LARGE-LOAD GRID INTEGRATION: A PRIMER

The Eight Problems — and the Decade That Frames Them

ATX Watches

An analytical review of interconnection queue pressure, load forecasting, resource adequacy, cost allocation, dynamic load behavior, co-location, demand flexibility, the federal-state jurisdictional divide, and the off-grid escape across the North American bulk power system.

July 13, 2026 · Updated August 1, 2026 (v1.31)

Data current through July 21, 2026. Regulatory positions in this area are moving quickly; several proceedings cited here have filing deadlines within the next 60 days. Changes since the original July 13 publication are recorded in **Latest Updates** and integrated into the body at the point of use.

DOI 10.5281/zenodo.21464969

Cite this version as: Zavadsky, V. (2026). *Large-Load Grid Integration: A Primer: The Eight Problems — and the Decade That Frames Them* (v1.31). Zenodo. https://doi.org/10.5281/zenodo.21464969

Abstract

Large-Load Grid Integration: A Primer examines how gigawatt-scale electricity demand from data centers, artificial intelligence, cryptocurrency mining and other large industrial loads is reshaping electric system planning across the North American bulk power system. It treats eight interrelated problems on a 2018-2035 timeline: interconnection queue integrity and speculative load; resource adequacy against development timelines; cost allocation and cost-shifting to existing ratepayers; large-load ride-through and dynamic performance; co-location and behind-the-meter generation; flexible and curtailable service; the federal-state jurisdictional divide; and off-grid development. The report argues that these constitute stages of one planning process rather than eight independent failures. It draws engineering principles, planning methodology, reliability assessment, regulatory proceedings and market reform at FERC, NERC, ERCOT, PJM, MISO, SPP, CAISO, NYISO and ISO-NE into a single technical reference, and adds comparative regional and international assessments, a forward-looking regulatory calendar, and a technical annex. Written for utility and ISO/RTO engineers and planners, regulators and their staff, policymakers, developers, consultants, researchers, investors and counsel.

Keywords large load interconnection; data center electricity demand; interconnection queue; load forecasting; resource adequacy; cost allocation; ratepayer impact; load ride-through; dynamic load modeling; co-location; behind-the-meter generation; demand flexibility; curtailable load; bulk power system reliability; electricity regulation; capacity markets; transmission planning; FERC; NERC; ERCOT; PJM; MISO

About This Report

Purpose. *Large-Load Grid Integration: A Primer* examines how gigawatt-scale demand from data centers, artificial intelligence, cryptocurrency mining and other large industrial loads is reshaping electric system planning across North America. It brings engineering principles, planning methodology, regulatory development and market reform into a single technical reference, on the view that the underlying engineering and the policy response cannot usefully be read apart.

At a glance

Who should read it	Utility and ISO/RTO engineers and planners, regulators and their staff, policymakers, developers, consultants, researchers, investors, and counsel.
What makes it different	It takes neither a single ISO nor a single discipline nor a single proceeding as its frame. Engineering, planning, operations, markets and regulation are treated together, across North America, because the problems do not separate in practice.
What it covers	Eight interrelated problems — queue integrity, resource adequacy against development timelines, cost allocation, ride-through and dynamic behaviour, co-location, demand flexibility, jurisdiction, and off-grid load — on a 2018–2035 timeline.
What it is not	Not an interconnection handbook, modelling manual, protection guide or design standard. It explains how studies, planning processes and regulatory frameworks interact, and where each fails under load of this scale and speed.
How to use it	End to end, or by entry point: Key Takeaways for the argument, Figure K1 and Table K1 to locate a problem in the interconnection sequence, the Technical Annex for the engineering beneath a section.
Sources	102 reference entries and a catalog of 87 orders, rules and directives — FERC, NERC, ISO/RTO tariffs and stakeholder material, state dockets, technical literature. Where a figure reaches the report through trade press rather than a filing, the text says so.
Status	Current through July 21, 2026. Changes since publication are dated and tagged in **Latest Updates** and worked into the body at the point of use, so the report always reads as one current document.

How it is organised

Part	What it does
Key Takeaways	The principal findings, each confidence-tagged.
Executive Summary	The evidence: load growth against build times, market outcomes, the reliability record.
Report at a Glance	The eight problems in one table, with what each causes if unaddressed.
A Short Primer	How the physical system and its institutions work.
The Interconnection Process	Figure K1 and Table K1 — the chain from request to ratepayer, and where each problem attaches.
Sections 1–9	The eight problems, then a 2018–2035 timeline of what each institution decided.
Synthesis and Central Findings	The eight issues drawn together as one planning problem, and the report's central conclusions — preceded by the international comparison and the cross-cutting institutional view.
Technical Annex T.1–T.11	Modelling, study practice, forecasting, adequacy, reliability metrics, operations.
Latest Updates	The running change log — every revision dated, tagged, already integrated above.
Catalog and References	87 orders, rules and directives cited, and the full bibliography.

THE GUIDING PRINCIPLE

The report's central premise is that these are not eight isolated problems but connected consequences of a planning framework built on three assumptions about demand: that load arrives slowly, behaves predictably, and stays small relative to the system. Large computational load breaks all three, and each break surfaces at a different point in the same process. Understanding those interactions is the precondition for durable solutions rather than local fixes undone one step later.

◆ A NOTE ON JUDGMENT

Analytical judgments are confined to **Assessment** boxes, each labelled and confidence-tagged — Settled, Likely, Emerging or Speculative — so interpretation reads apart from the record it rests on.

Contents

Section numbers 1–9 are the report's core; the Technical Annex (T.1–T.11) provides the engineering detail beneath them. **Upcoming Dates and Events** is the operational calendar; **Latest Updates** is the running change log.

Key Takeaways

A summary of the report's principal findings, organized as answers to three questions, with a confidence tag on each point so settled facts stand apart from informed expectations. Confidence key:

SETTLED adopted rule or occurred event

LIKELY proposal on a clear path / strong evidence

EMERGING proposal or position taken but not yet resolved

SPECULATIVE genuine forecast — direction supported, date or magnitude not

1. The three risks that dominate the record

- Load behaviour that the system was not built to absorb. Gigawatt-scale computational load disconnects itself in under a cycle during a normally cleared transmission fault, converting a disturbance the grid should absorb into a large, uncommanded loss of demand. The record is continuous rather than anecdotal: 26 ERCOT events above 100 MW between January 2023 and September 2025, and about 1,500 MW across 60 delivery points in northern Virginia in July 2024 during faults cleared correctly in 42 to 66 milliseconds — the archetype, because nothing malfunctioned. A larger west Texas event in December 2022 removed about 1,600 MW and took twelve and a half minutes to recover, though its clearing was abnormal and its load mixed (Section 4). ERCOT studies indicate that frequency excursions become significant near about 2,600 MW of instantaneous loss, and has already reduced operating limits on some interfaces because the possibility has itself become an operating contingency. Ramping is ungoverned everywhere: training clusters swing hundreds of megawatts in seconds, and no North American jurisdiction presently imposes a generally applicable ramp-rate limit on an individual large computational load, though individual service agreements may contain one. **SETTLED**
- Cost allocation that shifts large-load costs onto existing ratepayers. PJM's independent market monitor estimates roughly 40–45% of recent capacity-auction cost as attributable to data-center load, much of it for facilities not yet built. Capacity-driven increases already visible on household bills run approximately $10–20 a month in affected PJM zones. Because capacity prices clear against forecast load, the system socializes those costs today on the basis of projects that may never exist — which makes queue integrity a precondition for solving cost allocation rather than a parallel effort, and one of the principal areas of policy debate in this system expansion. **SETTLED**
- A timing inversion between load arrival and infrastructure delivery. Load reaches commercial operation in 18 months to three years; the generation, transmission, and rules meant to serve it take roughly four years, seven to ten years, and several years respectively. Planning cycles therefore commit capital and set tariffs against demand that arrives long before the system can serve it. NERC's own assessment places MISO and PJM in elevated shortfall risk from roughly 2028 and 2029 — before the remedies take effect. Two factors constrain infrastructure deployment: an equipment supply chain booked through 2029–30, and unresolved politics over who pays. **LIKELY**

2. The reforms that matter most

- Flexible, curtailable load service. Among the largest levers available. Giving up 0.25 to 0.5% of annual consumption — partial reductions falling in 85 to 177 hours a year, averaging about two hours each — could accommodate a modelled 76 to 98 GW of new demand on the existing grid, on the same timescale the load itself arrives. The underlying study states its participation and curtailment assumptions, and does not model network constraints. Current evidence indicates the principal barriers are contractual and market-design related rather than technical. **LIKELY**
- NERC registration of computational loads. Among the few levers that reach these loads even when they island off-grid, and the precondition for the rest: a Reliability Standard binds only the entities NERC has registered. NERC targets a first mandatory standard for year-end 2026; the registration criteria determine which entities that standard actually binds. **On July 16, 2026 FERC made this schedule binding** in Docket RD26-7-000, directing NERC to file

both the standards and computational-load registration criteria by December 31, 2026. The deadline is therefore settled; only its content remains open. **LIKELY**

- Mandatory ride-through obligations on load. ERCOT’s NOGRR282 and NPRR1308 impose frequency and voltage ride-through on Large Electronic Loads — the first such requirements placed on demand in North America — alongside dynamic modeling and mandatory registration under NPRR1325. Facilities that had cleared energization approval or completed their interconnection study by November 14, 2025 are exempt, a population ERCOT identifies as a continuing reliability exposure. **SETTLED**
- Queue and forecast integrity. ERCOT's Batch Zero queue screen (live July 11, 2026), the $50k/MW financial security requirement, and a PUCT duplicate-request disclosure rule due by December 2026. Addressing speculative queue entries could substantially reduce future cost-allocation disputes, since costs are socialized today on the basis of load that may never exist. **SETTLED**

3. What to watch over the next 12–24 months

- Year-end 2026. The first NERC large-load reliability standard and the PUCT duplicate-disclosure rule — both required to convert guidance into obligation. Under FERC's July 16, 2026 order in RD26-7-000 the NERC filing (standards *and* registration criteria) is now a compliance deadline, with a Phase II work plan due March 1, 2027. **LIKELY**
- Across 2027. FERC's large-load rulemaking maturing from its advance notice (ANOPR) to a proposed and then final rule, and the co-location and flexible-load service terms that follow. **EMERGING**
- 2028–2029. The years NERC's assessment marks as the onset of elevated shortfall risk (MISO ~2028, PJM ~2029). That interval — between the onset of the risk and the readiness of the remedies — defines the period to plan around. **EMERGING**
- An early indicator. Whether any jurisdiction writes a bounded, priced, predictable curtailment product — a stated annual hour cap, limits on event duration and frequency, a notice window, and compensation. A jurisdiction that does so first should attract a disproportionate share of investment, for two reasons established later in this report: time to energization, rather than electricity price, has become the principal siting constraint for many hyperscale data-center developers (Section 2), and developers are already accepting curtailment terms bilaterally in exchange for earlier connection (Section 6). A tariffed version converts a negotiation available only to the largest firms into a published product any developer can price. **EMERGING**

The Assessment boxes throughout carry the author's reasoned judgment rather than neutral summary or cited fact — each labeled and confidence-tagged so it reads apart from the record. Points tagged Emerging in particular remain contestable, and reasonable analysts weighing the same evidence may land differently.

Executive Summary

The North American grid now absorbs the fastest load growth in decades. Within the last twelve months, the major North American reliability institutions that govern it — FERC, NERC, and the RTOs/ISOs — reached the same conclusion: the rules written for a flat-demand system are no longer sufficient for today's pattern of large-load growth. A single customer can now request more power than a mid-sized city. The available data indicate a substantial mismatch between projected load growth and infrastructure development timelines.

The scale of the problem

ERCOT now tracks more than 438 GW of large-load interconnection requests — roughly five times the all-time system peak of 85.5 GW — with close to 90% of that volume coming from data centers. Over the same period it approved on the order of 2.2 GW to energize, and about 9 GW cumulatively as of April 2026 — two different quantities, and the report uses both. (ERCOT notes that its own 438 GW queue figure carries duplicate and speculative requests — see Section 1 — so it indicates pressure rather than demand.)

Why existing rules no longer fit

ERCOT is not alone in recognizing the problem. In responding to FERC's June 18 order, MISO indicated that its tariff provides no consistent or transparent framework for evaluating large loads — an admission that carries particular weight because MISO has seen the fastest data-center growth of any region. FERC staff put that growth at 43% compound annual growth since 2020 in the 2025 State of the Markets report of March 19, 2026, against 24% nationally over the same period, and the Commission repeated the finding in the June 18 orders. Every use of the figure in this report draws on that one dataset. The queue-integrity problem is therefore not a Texas peculiarity arising from ERCOT's connect-and-manage model. It appears wherever load growth outruns the processes built to evaluate it.

Capacity-market outcomes point in the same direction. PJM's Base Residual Auction cleared at $28.92/MW-day for the 2024/25 delivery year, then at $269.92 for 2025/26, $329.17 for 2026/27, $333.44 for 2027/28 and $325 for 2028/29. Two of those auctions fell in 2025 because PJM was compressing its schedule back toward a three-year-forward cycle, so the series covers five delivery years rather than five calendar years. The last three cleared at the price cap in force at the time, and the nominal decline in the most recent one reflects the extended collar rather than any supply response. PJM's independent market monitor attributes roughly 40–45% of recent capacity-auction cost to data-center load, most of it not yet built.

Reliability data are consistent with the same conclusion, and this is the part of the problem least represented in public debate. NERC has documented disturbances in which about 1,500 MW of computational load disconnected itself in response to a fault that never threatened it, and ERCOT has recorded 26 further events above 100 MW between January 2023 and September 2025. These are not outages in the ordinary sense: no equipment failed, protection operated exactly as designed, and the load left anyway. The consequence is that a single customer class can now produce contingencies approaching the loss of a large generating unit. ERCOT has therefore reduced operating limits on some interfaces, and NERC issued its most serious non-standard instrument, a Level 3 alert, in May 2026.

Some headline figures in this report come through trade press or analyst summaries rather than from the filing itself. Where such a figure carries weight, the text says where it came from and notes any dispute about it, both at the point of use and in the References.

This report isolates the eight issues that dominate the record and treats each in its own section. The first four concern the physical and financial system: (1) queue congestion and speculative load, including how load forecasting turns requests into planning numbers; (2) the resource-adequacy and development-timeline gap; (3) cost allocation and cost-shifting to existing ratepayers; and (4) large-load ride-through and dynamic performance. The remaining four concern the terms on which load connects: (5) co-location, behind-the-meter generation, and electrically proximate load; (6) flexible and curtailable load service, the one issue that also supplies the leading solution to several of the others; (7) the federal-state jurisdictional divide, which determines whether anyone holds authority to impose the first six; and (8) off-grid and islandable load, where on-site supply shifts the system impact rather than eliminating it.

A final section (9) sets all of this on a single 2018–2035 timeline — the actual decisions of each institution alongside the proposed and projected milestones — to show how the reforms and the reliability risks run on different clocks. Four structural observations run through all eight:

- Load now behaves as a dynamic bulk-system element rather than a passive boundary condition. AI and crypto facilities ramp, oscillate, and trip. The planning and standards framework still treats load as an inert P-Q injection, and NERC has said explicitly that this is inadequate.
- Time remains the principal constraint, and demand flexibility appears to offer one of the few responses that operate on comparable timescales. Every serious proposal — SPP's high-impact large-load (HILL) process, PJM's non-firm service, ERCOT's Controllable Load Resource pathway, MISO's expedited-study routes (ERAS/EPR) — trades some firmness for a faster energization date.

SPP has since taken the further step. On June 5, 2026, FERC accepted Conditional High Impact Large Load Service (CHILLS), which offers large loads long-term transmission service ahead of the upgrades normally required, conditioned on the load accepting curtailment and on separate telemetry and billing that expose its real-time draw to the operator. A load that exceeds a declared cap may be curtailed. CHILLS is the closest thing in force to the bounded, priced, predictable curtailment product this section identifies as the missing instrument — and the first place to look for evidence on whether developers will actually accept one.

- The regulatory posture shifted from study to compliance filing in June 2026. FERC's six simultaneous Section 206 show-cause orders put every RTO on a 60-day clock, with responses due August 17, 2026 — a date that may move in two regions: ISO-NE and its transmission owners have said they will ask for the 90-day abeyance the orders allow, and CAISO has published a compliance schedule that assumes one and ends in a November 16 filing.
- No single institution has clean authority over the problem. FERC can reach the wires but not the retail customer; the states can reach the customer but not the interstate system; NERC can reach reliability only over entities it first registers. The fragmentation of the remedies follows directly from the fragmentation of the authority.

Figure 6 — Regulatory sequence, Oct 2025 → year-end 2026

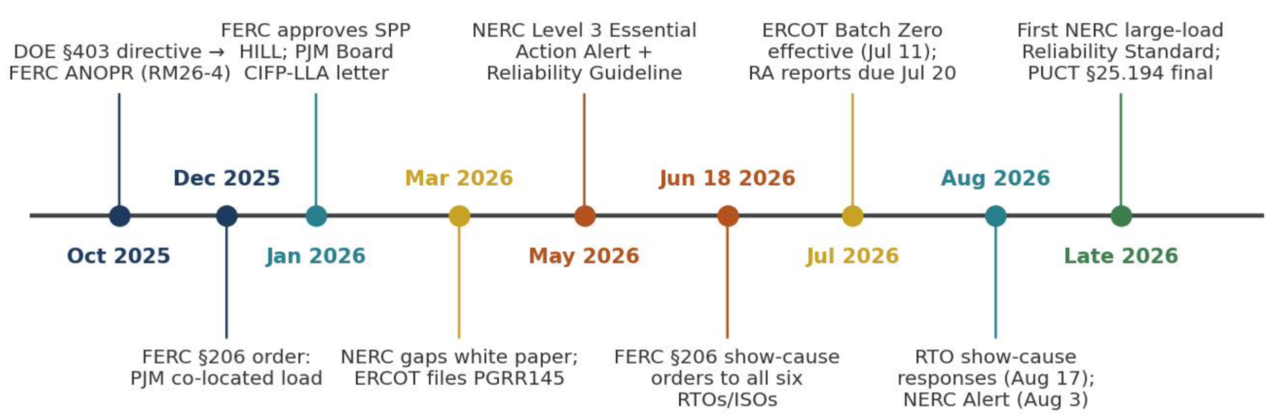


Figure 6 — The compressed regulatory sequence, from the DOE §403 directive through the standards and tariff filings due in late 2026. The interval matters because each step depends on the one before it: registration criteria determine which entities a standard can bind, and a tariff filed before the standard exists cannot reference it. The sequence leaves little slack, so a delay at any point moves every subsequent date rather than compressing the remainder.

