Hardware & Networking

Data Center Watch: AI Demand Supported as LLM Usage and GPU Pricing Climb; Meanwhile, Memory Spot Prices Diverge

IT Hardware/ Telecom & Networking Equipment

Joseph Cardoso [AC]
(1-212) 622-9036
joseph.cardoso@jpmchase.com

Manmohanpreet Singh
(1-212) 622-4527
manmohanpreet.singh@jpmchase.com

Marc Vitenzon
(1-212) 622-3342
marc.vitenzon@jpmchase.com

Akanksh Chauhan
(1-212) 622-0045
akanksh.chauhan@jpmorgan.com
J.P. Morgan Securities LLC

In the Data Center Watch report, we outline key data points relevant to AI infrastructure demand, including LLM token volumes and prices, GPU rental prices, and memory costs. Taken together, the latest data suggest positive trends for the AI infrastructure demand environment, including the expansion of LLM usage, which is more than offsetting year-on-year LLM token price declines, while GPU rental prices continue to grind higher month-on-month. Importantly, based on our sample and simplified assumptions, we estimate that token economics are directionally tracking favorably for model providers, even assuming conservative token prices. Meanwhile, DRAM prices continue to rise, albeit at a slower rate, while NAND spot prices have contracted modestly for three months in a row. *Key highlights from Data Center Watch report*:

- **LLM trends re-accelerate on OpenRouter for the month of June.** LLM trends on OpenRouter in June broadly re-accelerated, with token volume (*+70% m/m, 20x y/y*), pricing (*up modestly m/m but still -5% y/y, with U.S. models bucking the trend by rising y/y*), and spending (*+70% m/m, 16x y/y*) all increasing. Interestingly, while U.S.-based models lost volume share (*35%*), they dominated spend (*85%+*), with Anthropic's Claude Opus 4.7 standing out as the only model ranking top five in both volume usage and spending.
- **GPU rental prices rise across all tiers in the month of June.** GPU rental price sequential growth continued across the board in June for non-hyperscaler capacity, with prices for all three tiers of Nvidia GPUs (A100, H100, B200) tracking higher m/m. A100s led the m/m gains, posting the steepest sequential increase among the three, while H100s and B200s expanded at a more measured pace. Pricing for B200s was ~2x that of H100s, and H100 pricing was ~1.7x that of A100s, with both ratios ticking marginally lower for the second month in a row.
- **Memory prices remained elevated in the month of June but diverged, with DRAM and NAND moving in opposite directions.** DRAM spot prices climbed sequentially again in June, a third straight monthly gain that puts it up more than 8x y/y, while NAND spot prices slipped modestly, a third consecutive small monthly decline after a strong 1Q26 run-up, yet still sitting over 5x higher y/y.

LLM Token Tracker

Our LLM Token Tracker captures a single week of usage and pricing data each month, providing a monthly snapshot or sampling of trends across ~300 LLMs listed on OpenRouter, an aggregation platform that routes API requests to models through a single endpoint. Keep in mind, OpenRouter skews towards developer, startup, and agentic-coding traffic and excludes first-party API volume (direct OpenAI/Anthropic, hyperscaler-hosted endpoints).

Key takeaways:

- **Token volume growth accelerated in the month of June, following a**

See page 5 for analyst certification and important disclosures.

Joseph Cardoso [AC]
(1-212) 622-9036
joseph.cardoso@jpmchase.com

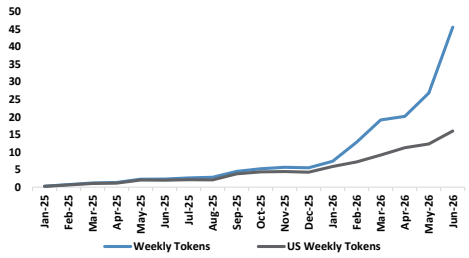
J.P.Morgan

moderation in growth through the start of 2026. Token volume growth accelerated in June, expanding +70% m/m (vs. +33% in May and +5% in April) and 20x y/y (vs. 12x in May and 15x in April). By comparison, volumes for U.S.-based models (OpenAI, Anthropic, Google, xAI, etc.), which represent 35% of the total (vs. 46% in May and 56% in April), also saw growth accelerate in June, expanding +30% m/m (vs. +9% in May and +23% in April) and 8x y/y (vs. 6x in May and 10x in April).

