US Data Center Daily Briefing

March 04, 2026

KEY THEMES

- Akamai deploys Nvidia Blackwell GPUs across 4,000+ locations
- Nvidia invests $2B each in Lumentum and Coherent
- Brookings warns PJM data center strain drove 14% bill rise
- Caterpillar-OnePWR-Vero3 launch first 500MW low-carbon prime power

Akamai is making a very loud bet that AI inference won't live only in hyperscale campuses. With a plan to deploy "thousands" of Nvidia Blackwell GPUs across **4,000+ locations**, it's pitching a decentralised inference fabric that it claims can cut latency by **up to 2.5x** and reduce inference costs by **up to 86%** versus hyperscalers. If that number is even directionally right, it's less a product launch than a direct challenge to where AI margins sit—and who captures them.

The Big Stories

Akamai to deploy Nvidia Blackwell GPUs across 4,000+ locations is the most consequential move in today's stack of announcements because it reframes "AI infrastructure" as an edge distribution problem, not a campus arms race. Akamai says it will roll out Blackwell GPUs, DPUs, and servers across more than 4,000 sites to build a decentralised AI inference platform, building on its October launch of Akamai Inference Cloud. The headline claims—2.5x lower latency and up to 86% lower inference cost versus hyperscalers—are the kind of numbers that force every cloud, CDN, and colo provider to ask a hard question: is inference becoming a placement game where the winner is whoever already owns the last-mile footprint?

Nvidia, for its part, is putting money behind the less glamorous bottleneck: the fabric. In Nvidia invests $2B each in optics firms to boost AI fabric, the company said it will invest **$2 billion** in **each** of Lumentum and Coherent, alongside multi-billion purchase commitments and access rights for advanced laser components. Analysts frame it as an effort to accelerate photonics-based interconnects and ease networking constraints in AI data centres, while also supporting US-based manufacturing expansion to reduce geopolitical supply risk. The connective tissue to Akamai is obvious: whether inference is centralised or distributed, it still lives or dies on interconnect economics and supply.

Grid politics are getting sharper—and the bill is showing up at the customer meter. A Brookings panel in Data centers strain PJM grid, driving up consumer bills warned that rapid data centre additions are straining PJM and have contributed to a **14% rise in customer bills** because new generation hasn't kept pace. The story also flags the policy churn around this: proposed or promised fixes range from on-site energy commitments by Microsoft, Google, and Amazon to a legislative "PRICE Act." The point investors should take away isn't the specific bill (yet); it's that data centres have moved from "economic development" to "ratepayer issue," and that tends to end with constraints, cost-allocation fights, and slower timelines.

Michigan is the PJM argument in a single state-shaped case study. In DTE's 4.4GW pipeline strains Michigan grid, sparks moratoria, DTE is seeking allocation of **over 4.4 GW** to proposed data centre projects—described as equivalent to **six Palisades Nuclear Plants**—and the reporting points to an aging grid and intensifying local/state responses, including moves like Ypsilanti heading toward moratoria. There's also a striking contrast: Indiana Michigan Power credits revenue from AWS's **2.2 GW** complex for planned rate reductions. Put together, it's the emerging US pattern: where the utility and the politics align, data centres are a fiscal lever; where they don't, they're a moratorium headline.

Against that backdrop, the "bring your own power" trend keeps hardening into actual project structure. Caterpillar, OnePWR, Vero3 to develop 500-MW low-carbon power lays out an integrated approach—natural gas prime power plus carbon capture, battery storage, and geological sequestration—with a first **500 MW** prime power project expected to launch this year. Caterpillar will lead FEED and supply generation equipment; OnePWR will build/own/operate under long-term commercial arrangements; Vero3 will develop and operate CO2 sequestration and manage tax credit monetisation. This matters because it's not a vague "decarbonisation" pledge—it's a financing and contracting template designed for customers who can't wait for grid upgrades.

Behind the Headlines

Europe's "sovereign AI" drumbeat is getting operational detail, not just policy language. In Telenor and Red Hat launch Norway-based sovereign AI factory, the companies unveiled a Norway-based Telenor AI Factory built on Red Hat OpenShift AI and OpenShift Platform Plus, explicitly aimed at keeping data inside Norway while providing high-performance AI compute. Skygard's Oslo data centre cluster is slated to support **up to 40 MW** when complete, and Telenor currently offers Nvidia H100 GPUs in DGX H100 systems (no counts or capex disclosed). The subtext is that "sovereign" is increasingly a packaging of three things

customers will pay for: data residency, predictable control planes, and a procurement path that doesn't require signing up to a hyperscaler's full stack.

Liquid cooling is now mature enough that compliance to someone else's spec is the product. Boyd debuts ROL4000 CDU for Project Deschutes liquid cooling says Boyd's new ROL4000 Coolant Distribution Unit meets Google's **Project Deschutes 2 MW CDU specification**, including a **3°C ATD** and **80 PSI** available pressure. It's modular and retrofit-capable, includes **0.2-micron filtration**, redundant power feeds, and even a 230VAC convenience port, with manufacturing and regional support across North America, Europe, and APAC. The interesting shift here is ecosystem control: when an operator spec becomes the market's reference design, CDU vendors are effectively competing on how quickly they can ship "compliant" capacity at scale—and how easily that slots into brownfield retrofits.

India is framing data centre growth as a power-infrastructure scheduling problem, not a demand forecast. In India pre-summit ties AI growth to power infrastructure, Hitachi Energy India and MeitY convened a pre-summit that argued India's AI ambitions require prioritising grid infrastructure, metering, and renewables, citing demand rising from **~8 GW to 16 GW** and India's **500 GW** renewable target by 2030. The value of this kind of convening isn't the headline statistics—it's what it signals about priorities: the bottleneck is being defined early (power and planning), and that tends to pull capital and policy attention toward transmission, interconnection process, and demand-side accountability. For data centre developers, it's a reminder that "market growth" can be gated by grid readiness long before it's gated by land or customers.

Track any ISO, state, county, or company in the US data center build-out — Telborg tracks power, permitting, new projects and legislation exclusively from trusted sources

Telborg Pro · $189/mo →

or book a 20-min call →