Report at a Glance

The eight issues, what each one causes if unaddressed, the reforms now in motion, and the bodies driving them. Section numbers link each row to its full treatment.

#	Issue	Core consequence	Reforms in motion	Lead bodies
1	Queue congestion & speculative load	Planning against load that may never appear; real projects delayed behind speculative ones	Batch Zero; $50k/MW security; forecast gating; duplicate disclosure	ERCOT/PUCT, FERC
1	Load forecasting (within §1)	Same queue yields forecasts differing 3×; over- or underbuild	Probability-weighted, project-based forecasts; agreement-gated inputs	PJM, ERCOT, NERC
2	Resource adequacy & timeline gap	Load builds ~3× faster than generation or wires; first PJM capacity shortfall	Expedited studies (ERAS); capped backstop procurement; BYOG	MISO, PJM, SPP
3	Cost allocation & cost-shifting	Data-center costs socialized to households; $10–20/mo already visible in affected PJM zones	“But-for” assignment; large-load rate classes; load-billed backstop	FERC, states, PJM
4	Ride-through & dynamic performance	GW-scale load trips uncommanded on faults it should ride through	NERC Level 3 Alert; Project 2026-02 standard; ERCOT NOGRR282	NERC, ERCOT
5	Co-location & behind-themeter	Tariff unclear on grid reliance; capacity double-counted	Firm/Non-Firm Contract Demand; CIR adjustment; WLPUN/PCLR	FERC, PJM, ERCOT
6	Flexible & curtailable load	Missed headroom: ~76–98 GW absorbable with modest curtailment	Non-firm service; Controllable Load Resource; DR expansion	All RTOs, FERC
7	Jurisdiction & the non-RTO grid	No single body has clean authority; a 50-state tariff patchwork	Transmission-only federal baseline; §205 invitation; NERC registration	FERC, states, NERC
8	Off-grid & islandable load	On-site generation lets GWscale load leave the grid uncommanded	Registration regardless of draw; coordinated islanding; standby tariffs	NERC, ERCOT, states

Table G1 — The eight problems in one view. Read the third column across and the shape of the report appears: every consequence is either a cost that lands on someone who did not cause it or a reliability exposure that no single body owns. The fourth column is what is already moving; the fifth is who has to move it.

Sources: ERCOT; PJM; MISO; SPP; FERC; NERC; PUCT. Each row is treated in the section indicated.

Reform maturity varies by row: queue gating is adopted and in force; ride-through has moved past NERC's May 2026 Level 3 Alert to mandatory rules in ERCOT, with a NERC standard directed for year-end 2026; co-location and jurisdiction remain contested.

Each section's Assessment box carries a Settled / Likely / Emerging / Speculative confidence tag (see the Key Takeaways key). The tag records how well supported a judgment is, not how far the reform it concerns has matured: an adopted rule can carry an Emerging assessment of its effect, and a contested reform can rest on a Settled fact. Speculative is reserved for genuine forecasting — a projection of what a body will do or a market will produce more than about a year out, where the evidence supports a direction but not a date or a magnitude.

A Short Primer: How the Grid Serves a Large Load

Readers who work in power systems can skip ahead to Section 1. For everyone else, a few paragraphs on how electricity actually reaches a data center will make the regulatory detail that follows far easier to read. Electricity moves in one direction through the four stages shown in the figure below. Generators — gas plants, wind and solar farms, nuclear units, increasingly batteries — produce the power. Transformers step it up to high voltage, and the transmission grid — the network of large towers and lines that forms the “bulk power system” — carries it long distances. Substations near the point of use step the voltage back down. Loads take it at the end of the chain — historically a home, a factory, or an office; now, increasingly, a data center that can draw as much as a mid-sized city. Because the system connects throughout and electricity resists storage at scale, supply and demand must balance instant to instant, everywhere. When they do not, frequency and voltage drift out of their narrow safe bands, and equipment protects itself by disconnecting — the root of the reliability problems in Sections 4 and 8.

Figure P — How the grid serves a large load: generation → transmission → delivery, and who governs each step

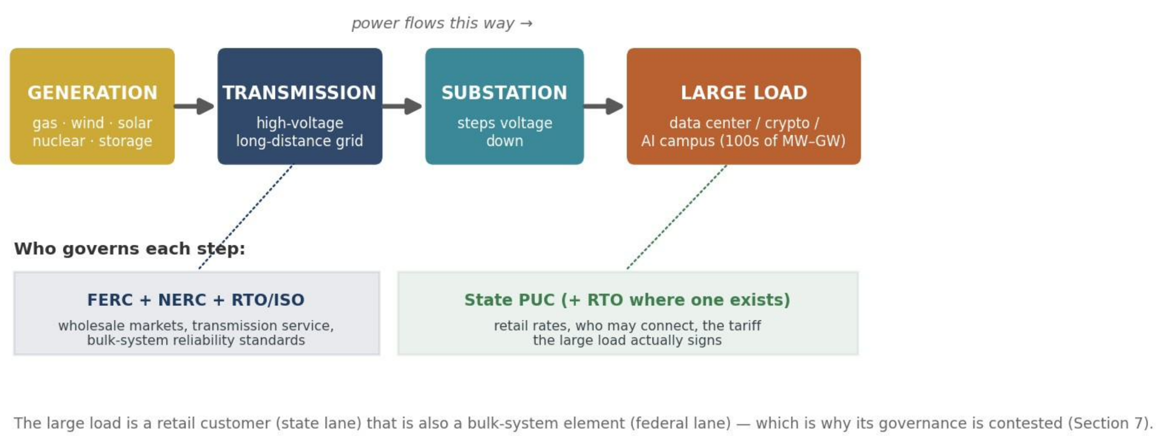


Figure P — The physical chain and its split governance. Power flows left to right through generation, transmission, distribution and load; regulatory authority divides down the middle, with federal jurisdiction over the transmission and wholesale segments and state jurisdiction over distribution and retail service. The division matters because a large load is simultaneously a retail customer and a bulk-system element, so no single authority governs it end to end. Readers should treat the dividing line, rather than any single stage, as the origin of most unresolved questions in this report (Section 7).

Three features explain almost everything that follows. First, the pieces move at very different speeds: a data center takes 18 months to three years to build, new generation about four years, and major transmission seven to ten years. Load can arrive long before the supply and wires to serve it (Section 2). Second, connecting a new large load requires an interconnection study to confirm the grid can serve it reliably, and thousands of pending requests have formed a queue (Section 1). Third, someone has to pay for any new wires and generation the load requires, and someone has to decide the terms of service — and here authority splits. Broadly, the federal government (through FERC and the reliability body NERC) and the regional grid operators (RTOs/ISOs such as PJM and MISO, with ERCOT as Texas's own operator) govern the wholesale markets, the transmission system, and bulk-system reliability; the states, through their public utility commissions, govern retail sales, local distribution, and the tariff a large customer actually signs. A data center sits squarely on that line — a retail customer that is also a bulk-system element — which is why, as Section 7 explains, no single institution has clean authority over it.

Several terms recur throughout. **RTO** or **ISO**: the entity that independently operates the grid and the wholesale market for a region. **Capacity market**: a market that pays generators years ahead to guarantee availability at peak; its prices (Section 3) carry large-load growth through to ordinary bills. **Firm service**: the grid commits to serve a customer at all times, whereas **non-firm** or **curtailable** service costs less and arrives faster but can be interrupted (Sections 6 and 8).

Ride-through: a device's ability to stay connected through a brief grid disturbance instead of tripping offline (Section 4). The Glossary at the back defines these and the many acronyms ahead.

One caution on this primer: it simplifies on purpose. The chain above runs in one direction through four stages because that is enough to follow the argument. Real networks are meshed, carry generation on the distribution system and behind customer meters, and have flows that reverse.

Simplifications that only shorten an explanation are left in place. Simplifications that could change a conclusion are flagged where they arise. Two of those are worth stating in advance.

First, a gigawatt of generating capacity is not a gigawatt of firm supply at the peak. The gap is large enough to reverse the apparent adequacy of a build-out, which is why Section 2 explains how the same generator carries three different gigawatt numbers.

Second, national totals average away regional concentration. A figure that looks modest for the United States can be severe in ERCOT or PJM. Where this report gives a national number, it establishes a trend rather than sizing the risk.

Why now: the demand curve, and what is bending it

For fifteen years U.S. electricity demand stayed essentially flat, growing about 0.1 percent a year between 2005 and 2020 as efficiency gains and the shift from manufacturing toward services offset population and economic growth. Around 2023 that changed.

Retail sales grew 0.8 percent a year on average from 2020 through 2024, and EIA forecasts about 2.2 percent in both 2025 and 2026; its 2026 Annual Energy Outlook records 2.1 percent average annual growth over the past five years — roughly twenty times the pre-2020 rate, and the first sustained increase in almost two decades. Those three figures come from three EIA products measuring three different windows — the *Electric Power Annual* for history, the *Short-Term Energy Outlook* current at July 2025 for the two forecast years, and the *Annual Energy Outlook 2026* for the trailing five-year average — so the 2.1 and the 2.2 agree rather than conflict. The inflection matters more than the level: planners built their institutions for a flat system and must now serve sustained growth of roughly 2 percent a year — on the order of 90 TWh of additional annual consumption, added each year to a system whose new generation takes about four years to build and whose new transmission takes seven to ten years.

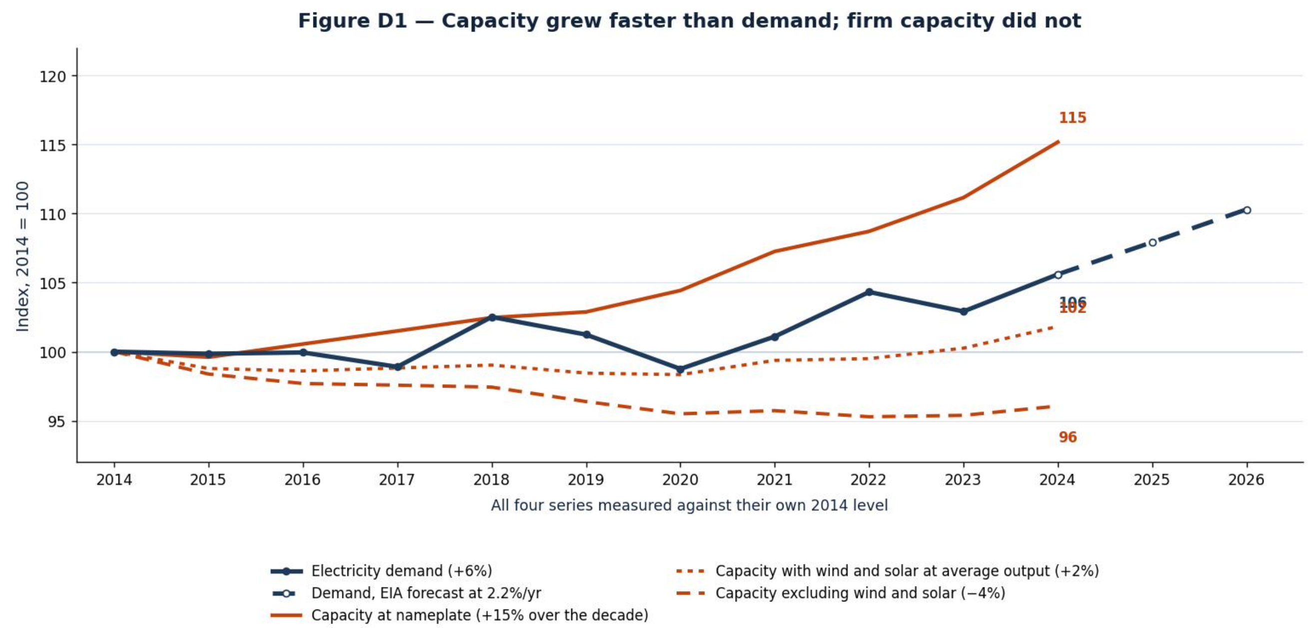


Figure D1 — Capacity grew faster than demand; firm capacity did not. All four series are indexed to their own 2014 level so they can be compared directly — an earlier version plotted terawatt-hours against gigawatts on twin axes, where the lines crossed for reasons of scaling rather than substance. Read plainly: installed capacity at nameplate grew about 15% over the decade while demand grew about 6%, so the system was not running down its margin and no shortage occurred. The concern is what the other two capacity lines show. Counting wind and solar at their average annual output, capacity grew about 2%; excluding them altogether it fell about 4%. Demand growth of 6% across a decade was comfortably served. The forecast 2.2% a year is a different proposition against a firm fleet that has stopped growing. Retail sales from EIA Electric Power Annual Table 2.5; capacity from Tables 4.2.A and 4.2.B; forecast from the Short-Term Energy Outlook.

The sales record shows the change more plainly than any growth rate does. Retail sales moved barely one percent in total over the seven years to 2021, then four percent in the three years to 2024 alone.

Year	Retail sales (TWh)	Change vs. prior year
2014	3,765	—
2015	3,759	-0.2%
2016	3,762	+0.1%
2017	3,723	-1.0%
2018	3,859	+3.6%
2019	3,811	-1.2%
2020	3,718	-2.5%
2021	3,806	+2.4%
2022	3,927	+3.2%
2023	3,874	-1.3%
2024	3,975	+2.6%
2025 (forecast)	4,063	+2.2%
2026 (forecast)	4,152	+2.2%

Table D1 — U.S. retail electricity sales to ultimate customers, 2014–2026. Actual years are EIA final annual data; 2025 and 2026 apply EIA's Short-Term Energy Outlook forecast of 2.2% annual growth to the 2024 base. Single-year swings of two to three percent are weather-driven and routine — the trend is what moved: about 0.2% a year through 2021, about 1.5% a year from 2021 to 2024, and 2.2%

forecast thereafter. From 2023, EIA improved its reporting of data center and cryptocurrency mining loads within the commercial and industrial sectors, which affects comparability across that boundary.

Source: EIA, *Electric Power Annual*, Table 2.5 (retail sales to ultimate customers). Forecast years apply the EIA *Short-Term Energy Outlook* growth rate to the 2024 base.

National energy consumption is not the appropriate planning metric for this report's subject, and the table should not be read as reassurance. EIA's 2.2% national forecast for 2025 and 2026 is composed of roughly 11% a year in ERCOT and 4% a year in PJM against near-flat demand almost everywhere else. Aggregate energy is also the wrong unit: what binds is the coincident peak in a particular place, on a particular network, and whether the wires and firm supply exist to serve it there. Table D1 establishes that the flat era ended. It does not measure the problem.

Data centers drive the headline, though not alone. Grid Strategies and others attribute the surge to four forces at once: AI and cryptocurrency data centers, a reshoring wave of industrial and manufacturing load spurred by federal industrial policy, building electrification (heat pumps and water heaters), and electric vehicles. Data centers dominate the near-term because they arrive large and fast, run around the clock, and cluster in specific localities — which is why their share of national peak understates the strain they place on the particular grids where they land.

Figure D2 — Two authoritative estimates of the same quantity, and the distance between them

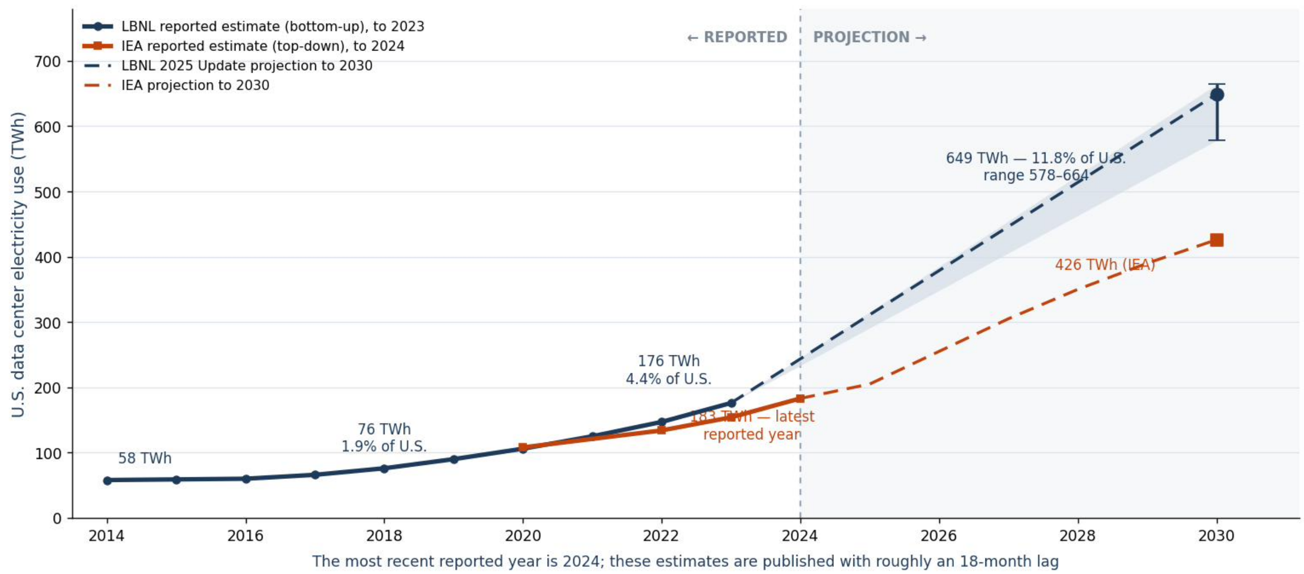


Figure D2 — Two authoritative estimates of the same quantity, and the distance between them. Lawrence Berkeley National Laboratory builds its estimate bottom-up from equipment shipments; the IEA models it top-down. On the reported record they differ by roughly 15% — LBNL puts 2023 at 176 TWh, the IEA at 154 — because they draw the boundary of a data centre differently. Both are shown rather than reconciled, since the gap is itself the finding. Everything left of the 2024 rule is reported; everything right of it is projection. That boundary is worth noting in its own right: at the date of this report the most recent year for which either body has published an estimate is 2024, because these figures are compiled with roughly an eighteen-month lag. LBNL's 2025 Update puts 2030 at 649 TWh in the reference case, 578–664 TWh across sensitivities, and 11.8% of US electricity within a 9.5–15.3% band. The IEA puts the same year at 426 TWh. ***Finding:*** *the disagreement between two authoritative forecasts is itself the planning problem. Fifty percent apart, five years out, working from a record that ends two years back — Section 1's difficulty, stated in one chart.*

The Large-Load Interconnection Process: Where the Eight Problems Arise

The eight issues treated in this report arise at identifiable points in a single sequence that runs from a customer's request to the tariff and operating limits that govern the load once it is energised. Table K1 sets that sequence out, with what each step decides, what this report finds going wrong there, and where the matter is treated. It is the shortest complete statement of the report's argument: the steps are conventional and well understood, and the difficulty is that gigawatt-scale computational load violates assumptions embedded in several of them at once.

