

Making Large AI Data-Center Loads Grid-Safe

A power-management approach to fault stability, load volatility, and faster grid interconnection with Lyten's Voltpack Mobile System

Lyten, Inc. | August 2026 | Classification: Public

Summary. AI data centers are the fastest-growing loads on the power grid, and the most difficult to safely interconnect. Left unmanaged, an at-scale AI facility becomes a severe threat to grid stability: it drops offline instantly during routine voltage disturbances (disconnection) and swings its power demand continuously, with individual transitions as fast as milliseconds (volatility). When a gigawatt-scale load trips or oscillates this way, the damage radiates outward: the sudden imbalance can seed the kind of cascading outage that reaches surrounding homes and businesses, while the low-frequency swings mechanically stress the utility's own turbines and transmission equipment. Grid operators are left with two costly options: accept the reliability risk these loads create or slow their development to protect the grid.

There is a third path: make the power load itself grid-safe. **Lyten's Voltpack Mobile System (VMS)** is an integrated power-management platform that transforms volatile data center demand into a predictable, fault-stable, and controllable asset at the point of interconnection (POI). By neutralizing sudden disconnections and high-frequency volatility, the platform addresses the exact reliability risks grid operators are flagging with growing urgency, clearing a faster path to interconnection while lowering costs for developers and ratepayers alike.

The problem

A large AI data center threatens the grid in two distinct ways: abrupt, uncontrolled **disconnection** during faults, and continuous high-frequency **volatility** during normal operation. Both stem from internal data center power architecture.

The first, *abrupt disconnection*, is a side effect of equipment protection: conventional server power supplies and rack-level UPS units are set to safeguard sensitive electronics, so they disconnect the facility at the first sign of a voltage disturbance rather than riding through it. When several large loads trip at once, the grid sees an immediate, systemic loss of load.

This has already destabilized the grid. Recent fault events have pulled well over a gigawatt of data-center load offline in seconds, briefly pushing system frequency up as generation outran

demand, the kind of imbalance that cascades at larger scale. Regulators have escalated: NERC has issued its highest-urgency alert, and its Large Loads Task Force is writing a mandatory ride-through standard for large loads, modeled on rules long imposed on inverter-based generation. Those pressures land directly on ERCOT, the grid operator now facing the largest concentration of this new load in the country.

The second threat is *load volatility*. AI training clusters step their power draw by 30 to 60% within milliseconds and repeat those steps throughout a training run.⁶ This stresses frequency and voltage stability and forces operators to hold expensive, inefficient reserves. A 2025 study by Microsoft, OpenAI, and NVIDIA locates the energy of these swings in a low-frequency band (roughly 0.2 to 3 Hz) that can destabilize long transmission lines and set generating units oscillating against one another (sub-synchronous resonance). Because the emitted frequency shifts as workloads change, utilities must protect a wider band, reaching the higher frequencies where turbine-generator shafts resonate; mechanical stress that shortens equipment life and can force protective trips.

Grid Reliability Risks from AI Data Centers

→ **Two failure modes:** uncontrolled **disconnection** during faults, and continuous **volatility** in normal operation.

→ **July 2024:** one 230 kV line fault caused six voltage dips in 82 seconds and pulled ~1,500 MW offline, spiking system frequency as generation briefly outran demand; the kind of imbalance that can cascade into an outage at a larger scale.¹

→ **May 2026:** NERC issued a Level 3 "Essential Actions" alert citing repeated events since 2022 across the Eastern and Texas interconnections, each with >1,000 MW of computing load disconnected or began oscillating within seconds.²

→ **Coming standards:** NERC's Large Loads Task Force is writing a mandatory voltage ride-through standard for large loads, modeled on the rules long imposed on inverter-based generation.²

→ **In operation:** power draw steps by 30-60% in milliseconds,⁶ recurring in a resonance-prone **0.2 to 3 Hz band**.³

The impact

The same growth that makes these loads valuable makes them a problem for the grid, and it leaves ERCOT with two paths, both costly. They can keep interconnecting large loads as they behave today and absorb the reliability risk. That is the outcome NERC is now warning against: facilities that trip during faults and swing their demand continuously, forcing the system to carry more reserves and build more transmission, much of which is paid for by the ratepayers.