- **Token pricing continued to trend lower year-over-year into June, albeit more modestly than in recent months.** While average token pricing rose +7% m/m in June (vs. +5% in May and +19% in April), year-over-year declines continued, albeit at a more moderate rate, tracking down -5% y/y (vs. -21% in May and -24% in April). We believe this moderation in the rate of decline was led by strong pricing trends for U.S.-based models, as evident in volume-weighted average pricing, which rose +27% m/m for U.S. models (vs. +19% in May and +10% in April) and +77% y/y (vs. +36% in May and +4% in April) vs. flat m/m for total models (vs. -8% in May and +15% in April) and -20% y/y (vs. -25% in May and -23% in April).
- **Token spending also accelerated in June, following a moderation in growth over the past two months.** Token spending growth accelerated in June, expanding +70% m/m (vs. +23% in May and +21% in April) and 16x y/y (vs. 9x in May and 12x in April). By comparison, spending for U.S.-based models, which represent 85%+ of the total (vs. 90% in May and 85% in April), also saw growth accelerate in June, expanding +65% m/m (vs. +30% in May and +35% in April) and 14x y/y (vs. 8x in May and 10x in April).
- **Opus 4.7 ranks as a top five model across both usage and spending.** The top five models in June based on token usage were: 1) DeepSeek V4 Flash; 2) MiMo-V2.5; 3) MiniMax M3; 4) Hy3 preview; and 5) Claude Opus 4.7, accounting for 45% of total token volume. However, on a spending basis, the top five models in June were: 1) Claude Opus 4.7; 2) Claude Opus 4.8; 3) GPT-5.5; 4) Claude Sonnet 4.6; and 5) GLM 5.2, accounting for 75% of total token spend.

Figure 1: Token Volumes

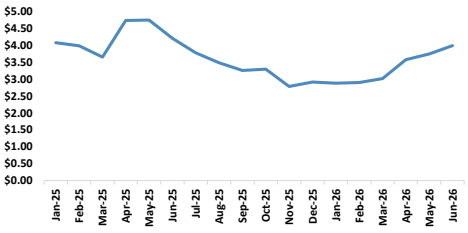
Weekly Tokens in Trillion



Source: OpenRouter, J.P. Morgan Estimates
Note: "U.S. Weekly Tokens" includes U.S.-based models

Figure 2: Average Token Pricing

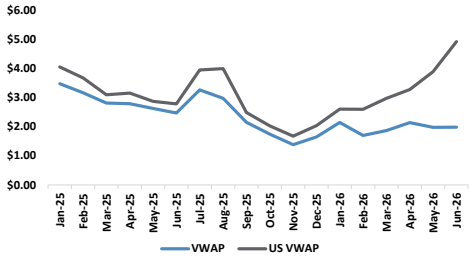
$ per Million Tokens



Source: OpenRouter, J.P. Morgan Estimates

Figure 3: Volume-Weighted Average Token Pricing

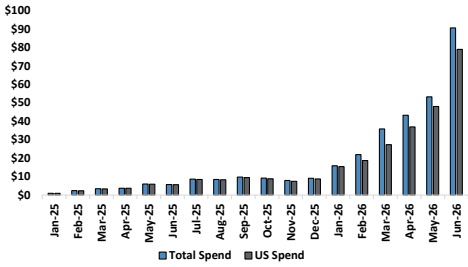
$ per Million Tokens



Source: OpenRouter, J.P. Morgan Estimates
Note: "U.S. VWAP" includes U.S.-based models

Figure 4: Snapshot of Weekly Token Spend

$ in Million



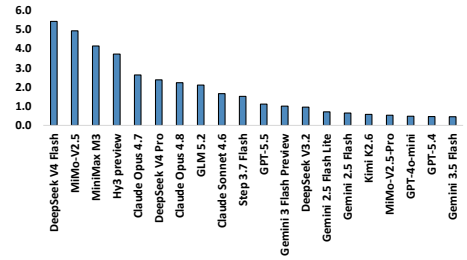
Source: OpenRouter, J.P. Morgan Estimates
Note: "U.S. Spend" includes U.S.-based models