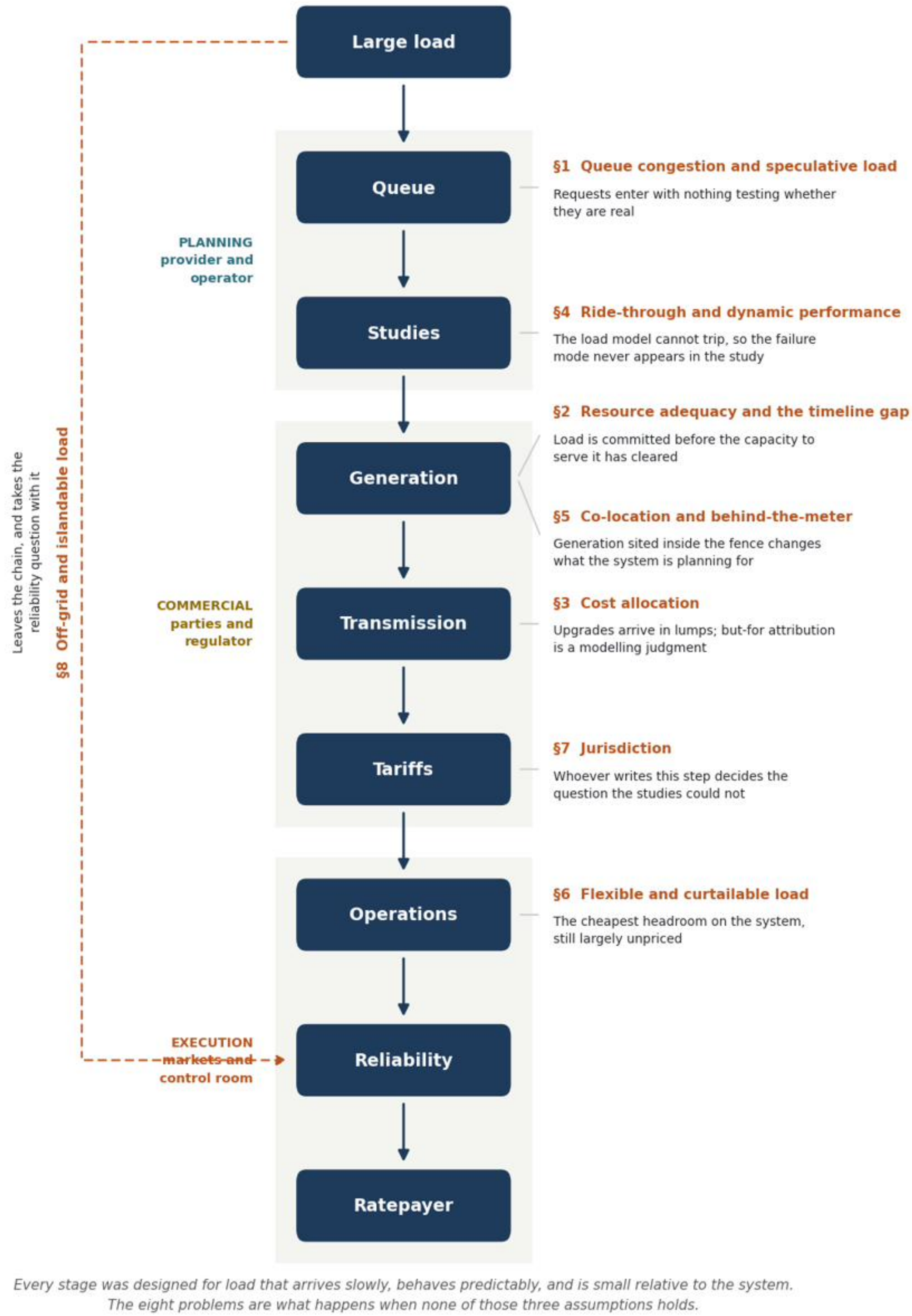


Figure K1 — The chain, and where each problem attaches. One spine runs from the customer's request to the ratepayer who ultimately pays for it, banded by the three institutional phases: planning belongs to the transmission provider and the operator, commercial to the parties and their regulator, execution to the markets and the control room. Seven of the report's eight sections attach to the stage where their problem arises. Section 8 is drawn as a dashed bypass because that is what going off-grid is — leaving the chain, and taking the reliability question along with it rather than resolving it. The stages are unremarkable in themselves; what fails is the assumption each was built on.

Step	Engineering question	Large-load problem	Sections
Planning — the transmission provider and the operator			
Load request	How much capacity, where, and by when?	Nothing tests whether the request is real; duplicate and speculative requests inflate the pipeline and crowd genuine ones	§1
Power flow	Does the network carry the load at peak and under contingency?	Volume overwhelms study capacity; a single peak case and a static load assumption answer a narrower question than the one asked	§1; T.3
Short circuit	Is fault duty at the point of interconnection within breaker ratings?	Routinely screened for generation and overlooked for load, though converter front ends and on-site generation both change the duty	T.3
Transient stability	Does the system stay in synchronism and recover voltage and frequency after a normally cleared fault?	Load represented as static and unable to trip, so the dominant failure mode — uncommanded loss of demand — cannot appear in the result	§4; T.4
EMT	What happens inside the cycle at a weak or converter-dense interface?	Treated as discretionary; vendor models used unvalidated, at settings that may not match what was commissioned	§4; T.4
Commercial — the parties and their regulator			
Upgrade list	Which network additions does this request trigger?	Upgrades arrive in lumps and physics shares them, so but-for attribution is a modelling judgment — the point at which cost-shifting begins	§3; T.3
Interconnection agreement	Who pays, on what schedule, under what performance terms?	Written before the standards that would define its performance terms exist; flexibility is left unpriced and the exit is unsecured	§3; §6
Execution — construction, markets, and the control room			
Construction	Can the wires be built in time?	Five to ten years against a two-to-three-year data-center build, with transformer and breaker lead times that no tariff shortens	§2; §5
Resource adequacy	Does generation exist to serve the load now committed?	Commitment precedes confirmation: the load is in the forecast and the agreement before the capacity to serve it has cleared	§2; T.6
Tariff and operations	What rates, curtailment obligations, and operating limits govern it?	The residual risk lands here, on ratepayers or on the load, and the choice is made by whoever writes this step rather than by the study that preceded it	§3; §6; §7

Table K1 — The sequence, and where it breaks. The three phases matter as much as the ten steps, and each fails differently: the engineering studies by modelling the load wrongly, the commercial agreements by allocating what the studies could not attribute, and execution by running out of time. The eight issues of this report are not eight separate problems so much as ten points on one process, which is why a fix applied at one step is so often undone at another.

Sources: As cited in the sections named in the final column. The sequence is standard practice; the third column states this report's findings.

Upcoming Dates and Events: A 12–18 Month Watchlist

Everything in this report eventually resolves into a calendar. This section is that calendar — the decisions, filings, and market events that will determine which of the eight problems get answered and which remain open a year and a half from now. It sits before Section 1 by design: a reader who takes nothing else from the report should still know what arrives and when.

Two cautions on reading it. First, the entries mix categories on purpose — compliance deadlines, market outcomes, rulemaking milestones, and the years NERC flags as the onset of elevated reliability risk — because developers and planners navigate that mixture as a single decision environment. Second, the confidence tag distinguishes a date someone must legally meet from a date someone currently intends to meet. Items tagged *Emerging* in particular signal direction of travel rather than schedule: rulemakings slip, standards get remanded, and a single court decision on the co-location or jurisdiction questions could reorder much of the 2027 column.

The next 90 days — July to September 2026

The densest stretch on the calendar. Four filings inside a month will show, for the first time, whether the RTOs intend to defend their tariffs or replace them.

Date	Event	What it decides	Tag
Jul 20, 2026	RTO/ISO resource-adequacy informational reports due to FERC — deadline passed	First formal statement from each RTO on whether it can actually serve its queue (§2)	SETTLED
Jul 24, 2026	ERCOT TSP/DSP submissions due — Batch Zero eligibility	Fixes the Batch Zero population and the base-load MW allocation (§1, §5)	SETTLED
Aug 3, 2026	NERC Level 3 Essential Action Alert responses due	Reveals the true extent of ride-through non-compliance across the fleet (§4)	SETTLED
Aug 3, 2026	§206 abeyance requests due (45 days)	Whether the common August 17 deadline holds across all six regions (§7)	LIKELY
Aug 17, 2026	**RTO/ISO §206 show-cause responses and briefs due (60 days)**	One of the year's most consequential filings: defend the existing tariff or propose the replacement (§5, §6, §7)	SETTLED
Sep 10 – Oct 9, 2026	PJM reliability backstop procurement window (per the June 30 design paper)	Whether backstop capacity can be procured outside the auction, and at whose cost (§2, §3)	LIKELY
Sep 15–16, 2026	NERC Data Center Load Modeling Workshop	Sets the modeling conventions the year-end standard will codify (§4, T.4)	SETTLED
Sep 16, 2026	Stakeholder comments on RTO show-cause filings due	Last broad intervention window before FERC acts (§7)	SETTLED
Oct 20 / 28, 2026	CAISO WEM Governing Body and Board of Governors votes on the large-load package	Whether the abeyance route produces a §205 proposal rather than a defence of the tariff (§7)	LIKELY

Table W1 — The next ninety days. Dates already fixed by an order, a rule, or a filing deadline. Confidence tags refer to the outcome, not the date: the calendar is settled even where what it produces is not.

Sources: FERC; NERC; ERCOT; PUCT; PJM; MISO; SPP; DOE, from the filings and dockets listed in the Catalog of Orders, Rules, and Directives Cited.

Q4 2026 through Q1 2027 — obligations begin to bind

Here voluntary guidance becomes enforceable. The December 31 date may prove the most consequential entry on the list, because reliability obligations then attach to the load itself rather than to the utilities around it.

Date	Event	What it decides	Tag
Mid-Nov 2026	CAISO filing on the show-cause order; §205 filings from any other region granted an abeyance	Whether the regions that take the pause file tariff proposals rather than defend the existing tariff (§7)	EMERGING
Q4 2026	First NERC large-load Reliability Standard (Project 2026-02 “bridge” standard)	Converts voluntary guidance into enforceable obligation (§4)	LIKELY
Dec 2026	PUCT duplicate-request disclosure rule final (16 TAC §25.194)	Determines the honest size of the ERCOT queue (§1)	LIKELY
Dec 31, 2026	**RD26-7-000 compliance:** NERC files Reliability Standards **and** Rules-of-Procedure registration criteria for computational loads	A central reliability reform lands on a date certain; the registry threshold remains the open question (§4, §8)	SETTLED
Dec 2026 – Jan 2027	NERC Long-Term Reliability Assessment (annual)	Whether the ten-year peak forecast is revised upward again (§1, §9)	LIKELY
Mar 1, 2027	**RD26-7-000:** NERC Phase II work plan for additional standards due	Sets the trajectory for standards beyond the bridge rule (§4)	SETTLED
Mar 2027	PJM reliability backstop — central pay-as-bid procurement	Price and cost allocation of backstop capacity billed to large loads (§2, §3)	LIKELY

Table W2 — Late 2026 into 2027, where guidance becomes obligation. These are the dates at which the reforms described throughout the report stop being proposals and start carrying compliance consequences.

Sources: FERC; NERC; ERCOT; PUCT; PJM; MISO; SPP; NYISO; state commissions.

Through 2027 and into 2028 — the rules mature, the risk window opens

The later entries grow increasingly uncertain because they depend on future rulemakings, and they also make the report's central timing inversion visible: the reliability-risk windows open in 2028 while the remedial standards are still being drafted and the major transmission is still years from energization.

Date	Event	What it decides	Tag
Across 2027	FERC RM26-4 matures from ANOPR to NOPR and then final rule	Whether a pro forma national large-load interconnection procedure exists at all (§1, §7)	SPECULATIVE
2027	NERC Phase II computational-load standards drafted; registry populated	Scope thresholds decide which facilities are actually bound (§4, §8)	SPECULATIVE
2027	Co-location and flexible-load service terms take effect after the rulemaking	Whether non-firm service is priced attractively enough to be chosen (§5, §6)	SPECULATIVE
Mid-2027	PJM Base Residual Auction for the 2029/30 delivery year	A third consecutive shortfall would confirm scarcity is structural, not cyclical (§2, §3)	LIKELY
Jul 2027	New York moratorium lapses; GEIS and Energize NY framework due	A framework other states may adopt, modify, or decline (§7)	LIKELY
2027–2028	ERCOT 765 kV backbone and Permian Basin plan construction milestones	The long-lead wires that do not actually arrive until the 2030s (§9)	SPECULATIVE
2028	MISO enters the NERC elevated/high reliability-risk window	Opening of the gap where risk bites before the fixes are enforceable (§9)	SPECULATIVE
Jun 1, 2028	PJM 2028/29 delivery year begins — the year procured ~6.8 GW short	First delivery year in PJM history to start below the reliability requirement (§2)	SETTLED

Table W3 — 2027 and beyond. Tagged Speculative where the entry depends on a rulemaking that has not yet been proposed, or on a market outcome three years out. The dates are indicative; the direction is not.

Sources: NERC *Long-Term Reliability Assessment*; ERCOT; PJM; MISO; FERC.

The near-term entries in this table are the operational watch list; Section 9 carries the full 2018–2035 view, including the long-lead physical builds beyond this horizon.

The Federal Frame: Six Regions, One Deadline

Every reform in the eight sections that follow is now running inside a single federal frame. On June 18, 2026, after reviewing more than 3,500 pages of comments in the large-load docket, FERC issued six simultaneous show-cause orders under section 206 of the Federal Power Act — one to each jurisdictional grid operator and its transmission owners. PJM (EL26-67), SPP (EL26-68), NYISO (EL26-69) and MISO (EL26-70) were joined by CAISO and ISO-NE, covering close to two-thirds of the load served under Commission-jurisdictional rates. The Commission preliminarily found every one of the six tariffs unjust and unreasonable for lacking clear provisions tailored to large and co-located loads.

The orders are tailored region by region but rest on the same five categories. First, clearer application and study processes, including consideration of alternative transmission technologies; protection against cost shifting, paired with new transparency about transmission costs; rules for co-location and behind-the-meter generation; new transmission services for loads that can limit their own use; and a study path for generation sited electrically close to the load it serves. Those five map onto Sections 1, 3, 5, 6 and 2 respectively — which is why this report treats them as one problem in eight parts rather than eight separate ones.

The orders run on one procedural calendar, and it has three exits rather than one. Interventions closed on July 9, 2026. Informational reports on generation adequacy were due thirty days after issuance — July 20, 2026, the nominal thirtieth day falling on a Saturday — and that deadline has now passed. Responses and briefs are due at sixty days, on August 17, with stakeholder comments thirty days after each filing, on or about September 16. The third exit makes the common deadline less common than it looks: within forty-five days, by August 3, an operator and its transmission owners may ask the Commission to hold the proceeding in abeyance for up to ninety days while a stakeholder process develops a section 205 filing instead — moving that region's substantive filing to roughly mid-November. ISO New England and the New England transmission owners announced on June 29 that they intend to make that request. CAISO went further on July 20, publishing an expedited stakeholder calendar in its informational report — straw proposal August 12, draft final proposal September 21, Western Energy Market Governing Body October 20, Board of Governors October 28 — ending in a filing on the show-cause order on November 16, a schedule the report states assumes a motion for abeyance. The Commission has said it will scrutinise such requests rather than grant them as a matter of course, will hold any abeyance to the ninety-day limit, and will look unfavourably on extensions. The practical effect is that a common deadline should be read as a starting position: the orders were simultaneous, but the responses need not be, and the first real test of how uniform this federal frame turns out to be arrives on August 3 rather than August 17.

Region	Where it stood on June 18, 2026	What FERC preliminarily found missing
PJM	Co-location provisions filed under earlier FERC directives; the Board has issued a Critical Issues Fast Path decisional letter setting out large-load initiatives.	No tariff provisions governing large-load interconnection or transmission service. Co-location is handled in a separate concurrent proceeding.
MISO	Expedited study routes (ERAS, EPR) but no dedicated large-load framework.	MISO indicated that its tariff gives no consistent or transparent basis for evaluating large loads — against the fastest data-center growth of any region, put by FERC staff at 43% compound annual growth since 2020 against 24% nationally (2025 State of the Markets, March 19, 2026).
SPP	HILL and HILLGA study processes live since January 15, 2026; Load Limited Resource Interconnection Service; CHILLS conditional service accepted June 5, 2026.	The narrowest set of deficiencies of the six, because SPP's reforms had already anticipated most of FERC's framework.
NYISO	Load interconnection procedures above 10 MW at 115 kV or higher, or 80 MW below; a reform initiative of its own running to a December 2026 board filing.	Application, study and operational requirements; cost shifting through the Transmission Service Charge when speculative requests drive local planning; no co-location terms.
CAISO	No traditional Order No. 888 transmission service, so the analytical framework differs structurally from the other five. Its July 20 report attributes California's position to an integrated CEC/CPUC/CAISO framework and sets a compliance calendar ending November 16.	Substantially similar deficiencies notwithstanding the different market design.
ISO-NE	No express provisions for large-load interconnection or transmission service.	The same gaps, but FERC recorded a less urgent concern in New England. ISO-NE and the New England transmission owners said on June 29 that they intend to seek the 90-day abeyance rather than respond on August 17.

Table 0 — The six jurisdictional regions on the day FERC acted. Responses and briefs are due August 17, 2026, except where a region obtains the 90-day abeyance the orders permit — ISO-NE has said it will ask and CAISO's published schedule assumes one; informational reports on generation adequacy were due July 20, 2026, and that deadline has now passed. What separates the regions is not whether they face the problem but how far each had travelled before the Commission set a common deadline.

Sources: FERC, six §206 show-cause orders, June 18, 2026 (EL26-67 et seq.; 195 FERC ¶ 61,211–61,216), and the operators' own tariff filings.

Two features of that table matter more than the individual entries. Acting early produced a measurable regulatory benefit. SPP drew the narrowest findings because it had already built the machinery — the clearest evidence in the record that anticipatory tariff work changes a region's regulatory exposure. The second is that the severity of FERC's findings tracks tariff readiness rather than load growth. ISO-NE has almost no large-load provisions and drew a mild order; MISO, whose data-center growth FERC staff measure as the fastest in the country, drew a pointed one. Regions where the load has not yet arrived are untested rather than prepared.

