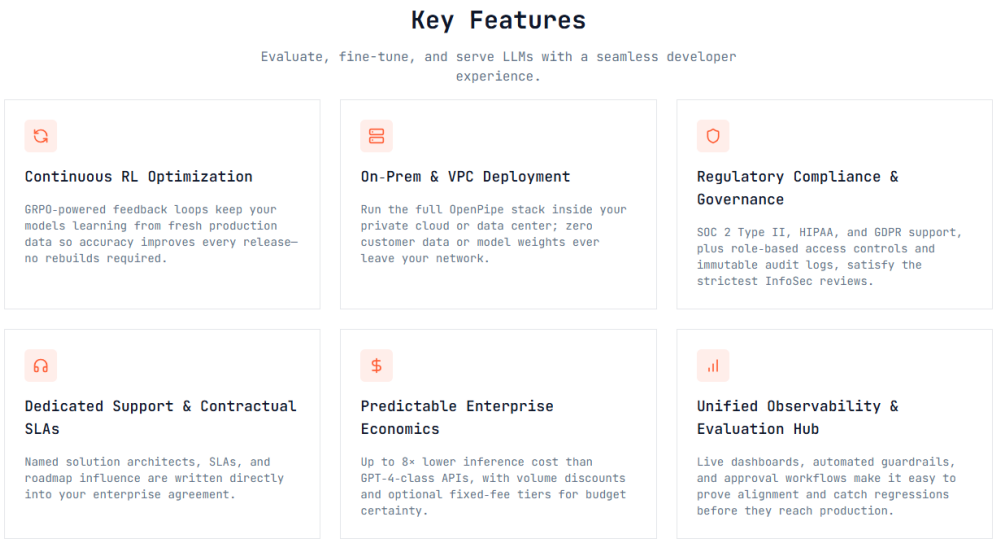
Source: Company Data

In particular, the integration of Weights & Biases into CoreWeave's ecosystem accomplishes several key objectives. First, it enhances AI development capabilities by combining CoreWeave's high-performance cloud infrastructure with Weights & Biases' tools for model training, evaluation, and observability, allowing customers to "iterate AI faster". Second, it grants CoreWeave access to Weights & Biases' enterprise customer base, which includes organizations like OpenAI, a major user of the platform. This positions CoreWeave to better compete in the enterprise AI segment. Lastly, it enables cost savings through streamlined operations and provides financing flexibility for future capital expenditures.

OpenPipe – Expanding End-to-End AI Capabilities

In September 2025, CRWV announced an agreement to acquire OpenPipe, a platform for training AI agents with reinforcement learning. OpenPipe's advanced machine learning techniques provide RL capabilities that enable agents to learn from experience, thus increasing accuracy, performance, and reliability over time. The deal builds on CRWV's strategy to deepen vertical integration across their tech stack.

Figure 25: OpenPipe Features



Source: Company Data

The CoreWeave Cloud Platform

CoreWeave's Cloud Platform is a fully integrated solution, purpose-built to support demanding AI workloads—including large-scale model training and inference—with exceptional performance, scalability, and efficiency. The platform comprises 1) Infrastructure Services, 2) Managed Software Services, and 3) Application Software Services, all of which are enhanced by CoreWeave's proprietary Mission Control and Observability software. This suite of tools enables seamless provisioning of infrastructure, intelligent orchestration of AI workloads, and comprehensive monitoring of training and inference environments. These capabilities ensure high availability, reduce operational complexity, and minimize downtime for end users. Architected on a microservices-based framework, the platform's components are fully composable and interchangeable, allowing customers to tailor their deployments based on specific performance, workload, and operational requirements. This modularity provides a high degree of flexibility, enabling clients to configure and optimize their use of the platform without compromising efficiency or reliability.