Joseph Cardoso [AC]
(1-212) 622-9036
joseph.cardoso@jpmchase.com

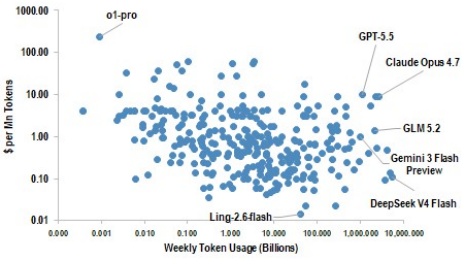
Figure 5: Top 20 Models Based on Volume

Weekly Tokens in Trillion



Source: OpenRouter, J.P. Morgan Estimates

Figure 6: Token Pricing vs. Volume



Source: OpenRouter, J.P. Morgan Estimates

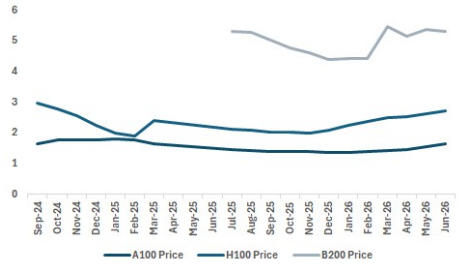
GPU Rental Price Tracker

Key takeaways based on Bloomberg indices:

- **A100 rental prices grew sharply m/m in June; H100-to-A100 ratio ticks lower.** A100 rental prices tracked to $1.63/GPU-hour on average in June, up +6.3% m/m (vs. +6.6% m/m in May and +1.8% m/m in April), representing a fifth straight month of sequential growth.
- **H100 rental prices continued m/m gains in June.** H100 rental prices tracked to $2.72/GPU-hour on average during June, a +3.7% m/m increase (vs. +3.2% m/m in May and +2.1% m/m in April), marking a seventh straight month of sequential growth. The H100-to-A100 price ratio tracked to 1.67x on average in June (vs. 1.71x in May and 1.77x in April).
- **B200 rental prices expanded m/m in June; B200-to-H100 ratio continued to move lower.** B200 rental prices tracked to $5.33/GPU-hour on average in June, up +2.7% m/m (vs. +0.4% m/m in May and +1.3% m/m in April), with the m/m increase trailing the H100 and A100 sequential gains. The B200-to-H100 price ratio tracked to 1.96x on average in June (vs. 1.98x in May, 2.04x in April, and 2.58x around index launch in Sep-25).

Figure 7: Avg. Monthly Rental Prices (Non-Hyperscaler)

$



Source: Bloomberg Finance L.P.

Figure 8: Avg. Monthly Rental Prices (M/M %Change)

%



Source: Bloomberg Finance L.P.

Memory Price Tracker

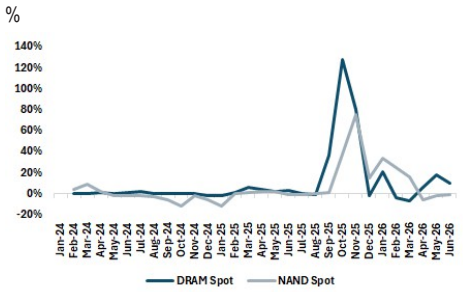
Key takeaways based on Bloomberg data:

- **DRAM spot prices continued to rise in June, with DDR5 16Gb up more than 8x y/y.** DDR5 16Gb spot prices reached $43.14 in June, rising +10% m/m (vs. $39.24 in May and $33.12 in April) and +740% y/y (vs. $5.13 in June-25), marking a third consecutive m/m increase following modest sequential declines in February and March.
- **NAND spot prices decline modestly m/m in June; NAND 1Tb still up over 5x y/y.** NAND 1Tb spot prices tracked to $27.03 in June, a modest decline of -0.3% m/m (vs.

Joseph Cardoso [AC]
(1-212) 622-9036
joseph.cardoso@jpmchase.com

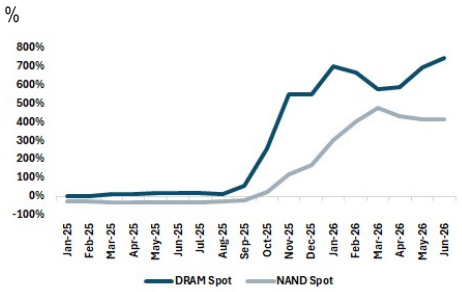
$27.12 in May and $27.43 in April) and an increase of +412% y/y (vs. $5.28 in June-25), marking a third consecutive month of modest m/m declines following the strong run-up through 1Q26.

Figure 9: Memory Prices (M/M %Change)



Source: Bloomberg Finance L.P.

Figure 10: Memory Prices (Y/Y % Change)



Source: Bloomberg Finance L.P.