1. Queue Congestion and Speculative "Phantom" Load

The issue. The large-load interconnection queue has become an increasingly uncertain planning instrument. Unlike generator interconnection, load interconnection historically carried limited standardized transparency, no consistent site-control test, and little financial commitment. In practice, developers shop the same project across multiple utilities, multiple points of interconnection, and sometimes multiple states, and take whichever offer clears first. The queue therefore cannot be read as a direct measure of demand; it reflects a mixture of project optionality and firm intent. ERCOT presents the extreme case. The queue went from roughly 63 GW in late 2024 to 226 GW by November 2025, and to more than 438 GW by mid-2026 — a figure ERCOT itself acknowledges is inflated. Its preliminary long-term load forecast implied a 2032 peak of 367,790 MW, more than four times the historical record, while ERCOT simultaneously told the PUCT that summer 2026 peak would land somewhere between 90,500 and 98,000 MW. Differences of this magnitude limit the forecast's usefulness for long-term planning and suggest that speculative queue entries materially influence the result. The resulting effects include transmission sized against load that may never appear (over-build, stranded cost, ratepayer exposure), capacity markets clearing against demand that has not materialized (Section 3), and genuine projects waiting behind speculative ones.

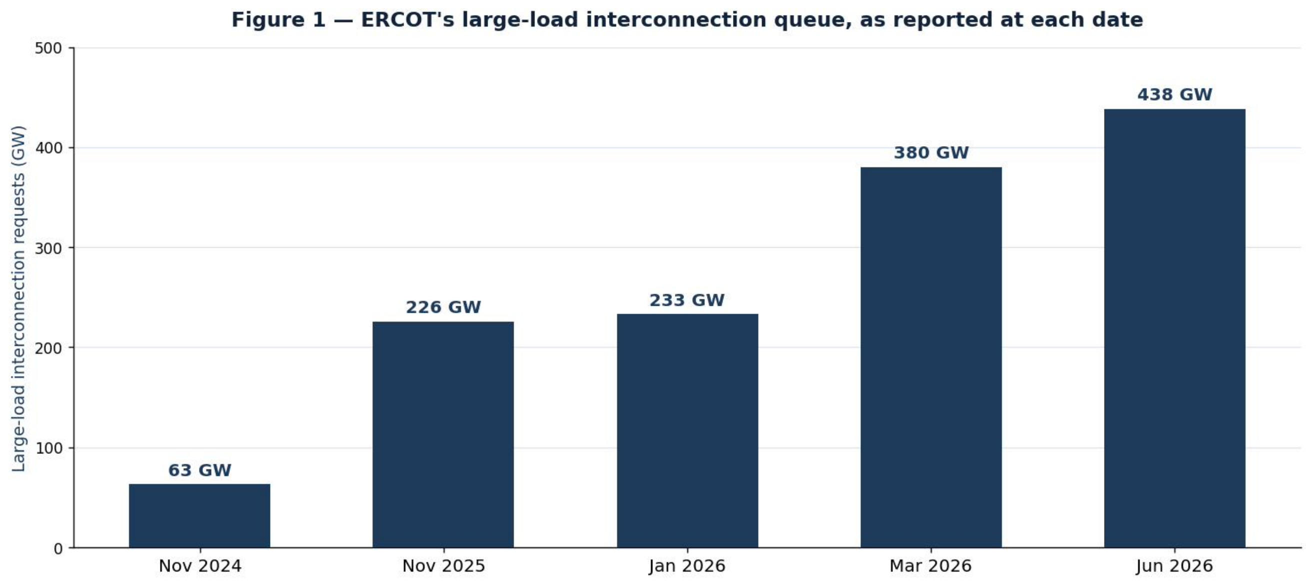


Figure 1 — ERCOT's large-load interconnection queue, as reported at each date. The queue grew roughly sevenfold in nineteen months. Set against it, ERCOT approved on the order of 2.2 GW to energize over the trailing twelve months to late 2025, and about 9 GW had been approved in total by the April 2026 Large Load Working Group — against an observed simultaneous peak near 3.7 GW in March 2026. The gap between requests and energizations is what defines the speculation problem, and ERCOT states that the queue figure carries duplicate and speculative requests, so it indicates pressure rather than demand.

How load forecasting actually works — and why the same queue produces different numbers

Because so much of this section turns on the gap between a request and a real project, the mechanics deserve specifics. No single method exists. Three families compete, and the choice among them explains why ERCOT and PJM can read similar data and publish forecasts differing by a factor of three. The traditional workhorse is the econometric (top-down) forecast: regress historical system load on weather, calendar, population, and economic variables, then project forward. It holds up for a slowly changing system and fails for a step-change: a 1 GW campus with no history in the data never enters the model. ERCOT and PJM both run econometric base models, PJM using Moody's Analytics economic inputs refreshed in its September 2025 release, and both have had to add something to them.

The second family supplies that addition. The bottom-up, or project-based, adjustment enumerates the actual named projects and weights each by its probability of materializing. The real judgment lives here, and the weighting scheme determines the result. PJM's Load Adjustment Request framework (published July 2025) does exactly this — it solicits largeload additions from each distribution company and reviews every one for significance, likelihood, and doublecounting risk. Utilities inside PJM apply explicit confidence tiers: PSE&G;'s 2026 filing, for example, counted expansions of existing data centers and new-service requests at 100% but feasibility-stage requests at 50%. The signed interconnection agreement gets full weight; the speculative inquiry gets zero. Texas encodes the same logic in rule rather than practice — after 2026, PUCT Project 58480 admits only large loads with a qualifying executed agreement into the ERCOT forecast.

The third family, still emerging, is the probabilistic or scenario forecast. Rather than a single line, it produces a distribution across many weather, economic, and project-realization combinations. NERC's May 2026 Reliability Guideline explicitly pushes planners toward this — resource-adequacy models that represent firm versus flexible load separately, account for behind-the-meter resources, and reflect AI-training operating windows, evaluated across probabilistic scenarios on a network-aware footprint. It answers a genuinely uncertain queue honestly, and defies explanation to a legislature that wants one number.

Figure 10 — Load-forecasting methodology: probability-weighting requests, and what it did to PJM's 2026 numbers

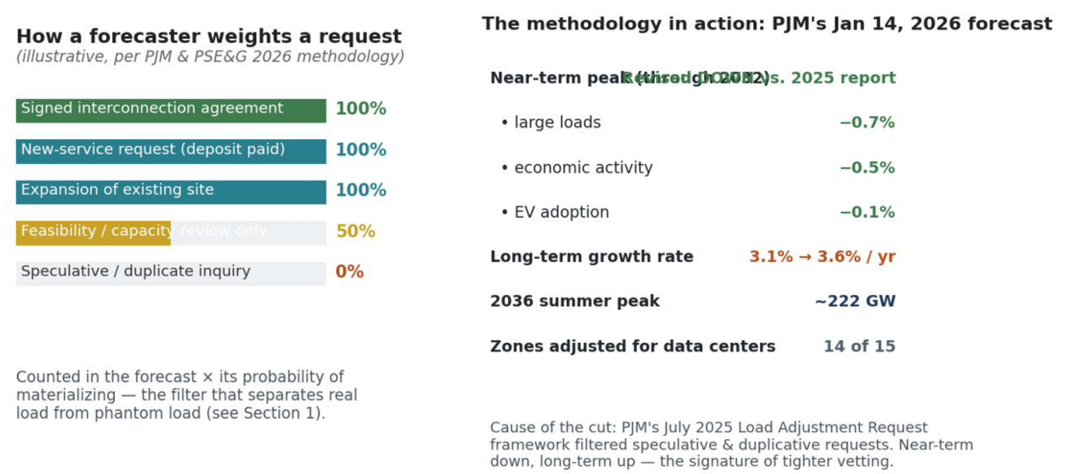


Figure 10 — The probability-weighting step, which separates a forecast from a queue tally. Each request is assigned a likelihood of proceeding rather than counted at full value. The step matters because the same queue can yield forecasts differing by a factor of three depending on how it is weighted. PJM's January 14, 2026 forecast shows the method in operation: tighter vetting pulled the near-term number down even as the long-term growth rate rose — which is the signature of a forecast improving rather than deteriorating.

The payoff is visible in PJM's own numbers. Its 2026 Long-Term Load Forecast, issued January 14, 2026, revised the near-term peak down versus the 2025 report — large loads –0.7%, economic activity –0.5%, EVs –0.1% for summer 2026 — precisely because the new vetting framework filtered speculative and duplicative requests out of 14 of 15 adjusted zones. At the same time it raised the long-term growth rate from 3.1% to 3.6% per year, to roughly 222 GW by 2036. A near-term revision downward paired with stronger long-term growth marks a forecaster who has stopped counting the queue and started weighting it. NERC's January 2026 Long-Term Reliability Assessment moved the opposite direction at the ten-year horizon, raising projected summer-peak growth by 224 GW — about a 69% jump over the prior year — which is what happens when the long tail of data-center load is finally taken seriously. For anyone reading a forecast, the number matters less than the method: what probability did the forecaster assign to feasibility-stage load, and did anyone check it against the neighboring utility's queue for the same project? Two forecasters with the same raw requests and different answers to that question will diverge more widely than differing weather assumptions would produce. The larger pattern behind all of this is that the forecasts themselves keep moving — and almost always upward. Grid Strategies, aggregating utilities' FERC Form 714 filings, found the nationwide five-year peak-growth forecast rise from about 23 GW (2022) to 39 GW (2023) to 67 GW (2024) on a consistent basis. (Its widely-quoted “128 GW” headline for 2024 is that 67 GW plus roughly 61 GW of further planning-document updates — a distinct, augmented metric, and a caution against comparing numbers that measure different things.) By the 2025 report the aggregate had reached about 166 GW projected by 2030, roughly a six-fold increase over the flat 2022 baseline. That is the macro version of the same story: as data-center load moves from speculative to contracted, it migrates into the official numbers, and each year's forecast rises to catch what the last one missed.

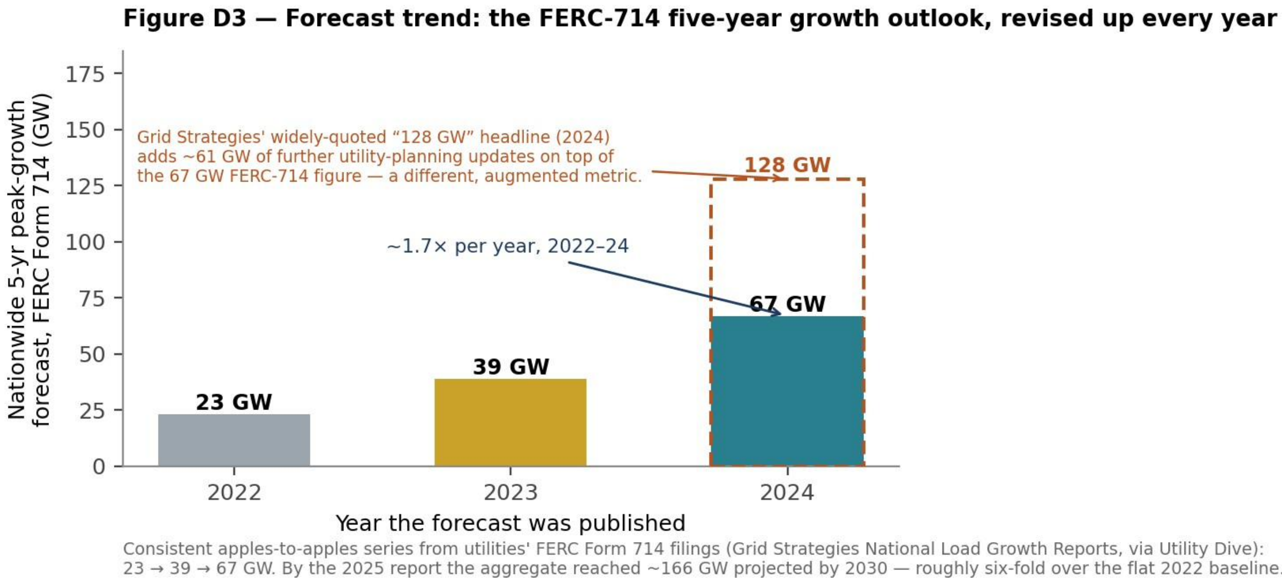


Figure D3 — The forecast trend itself tells the story: utilities' aggregate five-year peak-growth forecast (FERC Form 714 basis) has risen every year as data-center load hardens from inquiry into signed load — 23 → 39 → 67 GW. The dashed outline flags the widely-quoted 128 GW headline as a separate, augmented metric, not a continuation of this series. The risk cuts both ways: the mechanism that under-counted load in 2022 can over-count it now if the vetting (Figure 10) proves weak.

Proposed solutions

- ERCOT Batch Zero (PGRR145 / NPRR1325), effective July 11, 2026. Replaces the serial, project-by-project Large Load Interconnection Study with a system-wide batch study of all requests ≥75 MW, assessed against a common set of transmission limits. The study sorts loads into base load (studied, agreement executed), studied loads, and not-included. Base-load MW allocation is capped by a 'lesser-of' rule tied to transmission reliability limits, which forces the allocation to reflect physical headroom rather than request volume.
- Financial and site-control gating. PUCT Project 58481 (proposed 16 TAC §25.194, implementing SB6 / PURA §37.0561) sets a 75 MW threshold and requires an Intermediate Agreement supported by documented site control, backup generation disclosure, affiliate/end-use transparency, and financial security proposed at $50,000/MW — roughly 20% refundable at commercial operation. A 500 MW campus therefore posts $25 million to hold its place.
- Forecast gating (PUCT Project 58480, 16 TAC §25.370, effective March 1, 2026). After 2026, only large loads with a qualifying executed interconnection agreement enter the load forecast. This severs the link between speculative requests and planning cases.
- Duplicate-request disclosure. Texas law now directs the PUCT (by December 2026) to require large-load customers to disclose whether they have multiple similar requests under review in the state — among the most direct attacks on double-counting.
- FERC's Category 1 reform (June 18, 2026). Every RTO must justify or replace its application and study process for large-load transmission service, explicitly including readiness screens and alternative transmission technologies.

ASSESSMENT **analytical interpretation, not a cited fact** **confidence: LIKELY**

Experience with generator interconnection suggests financial gating can improve queue discipline, although its effectiveness for large loads remains to be demonstrated, and the mechanism lacks selectivity. In practice, a $50k/MW deposit costs a hyperscaler little and a mid-market colocation developer a great deal, so the screen selects for balance-sheet depth as much as for project credibility. Duplicate-request disclosure addresses the problem more directly and travels better: other states can adopt it without importing Texas's capital requirements. ERCOT's own commentary indicates that the published queue includes significant uncertainty, and the disclosed figure could settle at a substantially lower level.

2. Resource Adequacy and the Development-Timeline Gap

The issue. Even if every queued megawatt were real, the grid could not serve it on the schedule requested, because the three things that must arrive together do not move at the same speed. MISO illustrated the timing mismatch in its January 2026 Large Load Additions workshop: a data center goes from site selection to commercial operation in 18 months to three years; new generation takes about four years; major transmission takes seven to ten years.

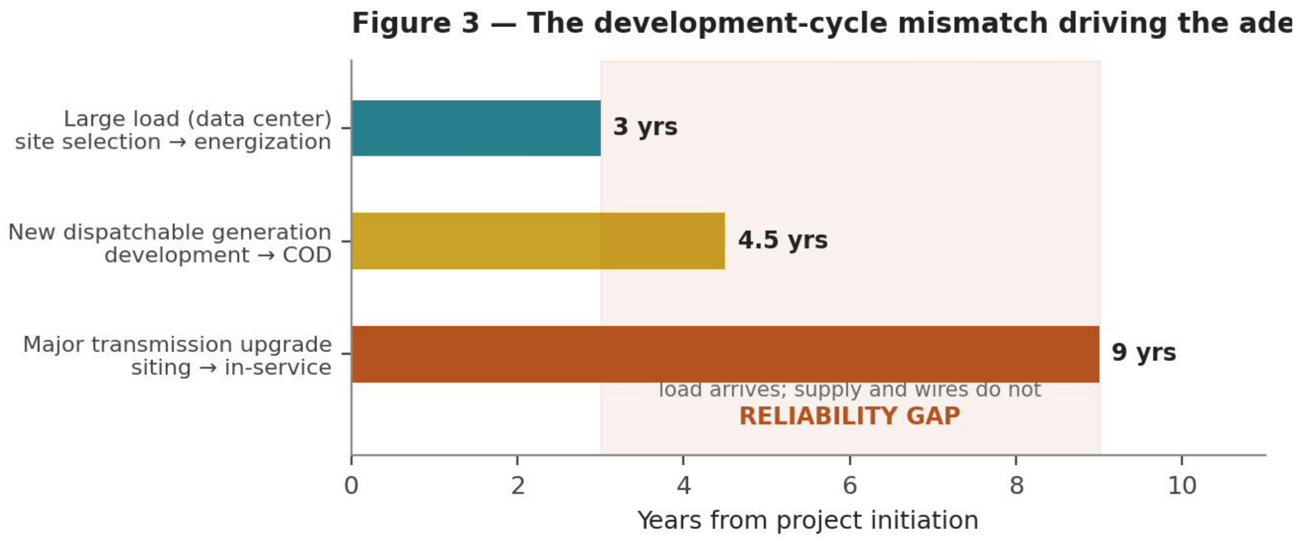


Figure 3 — Development timelines compared: large-load construction against the generation and transmission required to serve it. Load reaches commercial operation roughly three times faster than the supporting infrastructure, and the difference is structural rather than administrative, since it reflects equipment lead times and permitting rather than study queues. The shaded region is where reliability risk accumulates — the interval in which demand is connected and the resources planned to serve it are not yet in service.

Market outcomes are already consistent with this rather than leaving it a theoretical concern. PJM's December 2025 capacity auction for the 2027/28 delivery year cleared at the price cap and still fell 6,625 MW short of the reliability requirement — the first shortfall in the market's history. The 2028/29 auction of July 14, 2026 repeated it at 6,831 MW, leaving a reserve margin of 14.7% against a 20% target. PJM has said it will seek Commission approval for a special backstop procurement in September. The stakeholder design paper sets a procurement window of September 10 to October 9, 2026, and allocates the cost to the load additions that created the need. **The 2028/29 auction, released July 14, 2026, repeated it and widened the gap:** 138,318 MW UCAP procured, clearing again at the cap ($325/MW-day, about 2.5% below the prior cap), roughly 6.8 GW short of the reliability requirement, with only about 525 MW of new generation clearing. Two consecutive cap-clearing shortfalls with negligible new supply reflect the timeline gap appearing in market prices, and PJM's tariff treats a persistent shortfall of more than one percentage point as a trigger for a Reliability Backstop Auction — which moves the backstop below from a discretionary plan toward a mandatory pathway. PJM's load forecast calls for roughly 30 GW of new demand by 2030. In ERCOT, about 23 GW of generation was added across 2024–2025 with another 9 GW expected in early 2026 — real progress, and still an order of magnitude below the queue. Underneath the four-year generation figure sits a harder constraint: the equipment itself. On the order-book evidence below, gas-turbine manufacturing rather than developer demand now appears to be the binding limit. GE Vernova's combined gas backlog and slotreservation agreements reached 100 GW in the first quarter of 2026 (56 GW of it reservations), with near-term capacity effectively sold out through 2027 and delivery conversations now centered on 2029–2030; Siemens and Mitsubishi report comparable multi-year lead times, and large-power-transformer and high-voltage breaker backlogs run in parallel. The “four years” therefore reads optimistically, and the same supply constraints also affect the on-site alternatives of Sections 5 and 8 in a second way: a data center's behind-the-meter turbine order competes for the same slot, at the same handful of suppliers, as a utility's grid-serving order. On-site generation does not step outside the bottleneck — it bids against the grid inside it, and may widen the adequacy gap it was intended to address. (The market-wide view of this dynamic, and its mirror in the queue, is developed at Figure 15 in Section 9.)