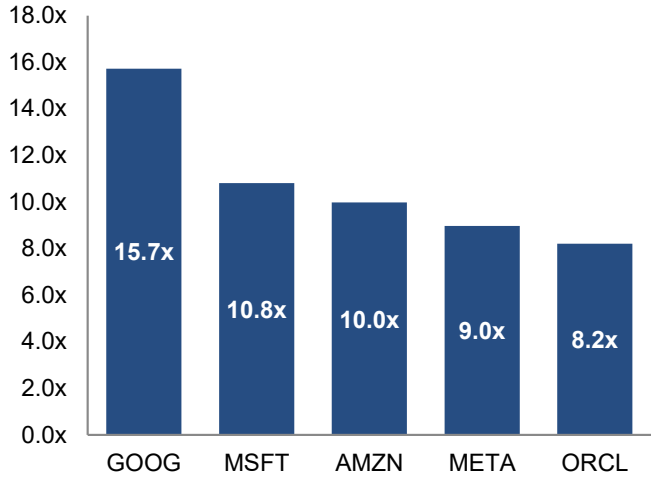
Security is a fundamental component of the CRWV platform. CoreWeave ensures that customers operate within a secure and resilient environment by implementing a Zero Trust model for data access and employing advanced security technologies, including Extended Detection and Response (XDR) and Data Loss Prevention (DLP) across all endpoints. In addition, the platform leverages Single Sign-On (SSO) and Multi-Factor Authentication (MFA) to defend against identity-based cyber threats and uphold the integrity of the CoreWeave Cloud Platform.

Valuation

There have been multiple frameworks discussed and we believe forward EV/EBITDA and EV/EBIT to be the most appropriate valuation methodologies. For Neocloud providers, we think enterprise value valuations based on forward profitability metrics should be burdened with the future debt/capex necessary to generate that profit. Based on that methodology, the neocloud comp set is trading at a high single digit EBITDA multiple. Investors/creditors also evaluate the "EBITDA power" of providers (similar to book but not billed EBITDA for data centers), particularly those with limited operating history. Hyperscalers are another comp set that have been discussed given that for some of them, the GPU (or TPU in the case of GOOGL) cloud business will represent a meaningful share of their revenue. This comp set trades at a low double digit CY27 EV/EBITDA multiple. Lastly, data center REITs have been discussed as a comp set as well given parallels in the business model between DCs and GPU clouds. Companies such as DLR and EQIX trade in the low 20s CY27 EV/EBITDA range. While GPU clouds trade at a discount vs. the hyperscaler and data center comp sets, we believe multiple expansion is possible as

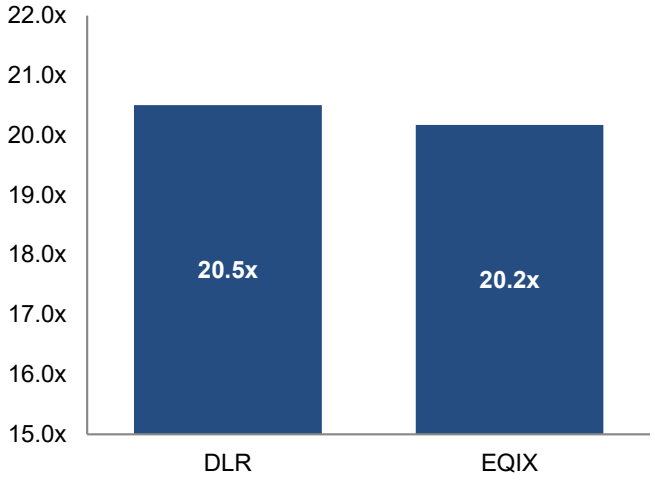
investors become more comfortable with their business models (especially if GPU hour pricing remains durable).

Figure 26: Hyperscaler CY27 EV/EBITDA Valuations



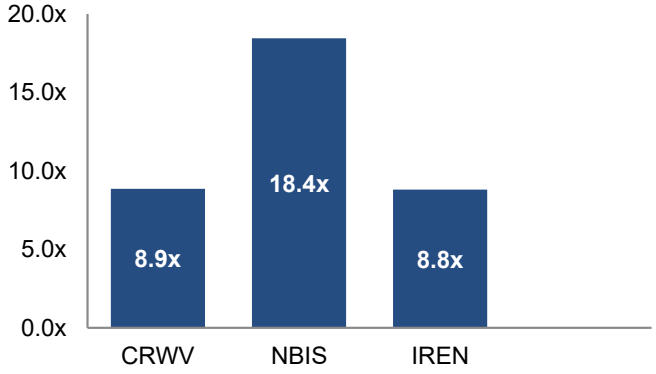
Source: Company Data, FactSet, Evercore ISI Research