Or they can slow the loads down to protect the grid, denying or delaying interconnection, capping the load a site can draw, and requiring curtailment during emergencies and peak-demand periods. That defends reliability, but it stalls economic development and pushes time-to-power out by years. The queue shows the stakes: as of June 2026, ERCOT was tracking approximately 474 GW of large loads seeking interconnection, up from 63 GW at the end of 2024,⁷ of which about 90% are data centers.⁴ A process built for a few dozen large loads is now managing hundreds, so wait times grow and viable projects stall.

Recent cost reforms compound the delay with financial risk. Under the rules implementing Texas Senate Bill 6, a large-load project must demonstrate site control and financial commitment, pay a minimum study fee of \$100,000, post financial security per megawatt, and cover the cost of the transmission needed to serve it. It is also exposed to a multi-billion-dollar build-out that risks being socialized to ratepayers, alongside new obligations to curtail during grid emergencies.⁵

Regulators are thus forced to choose between slowing economic development and accepting reliability risk and ratepayer cost. There is a third path: make the load itself grid-safe. An AI data center that stays stable at the point of interconnection does not trip during faults and does not swing the grid, so it does not have to be delayed, curtailed, or capped to protect reliability. That changes the equation for powering these data centers. The developer interconnects faster, contracts for less capacity, and reaches full power in a fraction of the time. The grid operator gets a load that supports stability instead of threatening it. The ratepayer avoids subsidizing transmission for a load that can stabilize itself. The rest of this paper describes how.

The solution

Lyten's Voltpack Mobile System (VMS) is a modular power-management platform that installs in parallel between the grid and the data center. It makes the facility behave, at the point where it meets the grid, as a smooth, fault-stable, controllable resource (Figure 1). The platform operates in grid-following mode when a grid is present and grid-forming mode when one is not. Within grid-following, Rapid Peak Shaving (RPS) is the mode that continuously reshapes the facility's volatile demand into a smooth, bounded, grid-safe load.

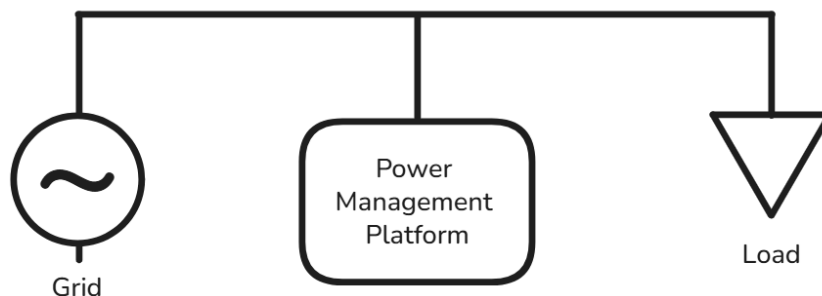


Figure 1. The integrated power management platform sits at the grid boundary, in parallel with the facility. VMS clusters connect at the low-voltage bus (400/480 VAC), or optionally at medium voltage, between the grid and the data-center load. Internal controls are not shown.

At the grid boundary, the VMS delivers:

- **Controlled fault survival (Figure 2).** Instead of cascading offline during a voltage disturbance, the host facility rides through it. The moment internal computing equipment disconnects, the platform draws an equivalent load in its place, preventing abrupt load rejection and the over-frequency excursion (the spike in grid frequency) that would otherwise follow. It then manages a controlled ramp-down as the servers safely return online. The VMS reaches full compensating load within a single grid cycle; this ride-through capability directly resolves the reliability threat now drawing regulators' highest concern.

