Americas Technology: Hardware: AI Project Pulse: August 2026

Katherine Murphy
+1(212)902-1151 | katherine.a.murphy@gs.com
Goldman Sachs & Co. LLC

Michael Ng, CFA
+1(212)902-8618 | michael.ng@gs.com
Goldman Sachs & Co. LLC



We summarize recent announcements and developments in the AI project space across neoclouds, sovereigns, and enterprise made in August 2026. This represents select project announcements and is not indicative of the total AI infrastructure opportunity.

Key highlights in August 2026 include:

- **TX, PA Governors tighten requirements for data center developers.** In August 2026, the Governors of Texas (August 3) and Pennsylvania (August 18) signed directives/executive orders that tightened the regulatory, energy, and environmental guardrails for new data center development in their respective states. While these headlines could present uncertainty and slower project development, GS Utilities analyst Carly Davenport expects the actual impact on high-quality, large-scale data center developments will be minimal, with the primary burden falling on speculative, under-capitalized projects. Within our US IT hardware coverage, DELL and SMCI have the most exposure to neoclouds building out greenfield AI data centers.
- **SMCI introduces F2027 guidance for +75% year-over-year revenue growth.** SMCI reported F4Q26 earnings on August 11, where the company introduced FY27 (year-end June 2027) revenue guidance of $65-$72 bn. SMCI's revenue outlook is supported by more than $60 bn of orders received in F4Q, with ~70% tied to "pure AI" deployments/GPU servers and ~30% related to CPU-based AI/CPU demand. Read-thrus from SMCI's results: DELL should also benefit from continued strength in GPU server demand across neoclouds/enterprise, while both DELL and HPE should see uplift from strong CPU-based server demand.

Neoclouds, AI labs, and data center operators

Headlines presented in reverse chronological order.

Goldman Sachs does and seeks to do business with companies covered in its research reports. As a result, 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. For Reg AC certification and other important disclosures, see the Disclosure Appendix, or go to www.gs.com/research/hedge.html. Analysts employed by non-US affiliates are not registered/qualified as research analysts with FINRA in the U.S.

- **IBM and Together AI Sign Multi-Year Agreement to Scale Open-Source AI Inference with NVIDIA AI Infrastructure on IBM Cloud** (*8/11/26*): Following a multi-year, $240 million agreement, IBM and Together AI have partnered to deploy a large-scale artificial intelligence inference cluster on IBM Cloud using NVIDIA HGX B300 systems and NVIDIA Spectrum-X Ethernet networking, with expected availability in the first quarter of 2027.
  - **Key Details:** NVIDIA HGX B300 systems, NVIDIA Spectrum-X Ethernet networking, open-source model inference, 30x more AI factory output, GPU-native data analytics, unstructured data extraction, and Red Hat OpenShift.
  - **Timing:** Expected cluster availability in Q1 2027.
- **Fermi Announces Binding Lease Agreement with TensorWave** (*8/10/26*): Fermi has partnered with AI cloud provider TensorWave to deliver a 222-megawatt turnkey data center solution in its first phase, expected to generate approximately $6.5 billion in total contracted revenue over a 15-year term starting in the second half of 2027.
  - **Key Details:** Turnkey data center lease, 222 MW total facility power for phase one, expansion rights up to 650 MW, tens of thousands of next-generation AMD Instinct GPUs, private electric grid, behind-the-meter power (combined-cycle natural gas, nuclear power, utility grid, solar, battery storage).
  - **Location:** Carson County, Texas.
  - **Timing:** Phased delivery beginning in the second half of 2027; first power targeted for 2026; initial lease term of 15 years with two 5-year renewal options.

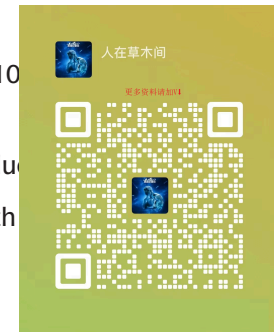
Sovereign

Headlines presented in reverse chronological order.

- **AMD, Cisco and HUMAIN Expand Saudi Arabia's AI Infrastructure as AMD Instinct Systems Go Live** (*8/31/26*): Saudi Arabia's PIF-backed AI company HUMAIN, alongside AMD and Cisco, has brought live an open, large-scale AI platform in the Kingdom powered by AMD Instinct MI355X GPUs, AMD EPYC CPUs, and Cisco Silicon One-based AI networking, with plans to deploy up to 250 MW of next-generation AI infrastructure starting in 2027 as part of a joint venture targeting up to 1 GW by 2030.
  - **Key Details:** AMD Instinct MI355X GPUs, AMD EPYC CPUs, Cisco Silicon One-based AI networking, Cisco N9000 Series platform, Cisco 800G optics, AMD Instinct MI400 Series GPUs, AMD ROCm open software, and GPU-as-a-service.
  - **Key Partners:** AMD, Cisco, HUMAIN, Public Investment Fund (PIF), Dr. Lisa Su, Tareq Amin, and Chuck Robbins.