Three ways to count a gigawatt

Capacity figures are quoted in gigawatts, and the same generator has three different gigawatt numbers attached to it. Confusing them is the most common error in reading this data, and it runs in both directions. **Nameplate capacity** is the maximum output the machine can produce under design conditions — the number in the press release and in EIA's addition totals. **Accredited capacity** is how much of that counts toward meeting the system's peak obligation, set by the operator through an effective load carrying capability analysis: PJM multiplies nameplate by a class rating reflecting how much the resource actually reduces shortfall risk in the hours when shortfall is likely. **Delivered energy** is the average output across a year, nameplate multiplied by capacity factor — what the resource contributes to total consumption rather than to peak.

Resource (per 100 MW nameplate)	Counts toward peak (MW)	Average output over a year (MW)	Why the two differ
Tracking solar	65	24	Output peaks at midday; PJM's shortfall risk concentrates in late-afternoon and winter hours.
Fixed-tilt solar	50	24	Same annual energy as tracking, delivered in a narrower daily window.
Onshore wind	30	34	Delivers more energy than it is credited for — output is weakest in the summer hours that set the peak.
Battery storage, 4–10 hour	100	net consumer of energy	Full credit for as long as its duration lasts; it moves energy in time rather than producing it.
Gas combined cycle	≈90–95	58	Dispatchable, so credit is near nameplate less forced-outage history; it runs below capacity for economic reasons, not physical ones.
Nuclear	≈95	91	Runs close to continuously; all three numbers nearly converge.

Table 2A — Three ways to count a gigawatt. Every figure here states a central tendency rather than a constant. Accreditation varies by region and by year: PJM's ratings are recalculated annually and ERCOT accredits on a different basis entirely. Average output varies at least as widely — with resource quality and climate at the site, with equipment type and vintage, with how hard a unit is dispatched, and with age, since output degrades over a plant's life and thermal efficiency falls as equipment wears. A modern combined-cycle unit in a merchant fleet and a forty-year-old peaker share a nameplate rating and almost nothing else. Read the relationships, not the numbers.

Source: Accreditation: PJM ELCC class ratings, 2026/2027 delivery year. Average output: EIA 2025 national capacity factors.

None of the three is the real number; they answer different questions, and the ordering between them is not fixed. Wind delivers more energy than it is credited for, while solar is credited for more than it delivers. A battery is credited at full nameplate but is a net consumer of energy. The 50.5 GW of capacity added across the United States in 2024 therefore does not net against demand growth in the way a single figure suggests: about 31 GW of it was solar, 11 GW batteries, 5 GW wind, and 2 GW natural gas. Against a data center drawing near its maximum around the clock, the question turns not on how many nameplate gigawatts were built but on how many firm gigawatts stand available in the hour the load runs — the quantity Figure D1 tracks, which has not grown.

For scale, U.S. summer peak demand reached about 760 GW in July 2024. Set against roughly 1,230 GW of installed capacity that looks like ample headroom. Set against the 955 GW that excludes wind and solar, and reduced again for accreditation and forced outages, the margin tightens considerably — which explains how PJM's capacity auctions clear at their cap three years running while national capacity totals rise.

The wires: extra-high-voltage build-out, and the capacity already in the ground

Transmission is the slowest of the four clocks, and the response to large-load growth has been the first extra-high-voltage build-out in decades. ERCOT is the clearest case. The Permian Basin Reliability Plan, approved by the PUCT in April 2025, authorised three 765 kV lines — the first in ERCOT history — covering roughly 1,255 miles, estimated at $10 to $14 billion depending on scope. The ERCOT board approved the remainder of the Strategic Transmission Expansion Plan in December 2025, and across the full programme the 765 kV element is estimated at $17.2 billion, with a further $4.7 billion of local Permian upgrades and about $11 billion of accompanying work elsewhere — a capital total near $33 billion. The two figures are not alternatives: the smaller covers the three Permian lines alone, the larger the whole extra-high-voltage build. MISO is doing the same thing on a comparable scale: Tranche 2.1, a $21.8 billion portfolio built around a 765 kV backbone across its multi-state footprint, with in-service dates in 2032–34.

The dates are the constraint. ERCOT expects its 765 kV lines in service around 2031 and MISO its own in the mid-2030s, while the demand they answer arrives now — Permian load alone is projected at roughly 23.7 GW in 2030 against about 11 GW in 2026. Analysis of the ERCOT plan finds the new lines do not materially change zonal imports and exports in the near term, because the west zone depends more on generation built locally than on power imported over the new backbone. This is not an argument against the build-out but the timing inversion expressed in physical assets: the wires answer the 2030s and do not answer 2028.

Which is why the more immediate question concerns the capacity already in the ground, and that quantity is now moving in both directions at once. It moves up through grid-enhancing technologies. Static line ratings were set on conservative worst-case weather assumptions, so a line rated for a still 40°C afternoon is carrying far less than it safely could on a windy winter night. FERC Order No. 881 required transmission providers to move to ambient-adjusted ratings by July 12, 2025, and Order No. 1920 requires providers to consider dynamic line ratings, advanced power flow control, advanced conductors and transmission switching in regional planning rather than defaulting to new lines. Full dynamic line rating goes further than ambient adjustment, combining sensors on the conductor with weather modelling and forecasting to set a rating hour by hour; demonstrations have reported capacity gains in the range of 25% and vendors claim up to 40% on favourable lines. Topology optimisation and power flow control add a second lever, rerouting flow around a constraint rather than raising the limit on it. None of this requires new right-of-way, and all of it can be deployed in months rather than years — which makes it, alongside flexible load, one of the few responses that operates on the load's own timescale.

Machine learning is beginning to appear in the same layer, and the symmetry is worth naming: the technology driving the load is also being applied to the network that must carry it. The applications are unglamorous and mostly concern speed. Contingency screening is a combinatorial problem that planners today address by studying a credible subset; learned surrogates can rank a far larger set quickly enough to shortlist cases for full simulation, which widens coverage rather than replacing the power-flow solution. Topology optimisation — finding the switching action that relieves a constraint — is a search problem of the kind reinforcement learning suits, and several operators have run trials. Dynamic line ratings depend on forecasting conductor temperature from weather, which is a forecasting problem before it is a grid problem. And short-horizon congestion and net-load forecasting feed both the market and the operator directly. Two cautions apply throughout. A learned model that cannot show why it reached an answer is difficult to place inside a planning process built on reproducible, auditable studies, and none of this changes the physics of what the conductor will carry: it finds headroom faster, and finds more of it, but it builds nothing. Set against a queue measured in gigawatts and lines that arrive in the 2030s, this belongs in the same category as the rest of this paragraph — real, deployable on the load's own timescale, and not a substitute for the wires.

Capability also moves downward, a direction the debate largely omits. ERCOT has reduced System Operating Limits on some interfaces because the sudden loss of a cluster of large loads could by itself cause a violation (Section 4). A System Operating Limit caps the power that may flow across an interface before a contingency occurs, so lowering it removes usable transfer capability from every generator and every consumer that depends on that path — not only from the facilities whose ride-through behaviour created the concern. The effect is the opposite of a grid-enhancing technology and arrives through the same variable. Software and sensors are adding transfer capability; unmanaged load behaviour is consuming it. A ride-through standard is therefore also a transmission-capacity measure, and Sections 2 and 4 are closer to the same subject than their separation in this report implies.

A third case sits between the two directions, and the Crane restart illustrates it. Constellation is returning the former Three Mile Island Unit 1 to service as 835 MW of carbon-free capacity, on a schedule far shorter than any greenfield equivalent in Figure 3, because the plant already stands. The binding constraint proved not to be construction. PJM determined that the 765 kV and 500 kV upgrades needed to deliver the unit's full output would not finish before December 2030, three years past the 2027 target, and a nuclear unit held below rated output for extended periods carries vibration and wear risks of its own. Constellation resolved the gap by moving 760 MW of Capacity Interconnection Rights from Eddystone, which Department of Energy emergency orders had left running as an energy-only resource, and FERC granted the waiver on June 1, 2026 (Section 5). The lesson generalises beyond nuclear restarts. Where existing capacity can return faster than the wires that would carry it, the binding step becomes an allocation question rather than a construction one, and it resolves on a regulatory clock rather than a supply-chain clock.

Proposed solutions

- Expedited generation study tracks. MISO's Expedited Resource Addition Study (ERAS) issues a Generator Interconnection Agreement in roughly 90 days for reliability-critical resources, on a quarterly serial basis with a project cap (68 projects, with carve-outs). As of the April 2026 IPWG, the ERAS queue held 53 requests totaling about 27 GW, predominantly LSE-sponsored. SPP runs a parallel process; PJM is developing an Expedited Interconnection Track.
- Expedited transmission review. MISO's MTEP Expedited Project Review (EPR) is the wires-side analogue, intended to move large-load-driven upgrades ahead of the normal MTEP cadence.
- Reuse of an existing interconnection position. Where a plant retires, derates, or runs below the service it holds, the interconnection rights attached to it can serve something else without building anything. Order No. 845 established surplus interconnection service for this purpose in 2018: a new customer may take the unused portion of an existing customer's interconnection service, so that total service at the point of interconnection stays the same, and only up to the level that needs no new network upgrades. The Eddystone-to-Crane waiver of June 1, 2026 applied the same logic across two stations, advancing full deliverability by roughly three years (Section 5). Two limits apply. The instrument reallocates deliverability rather than creating it, which is why PJM's market monitor protested the Crane waiver, and surplus service depends on the original interconnection agreement, so it ends when that agreement does.
- Reliability backstop procurement (PJM). On June 30, 2026, PJM stakeholders approved a two-part backstop plan: LSEs (and potentially data centers directly) request a one-time capacity procurement to cover the projected shortfall, with the procurement's average cost capped at $555/MW-day and billed to large loads. PJM has separately proposed adding roughly 14.9 GW through a mix of bilateral contracts and central procurement.
- Bring Your Own Generation (BYOG). Both PJM (per its January 2026 Board Decisional Letter) and ERCOT are building expedited tracks conditioned on the load bringing incremental generation with it — converting the load from a net drain on adequacy into a roughly neutral or additive one.
- Mandatory resource-adequacy accounting. FERC's June 18 orders required every RTO to file, by July 20, 2026, an informational report explaining how it will ensure adequate generation is available to serve existing and new large loads. That deadline has now passed. Each report was to set out the proposals under consideration in the region's stakeholder process, a schedule of milestones including the expected filing date, and any work under way to accelerate generation additions — which makes the six documents, taken together, the first region-by-region statement on the record of whether the queue can actually be served. MISO's filing, reported the same day, matches its position in the record: a new expedited study process for large loads conducted outside the regular interconnection queue, as the ERAS fast lane winds down, together with a non-firm transmission service option offered in response to the order, against a published target of approval within 120 days once studies are complete and interconnection agreements are signed. CAISO's is the counterpoint, though from a region carrying a fraction of the pressure — its own forecast is 1.8 GW of data-center growth by 2030. It reports that current planning

assessments identify no systemic generation-adequacy shortfall comparable to other regions, and attributes that to an integrated framework in which the CEC forecasts, the CPUC orders procurement, and the operator plans the transmission; it does not argue that no further work is required, and commits the additional work to its Large Loads stakeholder initiative. The PJM, SPP, NYISO and ISO-NE filings were not retrievable when this revision closed, and PJM's matters most.

ASSESSMENT **analytical interpretation, not a cited fact** **confidence: LIKELY**

Expedited tracks relieve the study bottleneck, not the physical one. Turbines, transformers, and high-voltage breakers carry multi-year lead times that no tariff shortens, and ERAS-style queue prioritization primarily reallocates available capacity rather than increasing physical supply. The backstop procurement reflects recognition that additional capacity may need to be procured at premium cost, and it charges that premium to the loads that created the need. Current evidence points to demand-side flexibility (Section 6) as among the most scalable near-term responses, and one of the few that operates on the load's own timescale. Two further points follow from the material above. First, capacity accounting is itself a variable: the same fleet counted at nameplate grew about 15% over the past decade, counted at average output about 2%, and excluding wind and solar not at all, so an adequacy argument that does not state which measure it uses cannot be evaluated. Second, the extra-high-voltage programs are correctly scoped and wrongly timed for this problem — they answer the 2030s — which places more weight on the measures that raise usable capacity within the existing network, whether by ambient-adjusted and dynamic ratings or by managing the load behaviour now reducing operating limits.

3. Cost Allocation and Cost-Shifting to Existing Ratepayers

The issue. Under traditional cost allocation, network upgrades and capacity costs are socialized across the load base, because historically load growth was incremental, broadly distributed, and roughly proportional to the customer base that paid for it. A 1 GW campus challenges many of those assumptions. It arrives lumpy and geographically concentrated, driven by customers whose investment economics differ substantially from those of residential consumers — and under traditional socialized cost allocation, residential customers may bear a portion of the resulting costs.

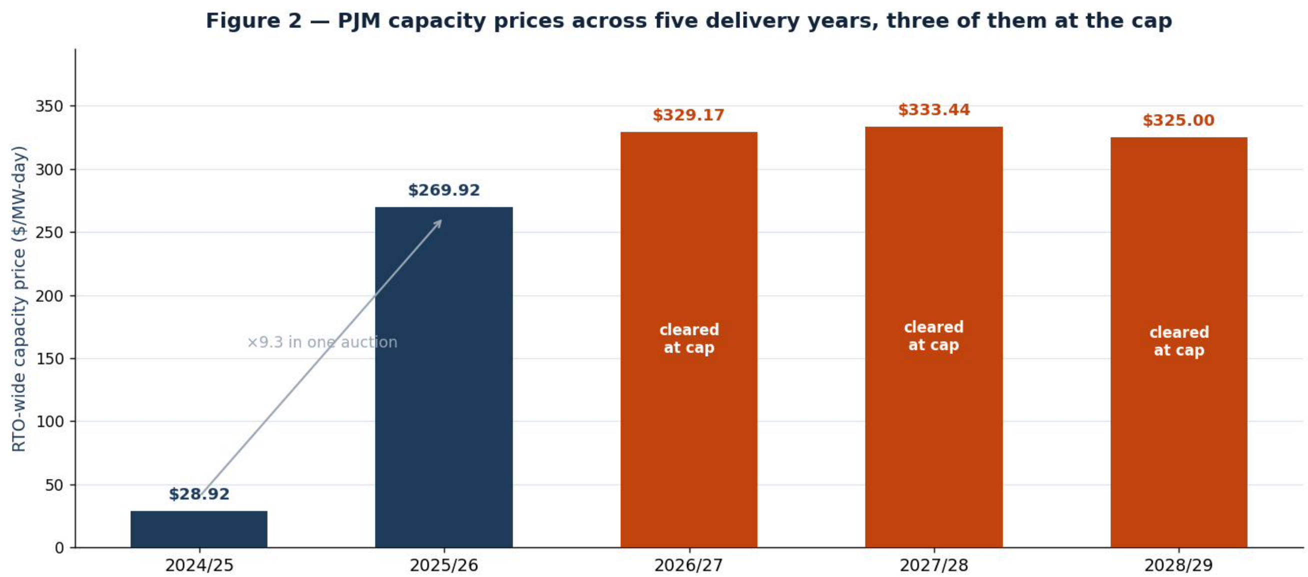


Figure 2 — PJM capacity prices across five delivery years. RTO-wide Base Residual Auction clearing prices. The 2025/26 auction, held in July 2024, raised the RTO price from $28.92 to $269.92/MW-day in a single round. The three auctions since — 2026/27 in July 2025, 2027/28 in December 2025, and 2028/29 in July 2026 — each cleared at the FERC-approved cap in force at the time: $329.17, $333.44 and $325/MW-day. Two of the three fell in the same calendar year because PJM was returning to a three-year-forward schedule. The escalation carries large-load growth through to household bills with essentially no buffer between auction and rate case, and the small decline at the end reflects the extended price collar rather than any supply response.

PJM's Independent Market Monitor found that data-center load accounted for $6.5 billion — about 40% — of the $16.4 billion cost of the December 2025 capacity auction, and that roughly $6.2 billion of that related to data centers that have not been built but could come online by the 2027/28 delivery year. Across the last three base auctions, data-center-related costs above existing data-center load totaled $21.3 billion, or about 45% of $47.2 billion. Forward estimates of the household impact turn on a single assumption — whether the capacity price collar holds. NRDC's September 2025 analysis, the source of the widely repeated $70/month figure, assumed the collar lapsed after 2027/28 and annual capacity costs rose to $27–30 billion, totaling $163 billion from 2028 through 2033; Synapse Energy Economics put the cumulative figure near $100 billion assuming mitigation continued, which implies roughly $40–45/month on the same arithmetic. That premise has weakened since: FERC accepted an extension of the $325/$175 per MW-day collar to the 2028/29 and 2029/30 delivery years on April 28, 2026, and the 2028/29 auction cleared at $16.4 billion rather than the $27–30 billion the higher projection assumed. Measured capacity-attributable increases are smaller and better documented — roughly $16/month in Ohio, $18/month in western Maryland, and about $10 of a $21/month Pepco increase in the District. The $70 figure is best read as the upper bound of a $15–70 range contingent on collar policy after 2029/30, rather than as a central expectation. The 2028/29 auction (July 14, 2026) cleared at the $325/MW-day cap for a third consecutive capped auction, so the escalation in Figure 2 has flattened against the ceiling rather than reversing — the small nominal dip reflects the price collar rather than a supply response. This has become one of the principal areas of policy debate in large-load system expansion, and it is likely to require near-term regulatory attention.

The argument for direct assignment reached the Commission in concrete form well before any of the figures above. Protesting the Susquehanna amended interconnection service agreement in June 2024, Exelon and AEP put the cost shift from that single arrangement at up to $140 million a year, and the exchange that followed shows why the principle proves harder to administer than to state. Their affiants derived the figure by costing 480 MW at a 98% load factor under PPL's LP-5 retail tariff — that is, by asking what the utility would have collected had the load connected to the distribution system in the ordinary way. Susquehanna answered that this measured forgone revenue rather than any cost actually incurred: PPL had built nothing for the arrangement, so no expense sat in rate base awaiting reassignment, and at most the utility lost an opportunity. Both accounts can hold at once, which is the difficulty. A cost shift measured against what a customer would have paid under an alternative tariff answers a different question from a cost shift measured against dollars the system has spent, and the two diverge most sharply in exactly the cases — co-location, behind-the-meter supply, off-grid campuses — where the load avoids the network it would otherwise have used.