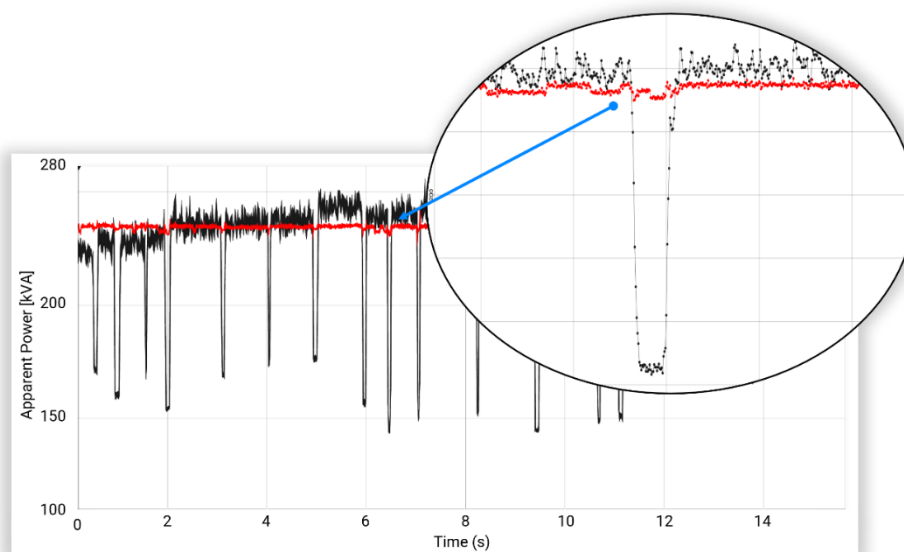


Figure 2. Measured data-center load apparent power (black) and the resulting grid-side apparent power at the point of interconnection (red), from Lyten testing. The platform holds the grid-side profile flat through large, rapid swings in facility demand.

- **A smooth, firmly bounded load.** Running continuously, RPS absorbs the data center's fast, volatile load steps so the grid operator simply sees a steady, predictable load held to a firm setpoint (Figure 2). In controlled testing to date, it has flattened up to 99.5% of load-step volatility and cut grid-side power variation (standard deviation) by up to 94.5%, holding grid voltage within about $\pm 0.4\%$, as much as 22 times tighter than standard power-quality limits require. It also corrects power-quality disturbances as it serves the load, keeping them off the grid. The smoothing is reactive rather than predictive, responding in real time so that any workload is smoothed without prior configuration. In RPS mode, compensation is local and automatic, with nothing exported to the grid, and response is fastest because control sits inside the system.
- **Optional grid-services participation.** In Grid Boost mode the platform instead operates as a standard energy storage system, taking commands from an external energy-management system to deliver frequency response, reactive support, and dispatchable curtailment, and to earn revenue in ancillary-service and reliability programs.
- **Grid-forming operation when there is no grid.** In RPS and Grid Boost the platform follows the grid's voltage reference. It can also create that reference itself, supplying backup power and black-start loads with no grid present, then resynchronizing when grid power returns. Transitions in both directions are seamless, with no interruption to the load.

Unlike an emergency buffer that sits idle until a fault, Lyten's platform works continuously. In testing, it was actively charging or discharging ~98% of the time at full data-center load, drawing power when the facility sheds load and releasing it when demand surges. The remaining fraction is the instants when facility demand and grid draw coincide and there's nothing to compensate.

The rest of the industry is converging on the same answer. The 2025 Microsoft, OpenAI, and NVIDIA study³ ranks energy storage as the most capable mitigation, and defines what such a system must do: measure the load directly, hold enough capacity for the workload, meet sudden rises and drops, and switch between charging and discharging quickly. The VMS meets all four, where the two in-facility alternatives it evaluates fall short:

- **Software smoothing** (at the GPU or server level) can typically raise the power floor only to about 90% of rated power, still leaving a peak-to-peak swing of roughly 20% of rated power; not enough where utilities impose tighter limits (e.g., $\pm 10\%$). Holding that floor also means running at a higher sustained power level, adding about 10% to continuous energy use as a **permanent operating-cost penalty**. A facility paired with the platform does not need to hold that elevated floor, because the fluctuations are absorbed before the grid sees them.

- **Hardware smoothing** at the rack (ultracapacitors, or uninterruptible power supplies (UPS)) can buffer short, fast fluctuations but cannot flatten the sustained load, and does nothing for mechanical loads such as cooling, which can be up to a third of a data center's total demand.