Figure 27: Data Center CY27 EV/EBITDA Valuations



Source: Company Data, FactSet, Evercore ISI Research

Figure 28: Neoclouds CY27 EV/EBITDA Valuations



Source: Company Data, FactSet, Evercore ISI Research

Figure 29: GPU Cloud Comp Sheet
($ in millions, except multiples, percentages, and per-share data)

Ticker	Company	Current Price	Market Cap	Enterprise Value	P/E			EV/Sales			EV/EBITDA			EV/EBIT		
					CY25	CY26E	CY27E	CY25	CY26E	CY27E	CY25	CY26E	CY27E	CY25	CY26E	CY27E
Hyperscalers																
MSFT	Microsoft Corporation	$390.49	$2,932,750	$2,883,640	25.6x	21.5x	18.5x	9.4x	8.1x	6.9x	15.9x	13.1x	10.8x	20.4x	17.3x	14.9x
GOOG	Alphabet Inc. Class C	356.18	4,455,970	4,403,029	32.9x	25.0x	24.2x	10.9x	9.0x	7.6x	25.5x	19.6x	15.7x	34.1x	25.9x	21.2x
AMZN	Amazon.com, Inc.	242.67	2,657,773	2,633,866	33.8x	27.5x	24.0x	3.7x	3.2x	2.8x	15.8x	12.7x	10.0x	32.9x	25.3x	19.9x
META	Meta Platforms Inc Class A	582.90	1,566,944	1,535,937	24.8x	17.7x	16.6x	7.6x	6.1x	5.1x	12.6x	10.5x	9.0x	18.4x	17.0x	14.7x
ORCL	Oracle Corporation	140.27	415,406	518,151	20.1x	17.8x	14.4x	8.2x	6.4x	4.6x	14.8x	11.4x	8.2x	19.0x	15.7x	11.9x
Hyperscalers Median					**25.6x**	**21.5x**	**18.5x**	**8.2x**	**6.4x**	**5.1x**	**15.8x**	**12.7x**	**10.0x**	**20.4x**	**17.3x**	**14.9x**
Data Centers																
EQIX	Equinix, Inc.	$1,002.02	$100,609	$120,854	72.8x	58.1x	52.7x	13.1x	11.8x	10.7x	26.7x	23.1x	20.8x	65.4x	50.1x	44.3x
DLR	Digital Realty Trust, Inc.	173.30	63,332	82,778	48.4x	84.0x	78.0x	13.5x	12.3x	11.0x	24.8x	22.4x	20.2x	125.8x	75.1x	63.9x
Data Center Median					**60.6x**	**71.1x**	**65.3x**	**13.3x**	**12.0x**	**10.9x**	**25.7x**	**22.8x**	**20.5x**	**95.6x**	**62.6x**	**54.1x**
Neoclouds																
CRWV	CoreWeave, Inc. Class A	81.75	50,426	72,208	NA	NA	NA	14.1x	7.7x	5.5x	23.7x	13.4x	8.9x	108.4x	99.7x	36.5x
NBIS	Nebius Group N.V. Class A	215.62	66,282	61,285	NA	NA	NA	115.7x	24.2x	10.4x	NA	59.0x	18.4x	NA	NA	NA
IREN	IREN Limited	38.82	15,798	17,525	NA	NA	NA	28.4x	12.3x	6.6x	NA	18.9x	8.8x	NA	NA	24.0x
Neoclouds Median					**NA**	**NA**	**NA**	**28.4x**	**12.3x**	**6.6x**	**NA**	**18.9x**	**8.9x**	**NA**	**NA**	**30.2x**

Source: Company Data, FactSet, Evercore ISI Research

TIMESTAMP

(Article 3(1)e and Article 7 of MAR)

Time of dissemination: July 06 2026 6:00 AM ET

ANALYST CERTIFICATION

The analyst, Amit Daryanani, primarily responsible for the preparation of this research report attest to the following: (1) that the views and opinions rendered in this research report reflect his or her personal views about the subject companies or issuers; and (2) that no part of the research analyst's compensation was, is, or will be directly related to the specific recommendations or views in this research report.