Proposed solutions

- "But-for" direct assignment. The emerging default: charge the large load for upgrades that would not have been needed but for its interconnection, irrespective of whether other customers incidentally benefit. DOE's §403 principles went further and proposed assigning 100% of such costs to the load.
- FERC Category 2 (June 18, 2026). Each RTO must demonstrate or propose enhanced network-upgrade cost transparency, pro forma cost-recovery agreements, and explicit mechanisms to prevent cost-shifting among transmission customers.
- Large-load rate classes and minimum-take contracts. State-level tariffs (Ohio, Virginia, Georgia and others) establishing separate classes with long terms (10–15 years), minimum demand charges, and exit fees — so that a load which departs early does not strand infrastructure built for it.
- Load-billed backstop procurement. PJM's June 30, 2026 plan bills the capped ($555/MW-day) procurement to large loads through LSEs rather than the general load base.
- Texas SB6 / PUCT gating. Financial security at $50k/MW plus the transparency rules ensure that transmission planned for a large load is backed by that load's own capital at risk.
- Governor-level coordination. The thirteen PJM-state governors and the Energy Dominance Council jointly committed to use their authorities to allocate costs to data centers and protect residential customers — a signal that if FERC and the RTOs do not solve this, states will.
- Voluntary commitments by the customers themselves. The seven signatories to the March 4, 2026 Ratepayer Protection Pledge — Amazon, Google, Meta, Microsoft, OpenAI, Oracle and xAI — undertook to pay for the new delivery infrastructure their facilities require and to pay for contracted capacity whether or not they consume it; Google implements this through a Capacity Commitment Framework that bills against power requested rather than power used, with minimum payments, financial security and stated cancellation fees. Two limits govern how much of the cost problem this reaches. The signatories are hyperscalers, which is to say the tenants: colocation developers, who build and own a large share of the capacity actually seeking interconnection, are not party to it, and neither are the utilities that decide how costs are divided in the first place. Hence the White House move to convene a second, wider pledge. And a commitment to pay is only worth the tariff that bills it: where the transmission provider charges embedded rather than incremental rates, upgrades driven by a large load are recovered from everyone regardless of what its owner has undertaken to pay. That makes the pro forma cost-recovery agreement of Category 2 the operative instrument and the pledge a statement of willingness to sign one.

ASSESSMENT **analytical interpretation, not a cited fact** **confidence: EMERGING**

Direct assignment is easy to state and hard to administer. Network upgrades arrive in lumps and physics shares them, so any “but-for” counterfactual rests on a modeling exercise that leaves wide discretion — and invites years of litigation over the baseline. The deeper problem runs upstream: capacity-market cost tracks forecast load, so today's socialized costs rest on projects that may never materialize. That makes queue integrity (Section 1) a prerequisite for solving cost allocation rather than a parallel workstream. Correcting the forecast should remove a large share of the cost shift before any allocation dispute begins.

4. Large-Load Ride-Through and Dynamic Performance

The issue. Most of the public debate misses this reliability issue; NERC treats it as among the most urgent. Power electronics drive large computational loads, and their protection schemes guard servers rather than the grid. Faced with a voltage sag — even a normally cleared fault hundreds of miles away that never threatened their service — they transfer to UPS and drop hundreds of megawatts off the system in under a cycle, uncommanded and unannounced. NERC's January 2025 incident review documented the archetype: on July 10, 2024, a shunted surge arrester on a 230 kV line triggered multi-shot autoreclosing, producing six recordable voltage depressions. About 1,500 MW of voltage-sensitive demand disconnected. With load gone and generation still spinning, frequency and voltage rose — frequency peaked near 60.047 Hz and took about four minutes to settle. NERC documented a comparable cryptocurrency-load event in January 2026, and has traced sub-synchronous oscillations to computational facilities in both the ERCOT and Dominion footprints.

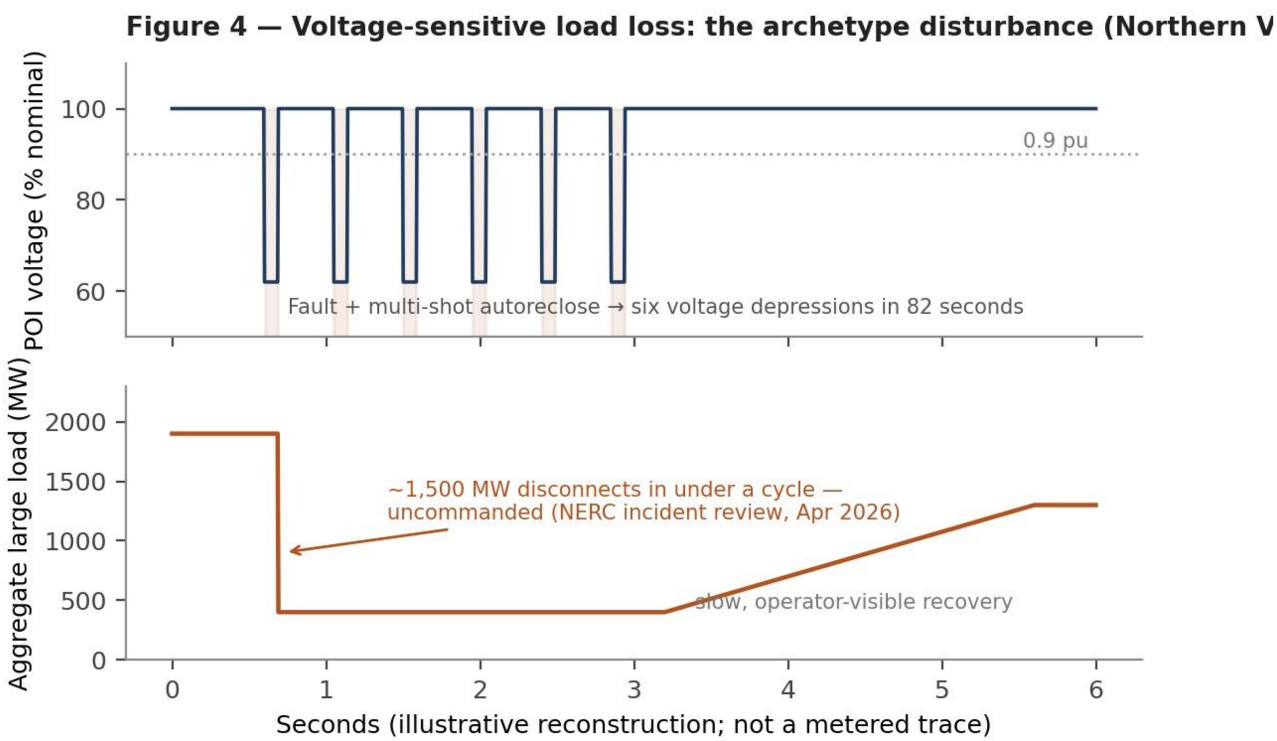


Figure 4 — The failure mode. A voltage disturbance that the load is physically capable of riding through instead produces a large, instantaneous, uncommanded loss of demand. This matters because the resulting contingency approaches the loss of a large generating unit, but arrives without notice and without any obligation attaching to the load. The asymmetry is the point: generators have been required to ride through such events since 2008, and loads have not (Section 4).

Why voltage, and why milliseconds

The sensitivity follows from what sits between the utility feed and the processor. Servers are powered by switched-mode supplies whose front end is a bridge rectifier converting incoming alternating current to a direct-current bus. That bus holds very little stored energy — milliseconds, not seconds — and the equipment downstream of it behaves as a constant-power load: as input voltage falls, current rises to hold output power steady. A motor slows and draws less when voltage sags, riding the disturbance out on its own inertia. A rack does the opposite, drawing harder into a weakening supply, which is why the protective response has to be fast and why it is set to trigger early.

The governing document is not a grid standard at all. Information-technology equipment is designed to the ITIC (Information Technology Industry Council) curve, published originally by the Computer and Business Equipment Manufacturers Association, which defines the voltage excursions such equipment should tolerate. At 0.7 per unit the curve permits equipment to disconnect within 20 milliseconds, and uninterruptible supplies built to IEC 62040-3 transfer to battery when voltage falls below roughly 0.7 to 0.8 per unit, within one to three cycles — 17 to 50 milliseconds. Now set that against the grid side: a transmission fault depresses voltage to 0.25–0.40 per unit for 42 to 66 milliseconds while protection clears it correctly. The two thresholds were engineered independently, decades apart, by different industries for different purposes. They coincide almost exactly. A fault cleared to specification lands squarely inside the window in which the facility is designed to remove itself.

What the grid actually sees then depends on the uninterruptible supply's topology, which is why identical-looking facilities behave differently in the same event. A unit operating in bypass or economy mode, with the load fed from the line for efficiency, reaches its threshold in about 20 milliseconds and transfers the entire site to battery in one step — a cliff edge, and the worst case for the system. A double-conversion unit, where the load always runs off the inverter, supplements from its own battery and the grid sees a reduced rather than eliminated draw, tripping only if the dip persists beyond roughly 150 milliseconds. A rotary or diesel-rotary system transfers to a flywheel and starts its engine — and does not hand the load back quickly once voltage recovers, which is why roughly 1,260 MW of the July 2024 event stayed off the grid for several hours after a disturbance lasting 82 seconds. Cooling plant is voltage-sensitive on its own account and can trip independently of the IT load, so a facility may shed in stages rather than at once.

Two consequences follow, and they frame the rest of this section. The first is that nothing malfunctioned. The facility performed exactly as designed, and its design objective — uninterrupted service to the computing load — was achieved by the very act that harmed the system. This is a collision between two correct engineering specifications, not a defect, which is why it cannot be fixed by enforcement of existing rules and requires a new obligation. The second is that the triggering disturbance can be small. A single-phase fault producing 0.7 per unit on the faulted phase is sufficient, and NERC's January 2026 reviews found facilities where neutral overcurrent protection tripped the whole site on exactly such an unbalanced sag, and others where transformer winding configuration passed the depression straight through to the equipment rather than tempering it. The remedies at that level are unglamorous — ground rather than neutral overcurrent protection, delta-wye staging, wider transfer thresholds — and considerably cheaper than the alternatives discussed below.

The record settles a question that decides how proportionate these rules are. If large loads disconnected only during extraordinary events — a cascading failure, a protection scheme that misoperated, a fault the system was never designed to survive — then requiring them to ride through would be asking for tolerance the grid itself does not offer. The documented record largely does not support that reading. In the great majority of the documented record the protection system worked correctly: in the July 2024 event protection detected and cleared the six faults in 42 to 66 milliseconds, well within design expectations, and the initiating failure of a lightning arrestor counts as ordinary equipment failure, which transmission systems experience continually. Two qualifications apply. The first event in ERCOT's tabulated record, on October 12, 2022, involved a breaker misoperation that delayed clearing to roughly ten cycles — abnormal, and a more severe disturbance than the system requires equipment to withstand. And ERCOT's largest large-load event, in west Texas on December 7, 2022, escalated: it began with a single-line-to-ground fault cleared normally in three cycles, then developed related faults including a three-phase fault caused by a breaker failure. The record therefore lacks uniformity, though those two events sit among dozens and the pattern holds in the rest, including the January 2023 west Texas event cleared normally in four cycles and the largest event outside Texas. ERCOT frames its own record the same way: these are reductions during *normally cleared* faults, with protection operating as designed. The system performed as designed, and about 1,500 MW of load disconnected regardless.

Date	System / location	Load lost	Trigger and clearing	System effect and recovery
Oct 12, 2022, 05:56 CT	ERCOT — west Texas, 138 kV	415 MW (about 100 loads in the area briefly cut ~500 MW)	Three-phase fault. Clearing **delayed to about 10 cycles by a breaker misoperation** — abnormal.	First event in ERCOT's tabulated record.
Oct 31, 2022, 23:12 CT	ERCOT — Dallas–Fort Worth	500 MW	Transmission fault.	Multiple large loads affected.
Dec 7, 2022, 03:50 CT	ERCOT — 138 kV, west Texas (Odessa area)	1,560–1,600 MW: data centers, oil and gas, other industrial	Single-line-to-ground fault cleared normally in 3 cycles, then **escalated: related faults including a three-phase fault caused by a breaker failure** — delayed clearing, not a normal disturbance.	Frequency spiked to 60.235 Hz, returning to 60 Hz in 12 min 30 sec. Two thermal generators tripped (112 MW). Solar unaffected — night. Only five loads were in ERCOT's interconnection process, contributing 212 MW of the total.
Jan 23, 2023, 12:19 CT	ERCOT — west Texas, 138 kV	177 MW across six tracked loads	Single-line-to-ground fault, **normal 4-cycle clearing**.	Typical of the recurring pattern.
Jan 2023 – Sep 2025	ERCOT — Central and far west Texas, Panhandle, North zone	26 events above 100 MW; typically 100–400 MW each	Normally cleared faults. Protection operated as designed; the loads, mostly cryptocurrency facilities, dropped anyway.	No customer outage. ERCOT reduced System Operating Limits on affected interfaces because the possible load loss had itself become a contingency.
Jul 10, 2024, 19:00 ET	Eastern Interconnection — northern Virginia (Dominion / PJM)	~1,500 MW across 60 delivery points and 25 substations	Lightning arrestor failure on a 230 kV line. Auto-reclosing, set for three attempts at each end, produced six successive faults in 82 seconds. **Each cleared properly, in 42–66 ms.** Voltage fell to 0.25–0.40 per unit.	Frequency rose to 60.047 Hz, back to 60.0 Hz in about four minutes. Voltage reached 1.07 per unit; operators removed shunt capacitor banks. About 1,260 MW stayed off for several hours. No customer lost service.
2024–2025 (multiple)	Eastern and ERCOT Interconnections	Several events at or above 1,000 MW	Voltage dips caused by ordinary grid faults.	The evidentiary basis for NERC's Level 3 alert.
Jan 2026	ERCOT — two NERC incident reviews	Crypto facilities; magnitudes not published	Temporary voltage dips from faults. Root causes: neutral overcurrent protection tripping whole sites on single-phase depressions, and transformer winding configuration passing sags through to the load.	No system consequence; produced design recommendations (ground rather than neutral overcurrent protection, delta-wye staging).
May 4, 2026	North America	—	NERC issues a Level 3 Essential Action Alert citing the record since 2022.	Responses due August 3, 2026 (Watchlist).

Table 4A — Documented large-load disconnection events. Two columns run thinner than they should: no one publishes facility-level identities, and recovery durations survive only where an operator reconstructed them afterwards. Both gaps trace to the same cause — large loads are not registered entities, so no one can compel the data.

Sources: ERCOT, *Large Load Loss/Reduction Events 2020–2024* (PDCWG, Nov 19, 2024); ERCOT Large Load Working Group, Sept 19, 2025; NERC incident reviews (Jan 2025; Jan 2026); NERC Level 3 Essential Action Alert, May 4, 2026.

Two patterns in that record deserve emphasis. The first is that severity is not driven by the size of the disturbance but by the concentration of load behind it: a 42-millisecond fault removed about 1,500 MW because 60 delivery points across 25 substations all saw the same sag and all responded identically. Load diversity, which has historically kept aggregate demand smooth and predictable, does not exist among facilities built to the same reference design with the same protection settings. The second is the asymmetry in recovery. Frequency recovered in about four minutes, but roughly

1,260 MW of the lost load did not return for several hours — so the system had to hold four minutes of surplus and then absorb a multi-hour reconnection ramp of a size no generator would be permitted to impose without scheduling it.

One distinction governs how the table should be read, because the figures invite a comparison that does not hold. Everything above records *uncommanded* load loss: facilities disconnecting themselves, on their own protection logic, without instruction. Uncommanded loss differs from involuntary load shedding, where an operator deliberately interrupts firm customers to save the system: the Southwest blackout of September 2011 at about 2,700 MW, the Northeast blackout of August 2003 at more than 61,000 MW, Winter Storm Uri in February 2021 at up to roughly 20,000 MW. In those events customers lost power. In the events in Table 4A, no customer did. The two categories are not additive.

They are, however, the right yardstick for scale. The uncommanded loss of about 1,600 MW in west Texas in December 2022 — the event isolated below, and the only one in the table large enough to make this comparison — is more than half the firm load interrupted in the 2011 Southwest blackout, and it arrived with no operator decision, no warning and no scheduling. ERCOT's modelled threshold of about 2,600 MW sits between the two. A disturbance class that has so far produced no customer outage is therefore operating within the same order of magnitude as the events that did.

One number in the December 2022 row deserves separate attention. Of about 1,600 MW that disconnected, only 212 MW came from the five facilities then inside ERCOT's large-load interconnection process. The remaining seven-eighths sat outside it — unstudied, unmodelled, and invisible to the planning cases that were supposed to anticipate exactly this. A ride-through standard binds the facilities an operator can identify, which is why registration (Section 4, and the December 31, 2026 filing) is not a separate reform from ride-through but a precondition for it.

READING THE DECEMBER 2022 WEST TEXAS EVENT

The largest event in the table also fits the pattern least well. Size alone has made that event the most quoted of the set, so three of its features warrant separate statement. Its clearing ran abnormally: an initial single-line-to-ground fault cleared in three cycles, but related faults followed, including a three-phase fault caused by a breaker failure. That disturbance therefore exceeded the one the ride-through debate concerns — a fault the protection system clears correctly. Its load ran mixed: ERCOT's own account records data centers together with oil and gas and other industrial load, and the oil and gas facilities contributed substantially — variable-frequency drives tripping on undervoltage, which fails differently from a server rectifier reaching the edge of the ITIC curve, and answers to different remedies. And its name is shared: the Odessa Disturbance reports NERC published for 2021 and 2022 concern solar inverters disconnecting, not load, and the June 4, 2022 event discussed below belongs to that series. This report therefore uses the December 2022 event for scale, and for the registration figure above, and takes the July 2024 northern Virginia event — every fault cleared correctly, the load overwhelmingly computational — as the archetype the ride-through rules actually address.

A precedent for imposing such obligations exists, because the generation side of the grid addressed the same question first. ERCOT introduced ride-through requirements for new generators in 2008, grandfathering some existing wind farms — the same structure, and the same grandfathering compromise, that NOGRR282 and NPRR1308 applied to load seventeen years later. Those generator requirements were tightened after the Odessa event of June 4, 2022, in which a single 345 kV single-line-to-ground fault caused 2,555 MW of generation to drop off — 1,711 MW of it solar — taking frequency down to 59.7 Hz and consuming 2,343 MW of the 2,442 MW of responsive reserve available at that moment. That left roughly 100 MW of unused responsive reserve, and the cause was equipment that did not ride through a fault of a type the system is designed to clear.