- **Size:** Up to 250 MW planned for the next phase, with a target of up to 1 GW of AI infrastructure by 2030.
- **Location:** Riyadh, Saudi Arabia.
- **Timing:** Production systems live as of August 31, 2026; next-phase deployment of 250 MW planned to begin in 2027 with capacity coming online in H2 2027; joint venture on track for 1 GW by 2030.

- **IntelliDB Enterprise Partners with Yotta to Deliver Sovereign & AI Database Infrastructure In India** (*8/26/26*): Following a strategic partnership, India's first AI-powered database management platform IntelliDB Enterprise has teamed up with Yotta Data Services to deliver sovereign, AI-ready database infrastructure hosted on Yotta's sovereign cloud and AI infrastructure in India, enabling enterprises and government organizations to modernize mission-critical workloads while ensuring complete data sovereignty and regulatory compliance.
  - **Size:** 20,736 liquid-cooled NVIDIA Blackwell Ultra GPUs (previously announced - February 2026) with 10,000 GPUs allocated to India AI Mission; Supermicro previously announced as one of Yotta's AI infrastructure partners.
- **Mistral and HUMAIN Announce Strategic Collaboration to Advance Sovereign AI in Saudi Arabia and Regionally** (*8/24/26*): Following a strategic collaboration valued in the hundreds of millions of Euros, French open-weight AI lab Mistral and Saudi Arabia's PIF-backed AI company HUMAIN have partnered to develop sovereign AI infrastructure, localize advanced AI models in cybersecurity and voice, and build high-performance Arabic language frontier models for regulated industries in the region.
  - **Key Details:** Sovereign AI infrastructure, advanced model development, localization of advanced AI models (cybersecurity and voice), Arabic language frontier models, open weights, Mistral Compute, and HUMAIN's data center infrastructure
  - **Location:** Riyadh, Saudi Arabia.
- **New Armenia Data Center** *(8/8/26)*: US-based AI cloud & infrastructure company Firebird announced that phase one of Sovereign AI infrastructure has come online
  - Size: $500 mn first-phase investment with plans to scale to more than 400 MW
  - Key Partners: Dell PowerEdge servers with NVIDIA Blackwell GPUs
- **AI Infrastructure startup sets up AI Pods across UK** (*8/1/26*): Following a strategic partnership with Hewlett Packard Enterprise (HPE), UK-based AI infrastructure startup Era4 is deploying a network of sovereign, lower-carbon modular data centers across the UK, utilizing HPE's AI Mod PODs configured with NVIDIA Blackwell B300 GPUs and direct liquid cooling to deliver high-performance compute powered by behind-the-meter renewable energy
  - **Size:** Initial GPU inventory scaling from 1,500 units to over 10
    12 months, with access to over 40 sites across the UK
  - **Location:** Chesterfield, Derbyshire, United Kingdom (first blu
  - **Timing:** HPE Private Cloud AI stack deployed in 40 days, with



onboarded into production within 50 days

Enterprise

Headlines presented in reverse chronological order.

- **Cisco Expands Secure AI Factory with NVIDIA with Supermicro partnership** (*8/25/26*): Following a strategic partnership with Supermicro, Cisco is expanding its Secure AI Factory with NVIDIA to deliver rack-scale, liquid- and air-cooled AI infrastructure, offering an NVIDIA Cloud Partner (NCP) compliant reference architecture designed to support trillion-parameter model training and high-throughput inference for neoclouds, sovereign clouds, and enterprises.
  - **Key Details:** Supermicro liquid- and air-cooled systems, rack-to-fabric liquid cooling, NVIDIA Vera Rubin NVL72, NVIDIA HGX Rubin NVL8, Cisco Silicon One-based switches (front-end), NVIDIA Spectrum-X based switches (back-end)
  - **Key Partners:** Cisco, Supermicro, NVIDIA
  - **Timing:** Supermicro compute solutions orderable through Cisco's authorized channel partner ecosystem beginning in October 2026.
- **Hudson River Trading to Build Next-Gen Research Platform Powered by NVIDIA Vera Rubin NVL72 on CoreWeave Cloud** (*8/20/26*): Following a multi-year, multibillion-dollar agreement, quantitative trading firm Hudson River Trading (HRT) will leverage CoreWeave's purpose-built AI cloud platform—including NVIDIA Vera Rubin NVL72 and HGX B200 GPU systems with Spectrum-X Ethernet networking—to power its next-generation AI-driven trading research and model development at scale
  - **Key Details:** NVIDIA Vera Rubin NVL72 and NVIDIA HGX B200 GPU systems, Spectrum-X Ethernet networking, Direct Connect high-throughput networking, and direct-to-expert technical support.
- **Upgraded AI infrastructure for low-latency financial operations at Indian FinTech** (*8/1/2026*): Following its standardization on Dell Technologies' unified infrastructure, Bengaluru-based fintech provider ZyBiSys has achieved 99.99% availability across six data centers in India, deploying multi-petabyte Dell PowerStore NVMe storage, PowerEdge servers, and PowerSwitch networking to deliver low-latency, real-time trading and compliance visibility to over 50 stockbrokers.
  - **Size:** Multi-petabyte NVMe storage environments, supporting more than 50 stockbrokers across six data centers
  - **Location:** Bengaluru, India.