IMPORTANT DISCLOSURES

This report is approved and/or distributed by Evercore Group L.L.C. ("Evercore Group"), a U.S. licensed broker-dealer regulated by the Financial Industry Regulatory Authority ("FINRA"), and Evercore ISI International Limited ("ISI UK"), which is authorised and regulated in the United Kingdom by the Financial Conduct Authority. The institutional sales, trading and research businesses of Evercore Group and ISI UK collectively operate under the global marketing brand name Evercore ISI ("Evercore ISI"). Both Evercore Group and ISI UK are subsidiaries of Evercore Inc. ("Evercore"). The trademarks, logos and service marks shown on this report are registered trademarks of Evercore.

The analysts and associates responsible for preparing this report receive compensation based on various factors, including Evercore's Partners' total revenues, a portion of which is generated by affiliated investment banking transactions. Evercore ISI seeks to update its research as appropriate, but various regulations may prevent this from happening in certain instances. Aside from certain industry reports published on a periodic basis, the large majority of reports are published at irregular intervals as appropriate in the analyst's judgment.

Evercore ISI generally prohibits analysts, associates and members of their households from maintaining a financial interest in the securities of any company in the analyst's area of coverage. Any exception to this policy requires specific approval by a member of our Compliance Department. Such ownership is subject to compliance with applicable regulations and disclosure. Evercore ISI also prohibits analysts, associates and members of their households from serving as an officer, director, advisory board member or employee of any company that the analyst covers.

This report may include a Tactical Call, which describes a near-term event or catalyst affecting the subject company or the market overall and which is expected to have a short-term price impact on the equity shares of the subject company. This Tactical Call is separate from the analyst's long-term recommendation (Outperform, In Line or Underperform) that reflects a stock's forward 12-month expected return, is not a formal rating and may differ from the target prices and recommendations reflected in the analyst's long-term view.

Applicable current disclosures regarding the subject companies covered in this report are available at the offices of Evercore ISI: 55 East 52nd Street, New York, NY 10055, and at the following site: https://evercoreisi.mediasterling.com/disclosure.

Evercore and its affiliates, and I or their respective directors, officers, members and employees, may have, or have had, interests or qualified holdings on issuers mentioned in this report. Evercore and its affiliates may have, or have had, business relationships with the companies mentioned in this report.

Additional information on securities or financial instruments mentioned in this report is available upon request.

Ratings Definitions

Current Ratings Definition

Evercore ISI's recommendations are based on a stock's total forecasted return over the next 12 months. Total forecasted return is equal to the expected percentage price return plus gross dividend yield. We divide our stocks under coverage into three primary ratings categories, with the following return guidelines:

Outperform- the total forecasted return is expected to be greater than the expected total return of the analyst's coverage sector.
In Line- the total forecasted return is expected to be in line with the expected total return of the analyst's coverage sector.
Underperform- the total forecasted return is expected to be less than the expected total return of the analyst's coverage sector.
Coverage Suspended- the rating and target price have been removed pursuant to Evercore ISI policy when Evercore is acting in an advisory capacity in a merger or strategic transaction involving this company and in certain other circumstances.*
Rating Suspended- Evercore ISI has suspended the rating and target price for this stock because there is not sufficient fundamental basis for determining, or there are legal, regulatory or policy constraints around publishing, a rating or target price. The previous rating and target price, if any, are no longer in effect for this company and should not be relied upon.*

*Prior to October 10, 2015, the "Coverage Suspended" and "Rating Suspended" categories were included in the category "Suspended."

FINRA requires that members who use a ratings system with terms other than "Buy," "Hold/Neutral" and "Sell" to equate their own ratings to these categories. For this purpose, and in the Evercore ISI ratings distribution below, our Outperform, In Line, and Underperform ratings can be equated to Buy, Hold and Sell, respectively.