Load reached comparable magnitude within six months. The December 7, 2022 west Texas event removed about 1,600 MW of mixed load and pushed frequency to 60.235 Hz — an excursion five times larger than the July 2024 Virginia event, taking twelve and a half minutes to recover rather than four. ERCOT drew the conclusion at the time, telling its board in June 2023 that the event showed the need for an improved interconnection process for large loads and better simulation models of new load types. The generators' obligation dates from 2008 and the loads' from 2025 — a seventeen-year interval across which the underlying physics did not change.

A substantial regulatory gap remains. Generators must ride through voltage and frequency excursions under the NERC protection and control standards PRC-024-4 and, for inverter-based resources, PRC-029-1. No mandatory NERC Reliability Standard presently applies to large loads at all. ERCOT has closed that gap inside its own footprint.

NOGRR282 and NPRR1308 impose frequency and voltage ride-through obligations on Large Electronic Loads, approved through the stakeholder process, and grandfather facilities that had cleared energization approval or completed their interconnection study by November 14, 2025. Those obligations bind by ERCOT protocol rather than by continent-wide standard, and ERCOT itself flags the exempt population as a residual risk. From the system's perspective a 1 GW load trip equals a 1 GW generator trip with the sign reversed — yet only one carries regulation as such across the interconnections.

ERCOT has put a number on the exposure. In a September 2025 study built on a 2030/31 case containing roughly 15.2 GW of modelled large electronic load, ERCOT screened three-phase faults and identified about 2,600 MW as the threshold of instantaneous load loss beyond which frequency excursions become significant — losses above that level could carry system frequency past 60.4 Hz. Measured against that threshold, the fleet is already large: by the April 2026 Large Load Working Group about 9 GW of large load had been approved to energize, against an observed simultaneous peak near 3.7 GW in March 2026. The gap between approved and observed is diversity in operating patterns, not head-room, and it narrows as facilities ramp toward their contracted demand.

The cost of that exposure is already being paid in operations rather than waiting on a future event. ERCOT has reduced System Operating Limits on some interfaces because a sudden loss of load could by itself cause a violation — which means the possibility that large loads may trip is now constraining how much power can be moved across the network before any fault occurs. Every user of those interfaces bears that constraint, on account of the ride-through behaviour of a subset of loads.

One feature distinguishes this problem from every other in this report: the rules are anticipatory. ERCOT characterises the impacts of the events observed so far as minimal, and it has not identified a case in which a conventional data center reduced consumption in response to a transmission-level voltage disturbance — the documented events involved other classes of large electronic load. No customer is on record as having lost service because a large load disconnected itself. The standards are therefore being written against a modelled threshold rather than an experienced failure. The anticipatory character constitutes both their justification and their principal political vulnerability: they impose cost now to prevent an event that has not yet occurred, and opponents can accurately observe that nothing has gone wrong yet. Every other section of this report describes a problem with a cost already incurred and measurable — auction prices, household bills, stranded assets. This one does not, and it is the one NERC treats as most urgent.

The evidentiary position is limited, for a structural reason. Because large loads are not registered Market Participants, ERCOT has limited means to perform thorough event analysis when these disturbances occur — it can see the aggregate signature on the system but cannot compel the facility-level data that would explain it. The registration work described below is therefore not procedural but a precondition for knowing whether the modelled threshold errs toward caution or optimism, and it explains why the December 31, 2026 filing deadline matters more than its procedural framing suggests.

Proposed solutions

- NERC Level 3 Essential Action Alert (May 4, 2026). Seven essential actions spanning modeling, studies, instrumentation, commissioning, operations, protection and control. Listed registered entities owed responses by August 3, 2026. A Level 3 Alert ranks as NERC's most serious non-standard instrument.
- NERC Reliability Guideline: Risk Mitigation for Emerging Large Loads (May 2026). Voluntary, but it is the template for the mandatory rule: treat clustered voltage-sensitive load loss under a single contingency as a reserve-sizing input comparable to generation MSSC under the NERC resource and demand balancing standard BAL-002; require Interpersonal Communication capability between large loads and their TO/DP; make large loads subject to Operating Instructions (TOP-001/IRO-001) with trained personnel (PER-005).
- Registration criteria for computational loads. NERC defines physical and electrical thresholds above which large loads become NERC-registered entities with direct compliance obligations. Draft criteria went out for comment through May 15, 2026.
- Project 2026-02 Reliability Standards. Standard Authorization Request posted April 1, 2026; an initial mandatory large-load Reliability Standard should arrive by the end of 2026.

- **FERC Docket RD26-7-000 (July 16, 2026) — the schedule becomes an order.** Acting on its own motion, FERC directed NERC to file one or more new or modified mandatory Reliability Standards addressing bulk-power-system risks from computational-load integration **by December 31, 2026**; to revise its Rules of Procedure to require **registration of computational-load entities** by the same date; and to file a **Phase II work plan** for further standards **by March 1, 2027**. The order rests on the record NERC assembled in RM26-4 and was framed by the Chairman as removing any doubt that NERC's own accelerated timetable is not optional.
- ERCOT NOGRR282 / NPRR1308. ERCOT has moved ahead of NERC with mandatory frequency and voltage ride-through obligations for Large Electronic Loads — the first such requirements imposed on demand in North America — plus dynamic modeling requirements and mandatory registration under NPRR1325. The rules grandfather facilities that had cleared energization approval or completed their interconnection study by November 14, 2025, an exempt population ERCOT has itself identified as a continuing reliability exposure.
- Modeling reform. Replacing static ZIP/constant-power load models with dynamic models that actually reproduce voltage-sensitive trip behavior; EMT-level representation at the point of interconnection (POI). NERC has a dedicated Data Center Load Modeling Workshop scheduled for September 15–16, 2026.
- Facility-side hardware. Wider UPS ride-through windows, coordinated protection settings between POI and facility switchgear, and POI voltage support (STATCOMs/SVCs). Research work is also examining circuit-breaker-operated braking resistors to absorb the energy imbalance when GW-scale load does drop.

ASSESSMENT analytical interpretation, not a cited fact confidence: LIKELY

Engineering, not economics, binds this section. Ride-through capability turns on protection settings and UPS design rather than new physics, and costs little at the margin — but it imposes bulk-system obligations on entities (hyperscalers, colocation tenants, crypto miners) that have never operated as regulated grid participants, and whose internal incentives favor protecting the compute. Registration therefore carries much of the weight of the NERC program: without it the standards bind no one. RD26-7-000 settles whether registration happens and by when, while leaving the registry threshold and the standard's substantive content open. The dispute should center on scope thresholds rather than on the technical requirements themselves. Two features distinguish this section from the others. Its rules anticipate rather than respond: no customer has yet lost service from a large-load disconnection, so the case rests on a modelled threshold rather than a realised failure — at once the justification for acting early and the argument opponents will deploy. And its costs are already being paid elsewhere — reduced System Operating Limits transfer the consequence to every user of the affected interfaces, which makes a ride-through standard a transmission-capacity measure as much as a reliability one. The evidence problem is structural rather than incidental: in the largest ERCOT event only 212 MW of about 1,600 MW sat inside the interconnection process, so the record itself understates what is not registered.

5. Co-Location, Behind-the-Meter Generation, and Electrically Proximate Load

The issue. If the grid cannot serve you, put the plant on your own site. Co-location — siting a large load directly at an existing generator, typically nuclear — offers the fastest route to gigawatt-scale power, and it broke the tariff. Every unresolved question turns on how much the arrangement still leans on the grid: whether the co-located load counts as a transmission customer at all, whether the host generator can still sell capacity, what happens when the load draws during a unit outage, and who pays for the transmission and ancillary services it still uses. The rules now in force trace back to a single rejected contract. PJM filed an amended interconnection service agreement for the Susquehanna nuclear station in June 2024, raising the load Amazon Web Services could take behind the generator meter from 300 MW to 480 MW, with a studied path to 960 MW once generation-deliverability violations were resolved. Exelon and AEP protested. FERC rejected the amendment 2–1 on November 1, 2024 on a narrower ground than the debate around it suggests: PJM had not met the high burden of showing its non-conforming provisions *necessary*, largely because those provisions tracked PJM's generally applicable co-located load guidance and therefore read as terms PJM would offer any similar customer rather than terms unique to Susquehanna. The Commission left the substantive questions expressly unresolved, and the concurrence recorded the rejection as without prejudice. The arrangement failed for its generality, in other words, rather than for any illegality. FERC opened a §206 show-cause proceeding on the tariff itself three months later, and answered a large part of that question on December 18, 2025, in an order under section 206 of the Federal Power Act declaring core portions of PJM's tariff unjust and unreasonable and directing a replacement rate. The order required PJM to establish clear interconnection and operational rules for generators serving co-located load; to require the transmission customer serving that load to choose among four service options — Network Integration Transmission Service (NITS), a new interim non-firm service available only to customers seeking NITS, a new Firm Contract Demand service, and a new Non-Firm Contract Demand service. The interim service warrants more attention, because it lets co-located load energise on non-firm terms while the network upgrades behind its NITS request remain unbuilt — an earlier energisation date in exchange for curtailability, which is the Section 6 bargain appearing inside the co-location tariff. The order further directed PJM to adjust Capacity Interconnection Rights to reflect reduced net injection when the unit serves on-site demand, to revise the behind-the-meter generation rules, and to grandfather certain existing contracts through a transition period. Rates and terms for the new services went to a paper hearing. Generator owners sought rehearing, arguing that mandating availability to operate amounted to a physical and regulatory taking; FERC issued its order on rehearing on June 18, 2026. The June 18 orders also introduced a concept worth watching: “electrically proximate large load” — load close enough to the generator's point of interconnection that the combined system impact is effectively the same as if they shared a substation (FERC's example: no more than two substations away). This closes the workaround of moving the load just off-site to escape co-location rules.

What happens when the on-site generator is not there

Co-location arrangements are typically described by their behaviour under normal operation. The regulatory questions concern the other case. A host generator is unavailable for scheduled maintenance several weeks a year, trips unexpectedly at a rate its forced-outage history sets, and may be curtailed or redispatched by the operator for reasons that have nothing to do with the load beside it. In each case the load’s demand does not disappear. It transfers to the grid, instantly and without notice unless the tariff requires otherwise.

That transfer is the mirror image of the Section 4 failure mode, and the same magnitude. A gigawatt of load appearing on the transmission system in one step because a host unit tripped is a contingency of the same size as a gigawatt of load vanishing, and the system has to hold frequency and voltage through it either way. The new service definitions address exactly this case. Under a Firm Contract Demand arrangement the grid commits to serve that fallback and the load pays network rates for the privilege, so the capacity is planned and reserved. Under Non-Firm Contract Demand the load has bought a cheaper, faster product and accepts that when its generator is unavailable the grid may decline to replace it — which is the same trade Section 6 describes, arriving through the co-location door. The Capacity Interconnection Rights adjustment closes the third case: a unit serving on-site load no longer injects that capacity into the market, so it cannot be counted as available to everyone else at the same time.

The practical question for a developer is therefore not whether on-site generation exists but how the arrangement behaves in the hours when it does not run. Three parameters decide that: the host unit's forced-outage rate and planned maintenance schedule; whether the fallback is contracted as firm or non-firm; and whether the site can reduce its own draw fast enough to ride out a host trip without leaning on the grid at all. A campus that can do the third has converted a reliability exposure into a flexibility product. One documented event indicates how demanding the third is. When a Susquehanna unit went out in November 2023, the co-located facility drew power from the grid for several hours rather than transferring to the other unit as intended; Susquehanna paid the transmission owner for the service used, and afterwards installed additional isolation equipment and developed further protective measures with that owner. The event stayed small and settled cleanly, but it remains the only operating record the co-location debate has, and it ran the way the protestants had predicted.

The generation itself spans a wider range of technology and size than the phrase "on-site generation" suggests, and the choice determines both the permitting path and the behaviour above. The classes now being built at data centers, with the size ranges typical of each:

Technology	Typical unit size	Site build-out	Start capability	Binding constraint
Reciprocating gas engines	2–20 MW	Dozens of units to reach 100–500 MW	Minutes; inherently N+1 because units are many and small	Air permitting scales with unit count; local emissions and noise opposition.
Aeroderivative turbines	30–60 MW	Two to ten units	Roughly 5–10 minutes to full output	Same manufacturer order book as grid-serving units.
Heavy-duty (frame) turbines, combined cycle	200–500 MW per block	One or two blocks	Hours; economics assume near-continuous running	Multi-year lead times booked to 2029–30 (Section 2).
Fuel cells	0.4–10 MW modules	Modular, to tens of MW	Continuous; poor at rapid cycling	Capital cost and firm fuel supply.
Battery storage	2–20 MW modules	50–500 MW	Sub-second; bridges rather than supplies	Duration. It shifts energy in time and produces none.
Existing nuclear (co-location at the host plant)	800–1,250 MW per unit	Host plant	Continuous baseload	The tariff and jurisdictional questions this section describes.
Diesel standby gensets	2–3 MW	Dozens, sized to full site load	Seconds, but permitted only for emergencies	Permits restrict annual run hours; not a supply resource.

Table 5A — Behind-the-meter and co-located generation classes. Sizes are the unit ranges typical of each technology rather than a survey of installed projects. Two patterns matter more than the individual rows. Reciprocating engines dominate fast-deployment projects because many small units give redundancy and can be permitted incrementally, at the cost of a larger emissions footprint per megawatt. And every gas-fired option above draws on the same constrained manufacturing base as the grid's own orders (Section 2), so on-site generation competes with utility procurement rather than escaping it.

Sources: Indicative equipment class ranges as deployed at data-center scale; not a survey of installed projects.

Why behind-the-meter generation still has to be visible to the operator

Co-location proposals frequently treat generation on the customer side of the meter as a private matter between developer and supplier. Four considerations, two physical and two regulatory, make it a system matter.

First, a planning case can only reflect what it contains. If several hundred megawatts of on-site generation and the load it serves are absent from the model, every study run on that case — power flow, short circuit, stability — answers a question about a system that does not exist. The December 2022 west Texas event in Section 4 demonstrates the point: seven-eighths of the load that disconnected was outside the interconnection process, and therefore outside the cases meant to anticipate it. Second, reserve sizing depends on knowing the largest credible loss. A co-located pair forms a single point of common failure: lose the host unit and the grid inherits the load; lose the grid tie and the arrangement islands. Neither event can be sized for if the operator does not know the arrangement is there.

Third, the arrangement uses grid services whether or not it buys grid energy. The system supplies voltage support, frequency regulation, and the reserve held against the host unit's forced outage; a facility invisible to the operator receives all three without paying for any — Section 3's cost-shift in a different form. Fourth, and most directly, reliability obligations attach to registration rather than to metering. NERC standards bind registered entities; a generator large enough and connected at high enough voltage falls inside the bulk electric system definition regardless of which side of a revenue meter it sits on, and the computational-load registration criteria due December 31, 2026 extend the same logic to the load. ERCOT has already made the operational version of this concrete: the WLPUN designation requires separate telemetry so that the facility's real-time draw is visible, and an imbalance unresolved within one minute lets ERCOT limit or suspend the arrangement.

Proximity decides which rules apply, measured electrically rather than in metres. FERC's June 18 orders introduced "electrically proximate" load for exactly this reason — load close enough to the generator's point of interconnection that the combined system impact is effectively the same as if the two shared a substation, with no more than two substations between them offered as the working example. Below that distance the pair behaves as one element under a fault and should be studied as one. Physical proximity is governed by a different and often tighter set of constraints: gas-fired generation must reach a pipeline of adequate pressure and capacity, needs cooling water or air-cooled condensers, and requires enough setback for emissions dispersion and noise. In practice these requirements, not the electrical ones, determine whether the generation can sit on the same parcel as the compute or must be sited a short distance away — at which point the electrical test decides whether it is still a co-location arrangement in the eyes of the tariff.

One behaviour of on-site generation deserves separate treatment, because it can convert a resource into a hazard. Generation configured to support voltage — a synchronous machine with its regulator in voltage-control mode, or an inverter running volt-var control — will inject reactive power when it senses a sag. During a transmission fault this is the same action the system operator is taking through capacitor banks, tap changers and the reactive capability of grid-connected generators, and the two are not coordinated with each other.

Four interactions follow. Control loops with comparable response times hunt against one another: an on-site regulator and a utility load-tap changer can chase the same voltage target and oscillate. Post-fault overvoltage is the sharper risk, and the July 2024 event already demonstrated it — voltage rose to 1.07 per unit after the load disconnected and operators had to remove shunt capacitor banks to recover it. Var injection from a large installed base of behind-the-meter generation, arriving at the same moment and unseen in the operator's model, works in the same direction as the disturbance rather than against it. Third, fault-current infeed from on-site generation alters the current a utility relay sees, which can desensitise protection coordinated on the assumption that fault current flows one way. And fourth, anti-islanding logic and ride-through requirements pull in opposite directions: one requires the unit to disconnect when it detects an island, the other requires it to remain connected through a disturbance, and the settings that satisfy both are narrow.

None of this argues against on-site generation providing voltage support, which is a genuine system benefit where it is coordinated. It argues that the capability cannot be dispatched privately. Reactive support is a system service, and a unit providing it unilaterally, on settings the operator has not reviewed and cannot see, is making an uncoordinated contribution to a system-wide control problem. This is the same conclusion the registration discussion above reaches by a different route, and it applies with more force here: a facility that merely consumes power can be modelled approximately, while one that regulates voltage must be modelled accurately.

What the two largest arrangements actually did

Neither nuclear arrangement that produced this section's rules ended as co-location. Talen and Amazon restructured in June 2025: Susquehanna injects its output into PJM, Talen serves the adjacent campus as a licensed retail electric generation supplier in Pennsylvania, and PPL Electric Utilities handles transmission and delivery. The contract reaches 1,920 MW at full quantity through 2042 and ramps to that volume no later than 2032. Talen timed the transmission reconfiguration to the spring 2026 Susquehanna refuelling outage and described the resulting structure as requiring no Commission approval. Read against the rejected ISA, the parties surrendered the advantage co-location was sought for and bought regulatory certainty with it, at network rates.

Constellation proposed no co-location at any point. Crane restarts as a grid-connected unit selling energy, capacity and clean attributes to Microsoft against that company's consumption across PJM, with no data center at the plant. The restart nonetheless turned on the same instrument the December 2025 order reaches. PJM found that the 765 kV and 500 kV upgrades needed to deliver Crane's full output would not finish before December 2030, three years past the 2027 target, and a nuclear unit held for extended periods below rated output carries vibration and wear risks of its own. Constellation therefore asked to move 760 MW of Capacity Interconnection Rights from Eddystone Units 3 and 4 — running as energy-only resources under Department of Energy emergency orders, and so unable to use those rights — to Crane. FERC granted the waiver on June 1, 2026 over the market monitor's protest. Capacity Interconnection Rights behave here as a transferable asset worth about three years of schedule, which measures what the co-location adjustment in the December order takes away.