Both approaches share a fundamental limitation: they are confined within the facility. While they mitigate internal variations, they fail to make the total load fault-stable or controllable at the grid boundary. The disconnection and load-rejection failures that triggered NERC's alert occur at the grid interface, and they can only be solved there, precisely where the VMS operates.

Independent testing confirms that the VMS's fault response already complies with some of the most stringent distributed-resource grid codes in the world, surpassing ERCOT's current ride-through requirements. Drawing a controlled load to offset disconnecting server equipment is exactly the certified ride-through behavior U.S. regulators are preparing to mandate. Because upcoming federal large-load standards are modeled on these exact inverter-based resource rules, the platform provides a field-proven, ready-to-deploy compliance pathway. Meeting local standards requires simple software tuning rather than a fundamental hardware redesign.

The platform ships as modular blocks that install in hours, so grid-stable operation is available from a project's earliest phase. They support construction and commissioning loads, let a facility energize in a controlled, phased way, and scale up incrementally as load grows. Every system is sourced and manufactured on a secure, localized supply chain.

Why it works: economically and for the grid

Time-to-power is the single largest cost data-center developers carry, and this platform solves it directly. The economics line up for the developer and the grid at the same time:

- **Faster Time-to-Power:** By installing in hours and enabling controlled, phased commissioning, the platform allows projects to energize and scale up without waiting years for deep transmission infrastructure buildouts.
- **Lower Contracted Capacity:** Capping grid draw to a strict, predictable setpoint reduces the peak capacity a site must contractually reserve. This lowers upfront interconnection fees, minimizes the need for socialized transmission upgrades, and eliminates costly demand-charge penalties. The same study notes that storage can lower a training cluster's peak power draw by charging during the low-power communication phases.
- **Deployable Fleet Flexibility:** Because the modular blocks are transportable, a single asset fleet can de-risk multiple sequential buildouts and early-stage operations rather than sitting stranded at a single completed site.

This is the third path: reliability for the grid and fast time-to-power for the developer, without interconnection denials, load caps, or emergency curtailment. Lyten welcomes active engagement with ERCOT and other system operators to transform massive AI data center loads from systemic grid liabilities into stable, predictable assets.

To learn more or discuss a deployment, visit lyten.com/products/energy-storage-system.



Sources

1. North American Electric Reliability Corporation, *Incident Review: Considering Simultaneous Voltage-Sensitive Load Reductions* (January 2025), [nerc.com](https://www.nerc.com).
2. NERC, Level 3 "Essential Actions" Alert on computational loads (issued May 4, 2026), [nerc.com](https://www.nerc.com); coverage in *Utility Dive* and *Renewable Energy World*, April–May 2026.
3. E. Choukse, B. Warrier, S. Heath, et al. (Microsoft, OpenAI, NVIDIA), "Power Stabilization for AI Training Datacenters," arXiv:2508.14318 (August 2025).
4. ERCOT, ERCOT Update, presentation to the Texas Senate Committee on Business & Commerce (Pablo Vegas, President and CEO, July 29, 2026), slide 6: "Large Load Interconnection Requests (as of June 2026)." https://www.ercot.com/files/docs/2026/07/29/ERCOT-Senate-July-29-Panel-1-Assessing-The-Grid.pdf?utm_source=twitter-x&utm_medium=post&utm_campaign=Lege-2026&utm_content=2026
5. Texas Senate Bill 6 (2025), implemented through PUCT rulemaking (Project No. 58317); analysis in Weil, Gotshal & Manges, "Senate Bill 6 Reforms Interconnection and Co-Location Rules for Data Centers and Other Large Loads in ERCOT" (July 2025).
6. Arora, C., Sorel, M., & Sachdeva, P. (2025). The next big shifts in AI workloads and hyperscaler strategies. In McKinsey & Company. <https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/the-next-big-shifts-in-ai-workloads-and-hyperscaler-strategies>
7. Baraniuk, C., "ERCOT's large load queue has nearly quadrupled in a single year," Latitude Media (November 2025), <https://www.latitudemedia.com/news/ercots-large-load-queue-has-nearly-quadrupled-in-a-single-year/>