Evercore ISI rating (as of 07/06/2026)

Coverage Universe			Investment Banking Services I Past 12 Months		
Ratings	**Count**	**Pct.**	**Ratings**	**Count**	**Pct.**
Buy	481	61	Buy	92	19
Hold	272	35	Hold	21	8
Sell	11	1	Sell	0	0
Coverage Suspended	8	1	Coverage Suspended	2	25
Rating Suspended	14	2	Rating Suspended	4	29

Issuer-Specific Disclosures (as of July 06, 2026)

Price Charts

GENERAL DISCLOSURES

This report is approved and/or distributed by Evercore Group L.L.C. (“Evercore Group”), a U.S. licensed broker-dealer regulated by the Financial Industry Regulatory Authority (“FINRA”) and by International Strategy & Investment Group (UK) Limited (“ISI UK”), which is authorized and regulated in the United Kingdom by the Financial Conduct Authority. The institutional sales, trading and research businesses of Evercore Group and ISI UK collectively operate under the global marketing brand name Evercore ISI ("Evercore ISI"). Both Evercore Group and ISI UK are subsidiaries of Evercore Inc. ("Evercore"). The trademarks, logos and service marks shown on this report are registered trademarks of Evercore Inc.

This report is provided for informational purposes only and subject to additional restrictions contained in this disclosure. It is not to be construed as an offer to buy or sell or a solicitation of an offer to buy or sell any financial instruments or to participate in any particular trading strategy in any jurisdiction. The information and opinions in this report were prepared by employees of affiliates of Evercore. The information herein is believed by Evercore ISI to be reliable and has been obtained from public sources believed to be reliable, but Evercore ISI makes no representation as to the accuracy or completeness of such information.

Opinions, estimates and projections in this report constitute the current judgment of the author as of the date of this report. They do not necessarily reflect the opinions of Evercore or its affiliates and are subject to change without notice. In addition, opinions, estimates and projections in this report may differ from or be contrary to those expressed by other business areas or groups of Evercore and its affiliates.

Evercore ISI has no obligation to update, modify or amend this report or to otherwise notify a reader thereof in the event that any matter stated herein, or any opinion, projection, forecast or estimate set forth herein, changes or subsequently becomes inaccurate. Facts and views in Evercore ISI research reports and notes have not been reviewed by, and may not reflect information known to, professionals in other Evercore affiliates or business areas, including investment banking personnel. Evercore ISI salespeople, traders and other professionals may provide oral or written market commentary or trading strategies to our clients that reflect opinions that are contrary to the opinions expressed in this research.

Our asset management affiliates and investing businesses may make investment decisions that are inconsistent with the recommendations or views expressed in this research.

Evercore ISI does not provide individually tailored investment advice in research reports. The financial instruments discussed in this report are not suitable for all investors and investors must make their own investment decisions using their own independent advisors as they believe necessary and based upon their specific financial situations and investment objectives. This report has been prepared without regard to the particular investment strategies or financial, tax or other personal circumstances of the recipient. Nothing in this report constitutes investment, legal, accounting or tax advice, or a representation that any investment or strategy is suitable or appropriate to the recipient’s individual circumstances, or otherwise constitutes a personal recommendation to the recipient. None of Evercore Group or its affiliates will treat any recipient of this report as its customer by virtue of the recipient having received this report.

Securities and other financial instruments discussed in this report, or recommended or offered by Evercore ISI, are not insured by the Federal Deposit Insurance Corporation and are not deposits of or other obligations of any insured depository institution. If a financial instrument is denominated in a currency other than an investor’s currency, a change in exchange rates may adversely affect the price or value of, or the income derived from the financial instrument, and such investor effectively assumes such currency risk. In addition, income from an investment may fluctuate and the price or value of financial instruments described in this report, either directly or indirectly, may rise or fall. Estimates of future performance are based on assumptions that may not be realized. Furthermore, past performance is not necessarily indicative of future performance.

Please note that this report was prepared for and distributed by Evercore ISI to its institutional investor and market professional customers.

Recipients who are not institutional investor or market professional customers of Evercore ISI should seek the advice of their independent financial advisor before considering information in this report in connection with any investment decision, or for a necessary explanation of its contents.

Electronic research is simultaneously distributed to all clients. This report is provided to Evercore ISI clients only and may not be, in whole or in part, or in any form or manner (i) copied, forwarded, distributed, shared, or made available to third parties, including as input to, or in connection with, any artificial intelligence or machine learning model; (ii) modified or otherwise used to create derivative works; or (iv) used to train or otherwise develop a generative artificial intelligence or machine learning model, without the express written consent of Evercore ISI. Receipt and review of this research report constitutes your agreement with the aforementioned limitations in use.

This report is not intended for distribution to or use by any person or entity in any locality, state, country or other jurisdiction where such distribution, publication, availability or use would be contrary to law or regulation or which would subject Evercore Group or its affiliates to any registration or licensing requirement within such jurisdiction. If this report is being distributed by a financial institution other than Evercore ISI or its affiliates, that financial institution is solely responsible for such distribution and for compliance with the laws, rules and regulations of any jurisdiction where this report is distributed. Clients of that financial institution should contact that institution to effect any transaction in the securities mentioned in this report or if such clients require further information. This report does not constitute investment advice by Evercore ISI to the clients of the distributing financial institution, and neither Evercore ISI nor its affiliates, nor their respective officers, directors or employees, accept any liability whatsoever for any direct or consequential loss arising from their use of this report or its content.

Evercore is not authorized to act as a financial entity, securities intermediary, investment advisor or as any other regulated entity under Mexican law, and no actions, applications or filings have been undertaken in Mexico, whether before the National Banking and Securities Commission (Comisión Nacional Bancaria y de Valores “CNBV”) or any other authority, in order for Evercore to carry out activities or render services that would otherwise require a license, registration or any other authorization for such purposes; neither is it regulated by the CNBV or any other Mexican authority. This document, nor its content, constitutes an offer, invitation or request to purchase or subscribe for securities or other instruments or to make or cancel investments, nor may it serve as the basis for any contract, commitment or decision of any kind.

For persons in the UK: In making this report available, Evercore makes no recommendation to buy, sell or otherwise deal in any securities or investments whatsoever and you should neither rely or act upon, directly or indirectly, any of the information contained in this report in respect of any such investment activity. This report is being directed at or distributed to, (a) persons who fall within the definition of Investment Professionals (set out in Article 19(5) of the Financial Services and Markets Act 2000 (Financial Promotion) Order 2005 (the “Order”)); (b) persons falling within the definition of high net worth companies, unincorporated associations, etc. (set out in Article 49(2) of the Order); (c) other persons to whom it may otherwise lawfully be communicated (all such persons together being referred to as “relevant persons”). This report must not be acted on or relied on by persons who are not relevant persons.

For persons in Switzerland: The distribution of these materials to persons or entities in Switzerland is not intended to, and does not, constitute a financial service under the Swiss Financial Services Act (FinSA). In particular, the distribution of these materials does not constitute the provision of personal recommendations on transactions with financial instruments (investment advice) within the meaning of Article 3(c)(4) of FinSA.

For persons in Australia: Evercore does not have an Australian Financial Services License ("AFSL"). This report is only to be distributed to persons who fall within the definition of "wholesale" investors under section 761G of the Corporations Act 2001 (Cth). This report must not be acted on or relied on by persons who are not wholesale investors. Evercore relies on the relief provided under ASIC Corporations (Foreign Financial Services Providers—Limited Connection) Instrument 2017/182 to provide this report to wholesale investors in Australia without an AFSL.

For persons in New Zealand: This material has been prepared by Evercore Group L.L.C. for distribution in New Zealand to financial advisors and wholesale clients only and has not been prepared for use by retail clients (as those terms are defined in the Financial Markets Conduct Act 2013 (‘FMCA’)). Evercore Group L.L.C. is not, and is not required to be, registered or licensed under the FMCA, and unless otherwise stated any financial products referred to in this material are generally only available in New Zealand for issue to those satisfying the wholesale investor criteria in the FMCA.

Applicable current disclosures regarding the subject companies covered in this report are available at the offices of Evercore ISI: 55 East 52nd Street, New York, NY 10055, and at the following site: https://evercoreisi.mediasterling.com/disclosure.

In compliance with the European Securities and Markets Authority's Market Abuse Regulation, a list of all Evercore ISI recommendations disseminated in the preceding 12 months for the subject companies herein, may be found at the following site: https://evercoreisi.mediasterling.com/disclosure.