	Susquehanna — Talen and AWS	Crane — Constellation and Microsoft
Host unit	Susquehanna, two units, 2,520 MW, Luzerne County, Pennsylvania.	Crane Clean Energy Center, the former Three Mile Island Unit 1, 835 MW, retired 2019.
Original structure	Co-located behind the generator meter; 150 MW authorised behind each unit in 2023.	Never co-located. A restart injecting to PJM, with no data center on the site.
Regulatory event	Amended ISA raising co-located load to 480 MW rejected 2–1, November 1, 2024; §206 show cause opened February 20, 2025.	Waiver granted June 1, 2026 over the market monitor's protest, transferring 760 MW of CIRs from Eddystone Units 3 and 4.
Structure now	Front-of-the-meter: output to PJM, Talen as licensed retail supplier, PPL delivering.	Front-of-the-meter: energy, capacity and clean attributes sold to Microsoft against its PJM consumption.
Contract	1,920 MW at full quantity through 2042, full volume no later than 2032; reconfiguration timed to the spring 2026 refuelling outage.	Twenty years from a 2027 restart; a DOE loan of about \$1B against a stated \$1.6B restart cost.
What it establishes	The tariff question can be avoided by paying network rates — at a price these parties were willing to pay.	A Capacity Interconnection Rights question can decide a project with no co-located load in it at all.

Table 5B — The two arrangements that produced this section's rules, and where each ended. Both began as tests of whether a hyperscaler could buy nuclear output without buying network service, and both now buy network service. The last row carries the comparison: the co-location dispute has so far produced more regulatory doctrine than co-located megawatts, and the instrument it turns on proved decisive in the case that never involved co-location at all. Neither outcome settles whether co-location is uneconomic or merely unresolved.

Sources: FERC, *PJM Interconnection, L.L.C.*, 189 FERC ¶ 61,078, Docket Nos. ER24-2172-000 and ER24-2172-001 (Nov. 1, 2024) (order rejecting amendments to interconnection service agreement), reh'g denied, ER24-2172-002 (Apr. 2025); 190 FERC ¶ 61,115 (Feb. 20, 2025) (show cause); 193 FERC ¶ 61,217 (Dec. 18, 2025); 195 FERC ¶ 61,162 (June 1, 2026) (CIR waiver). Talen Energy Corporation, Form 8-K, Exhibit 99.1 (June 11, 2025). Constellation Energy, Crane Clean Energy Center announcements and waiver request of March 31, 2026.

Proposed solutions

- Firm and Non-Firm Contract Demand service (PJM). Co-located load caps its grid withdrawals contractually; non-firm service is cheaper and interruptible. This architectural innovation converts “how much do you lean on the grid?” from an assumption into a priced, enforceable parameter.
- Capacity Interconnection Rights (CIR) adjustment. Those rights are reduced to reflect the generator's true net injection, preventing a unit from being counted twice — once for the data center and once for the capacity market. Those rights also transfer between plants, and the value of one is now observable: FERC's waiver of June 1, 2026 moved 760 MW of CIRs from Eddystone to Crane, advancing full deliverability for the restart by roughly three years.

- FERC Categories 3–5 (June 18, 2026). All six RTOs must justify or propose rates, terms and conditions for colocation arrangements; transmission service for flexible loads, co-located load, and load with BTM generation; and terms for interconnection customers serving electrically proximate load.
- ERCOT WLPUN and PCLR. Under NPRR1325, the Wholesale Load Point Un-Netted designation and PreCommercial Load Resource election formalize co-located and pre-energization arrangements, with Form X / Form W elections due July 10, 2026 and TSP/DSP submissions to ERCOT by July 24, 2026. If an imbalance is not resolved within one minute, ERCOT may limit or suspend operation of the WLPUN — a hard operational constraint.
- SPP HILL. The High Impact Large Load process, approved by FERC in January 2026, jointly studies the large load and the electrically proximate generation intended to serve it, rather than studying them as strangers.

Figure 9 — Four ways a large load meets its generator, and how much it leans on the grid (per FERC PJM order, Dec 18 2025)

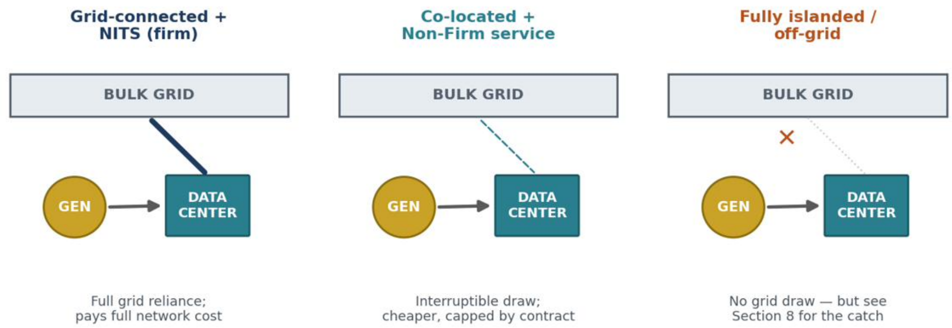


Figure 9 — The four supply configurations available to a large load, running from full grid reliance through co-location and behind-the-meter generation to complete islanding. The middle two matter most because they are where the regulatory question sits: how to price a connection that uses the system intermittently without paying for it as though it did so continuously. The fourth appears to escape the question entirely, and Section 8 sets out why it shifts many system impacts rather than eliminating them. A fifth arrangement sits outside the figure and has absorbed both of the largest nuclear transactions: full network service with a bilateral contract to an identified unit, which carries the commercial character of co-location and none of its system-impact character.

ASSESSMENT analytical interpretation, not a cited fact confidence: **EMERGING**

Co-location exposes the federal/state jurisdictional line more sharply than most other issues treated here. FERC holds authority over transmission service and wholesale sales; a load sitting behind a generator's fence touches neither on the face of the statute, and the takings arguments raised in the PJM rehearing carry weight. Commercially, the logic points most co-located loads toward the non-firm option, because firm service priced at full network cost erases much of the advantage of co-locating in the first place. The paper-hearing rates therefore largely decide the outcome: price non-firm service attractively and co-location converges with demand flexibility (Section 6) into a single product. The reliability question, as distinct from the commercial one, turns on how the arrangement behaves when the host unit runs unavailable. A co-located pair that falls back to the grid on a host trip presents the system with a step change in demand matching the uncommanded losses of Section 4 in magnitude. No operator can plan for it without knowing the arrangement exists, which is why registration and telemetry, rather than the tariff, carry the reliability weight here. The record supplies a partial answer on the commercial question. Both nuclear arrangements large enough to test it took network service instead of co-location before any replacement rate existed, which fits either reading: front-of-the-meter service costs less once regulatory delay carries a price, or the parties routed around an unsettled rule they intend to return to. The paper-hearing rates and the Category 3–5 filings separate the two readings: co-located volume that stays low after the rules land supports the first.

6. Flexible and Curtailable Load — Potential and Implementation Challenges

The issue. Every other section describes a constraint; this one describes a source of relief. Because a large computational load can often give up a small share of its annual consumption — a quarter to a half of one percent — without harming its core function, curtailable service could let the existing grid absorb far more demand than a firm-service forecast implies — a modelled 76 to 98 GW nationwide, under defined participation and curtailment assumptions. Machinery blocks it rather than physics or economics: the industry has not yet built the contractual, operational, and compensation terms needed to deploy flexibility at scale, and a lever of that size sits idle until it does. The grid carries sizing for a peak that occurs a few dozen hours a year. Duke University's Nicholas Institute quantified the implication: across the 22 largest balancing authorities, covering about 95% of U.S. load, roughly 76 GW of new load could be absorbed on the existing system at a curtailment rate of 0.25%, 98 GW at 0.5%, and 126 GW at 1%. Those percentages are shares of energy, not of time — the study defines the rate as annual curtailed megawatt-hours divided by the new load's maximum potential annual consumption — and the distinction matters to anyone signing the contract. Some curtailment occurs in 85 hours a year at the 0.25% rate, 177 at 0.5%, and 366 at 1%. Events average 1.7, 2.1 and 2.5 hours respectively, and most leave the load largely running: across the hours in which any curtailment is called, 88% retain at least half the new load, 60% at least three-quarters, and 29% at least nine-tenths. A 1–2% reduction in data-center peak demand tracks a 0.5–2.8% reduction in electricity rates.

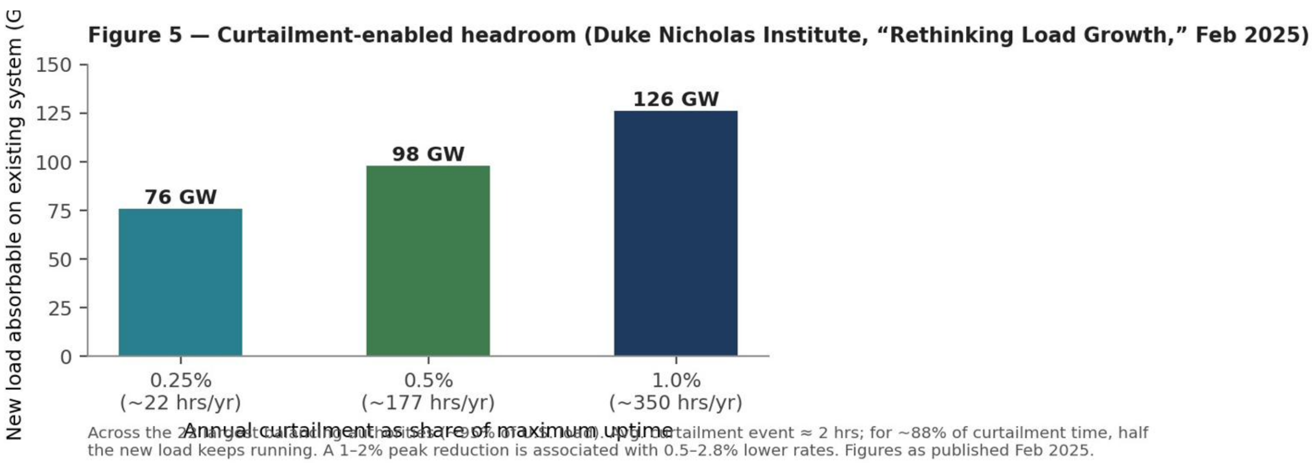


Figure 5 — Curtailment-enabled headroom: additional load the existing network could serve if that load accepts a limited number of curtailment hours each year. The quantity runs large because networks carry sizing for a peak that occurs in a few dozen hours, so a modest, well-timed reduction releases capacity out of proportion to its size. A planner should conclude that flexibility operates on the load's own timescale, where new generation and transmission do not. Three qualifications attach. The figure states a modelled result; it depends on the participation and curtailment assumptions given in Section 6; and it forms an upper bound in one specific respect, because the study does not model network constraints and lists them first among the factors that would reduce it.

The evidence base, and what it agrees on

The **Duke** estimate is the most cited figure in this debate and should not carry the argument alone, so it is worth setting beside the other work. **EPRI's DCFlex** initiative — a consortium of hyperscalers, utilities, grid operators and equipment suppliers — reports from its pilots that 10 to 40 percent load modulation is achievable at a data center without breaching customer service-level agreements, which is a facility-level measurement rather than a system-level model. **ICF**'s analysis anchors on roughly 80 hours of curtailment a year for flexible loads while noting a tail past 300 to 400 hours in extreme conditions. A 2025 **GridLab and Telos** Energy study of NV Energy found that if utilities account for flexible data center operations during long-term resource planning, rather than after investment decisions have been made, they can significantly reduce system costs. By modeling 1–2 GW of data center load that can be temporarily curtailed or shifted during periods of peak demand, the study estimated approximately $300–400 million in net present value savings over 2025–2050. Most of these savings come from avoiding or delaying investments in new generation, transmission, and other grid infrastructure, while requiring data center flexibility for only a limited number of hours each year.

Four separate studies, measuring four different things. The **Duke** estimate is a system figure: how much additional load the existing network could carry if that load accepts a limited number of curtailment hours a year. **EPRI's DCFlex** pilots measure something narrower — the share of its own draw a single facility can shed without breaching its service-level agreements. **ICF** estimates the hours such a commitment would require in practice, anchoring near 80 hours in a normal year with a tail past 300 to 400 hours in extreme conditions. **GridLab and Telos** convert the idea into money, estimating the system-cost savings that follow when flexibility is planned for in advance rather than treated as an afterthought. Because each answers a different question — network headroom, facility capability, hours required, and dollars saved — the four results are not additive, and a reader who stacks them will double-count.

The four agree on direction and rough order of magnitude: a modest, well-timed reduction in demand releases capacity out of proportion to its size, because the grid is built for a peak that occurs only a few dozen hours a year. They diverge on the tail — the number of curtailment hours a flexible load might face in a stressed year — and a developer's counsel reads the tail most closely. A figure near 85 to 177 hours of mostly partial reduction describes the average year, and a developer can plan around it; an obligation that is effectively unbounded in a bad year is a different contract entirely, and the service terms have to resolve the distance between those two readings.

Who directs the curtailment, and what that decides

Two different things travel under the word flexibility, and the studies above measure the first while most of the industry's own examples describe the second. In the first, the operator of the system calls the event: a dispatch instruction, a tariff trigger, a contractual notice, arriving when the system is short and not otherwise. In the second, the operator of the load schedules around the grid on its own initiative — deferring a training run, moving a batch job to a region with headroom, shifting work off the local peak. The second costs far less. Training and batch inference checkpoint cleanly and tolerate delay, the software to orchestrate them exists and deploys in weeks rather than permitting cycles, and the load surrenders throughput on its own schedule rather than someone else's. It also gives the system almost nothing it can plan against. A planner can count a resource only if it is obligated rather than voluntary, verifiable after the fact, available at the system's peak rather than at the load's convenience, and deliverable at the point where the constraint binds. Self-directed deferral satisfies none of those four by construction. The relation runs close to inverse: the less a flexibility costs the load, the less the system can rely on it. The contract exists to resolve that, converting an unobligated capability into an obligated one at a price both parties accept.

Two consequences follow for anyone writing that contract. The flexible share covers a fraction of the campus rather than the campus: interactive inference runs under a latency bound and cannot wait, while training, batch inference and offline processing carry the deferrable part, so a commitment sized against today's training-heavy mix may bind differently on a site that later shifts toward serving. And the same megawatts carry different flexibility depending on who holds the control rights. An operator running its own models can defer its own training; a colocation or cloud provider cannot pause a tenant's workload, and reaches the same obligation through batteries, on-site generation, and the shape of its own contract portfolio instead. That mechanism drives the fleet-depth asymmetry noted earlier: a term costing a global operator almost nothing, because it has somewhere else to send the work, becomes a firm curtailment obligation for a single-site provider with nowhere to send it, and a tariff modelled on the first will misprice the second.

Interruptible load is not a new idea

Curtailable service is neither an untested proposition nor an invention of the data-center era. Aluminum smelters, chlor-alkali plants and other electro-intensive industry have taken interruptible rates for decades, in exchange for prices unavailable to firm customers. ERCOT has dispatched load as a frequency resource since the 2000s through Loads Acting as Resources: on February 26, 2008, roughly 1,100 MW of contracted load was deployed automatically after a frequency decline — load acting as the remedy rather than the disturbance, and the mirror image of every event in Table 4A. Cryptocurrency mining has since become the largest voluntarily price-responsive load class on any North American grid, idling within seconds when prices spike.

The scale is not marginal either. Demand response accounts for about 5% of the capacity cleared in PJM's most recent auctions — comparable to the entire hydro fleet and larger than wind and solar combined in that market. The movement in that figure is instructive. Between the 2026/27 and 2027/28 auctions, cleared demand response rose from 5,531 MW to 7,299 MW with essentially no change in the megawatts offered. The increase came from an accreditation reform approved by FERC in May 2025 that raised demand response's effective load carrying capability. Nearly 1,800 MW of capacity appeared because the rules for counting it changed, which demonstrates the section's argument directly: the binding constraint on flexibility is how it is valued and contracted, not whether the capability exists.

What each region is actually building

The instruments differ by region, and the differences follow from market structure rather than from disagreement about the physics. Regions with centralised capacity markets already have a venue in which flexibility can be paid — so their work is accreditation reform. Regions without one are writing bespoke service products instead.

Region	Instrument	How it works	Where it stands
PJM	Demand response in the capacity market (RPM); Non-Firm Contract Demand for co-located load	DR bids into the Base Residual Auction and is accredited by ELCC like any other resource; non-firm service caps grid withdrawals by contract.	DR is about 5% of the cleared supply mix. An accreditation reform approved by FERC in May 2025 raised DR's ELCC and lifted cleared DR from 5,531 MW to 7,299 MW between the 2026/27 and 2027/28 auctions — with essentially no change in the megawatts offered.
ERCOT	Controllable Load Resource; Emergency Response Service; four-coincident-peak price response	A CLR registers as dispatchable and accepts ERCOT dispatch and ramp-rate instructions; ERS is a contracted emergency reserve; 4CP exposure gives every large load an unpriced incentive to curtail at the annual peaks.	CLR pathway formalised under NPRR1188 and NPRR1325. Cryptocurrency load has been the practical proving ground: highly price-responsive, and able to idle within seconds.
MISO	Load Modifying Resources; Expedited Resource Addition Study on the supply side	LMRs are registered demand-side resources counted toward the planning reserve margin and callable in emergencies.	MISO tightened LMR notification and availability requirements after the 2022–23 winter events, trading volume for dependability.
SPP	Conditional High Impact Large Load Service (CHILLS)	Long-term transmission service granted ahead of the upgrades that would normally be required, conditioned on curtailability and separate telemetry.	Accepted by FERC on June 5, 2026. The closest instrument in force to a bounded, priced curtailment product.
NYISO / ISO-NE / CAISO	Special Case Resources; Active Demand Capacity Resources; Proxy Demand Resource and the Demand Response Auction Mechanism	Aggregator-mediated programs that let retail customers offer curtailment into wholesale capacity and energy markets.	Long-established and modest in scale relative to peak; built for aggregations of small commercial load rather than for single gigawatt-scale customers.
Bilateral utility contracts	Custom demand-response terms inside a large-load service agreement	The utility and the customer negotiate curtailment terms directly, outside any market product.	The fastest-moving venue. Google has integrated about 1 GW across Indiana Michigan Power, TVA, Entergy Arkansas, Minnesota Power and DTE.

Table 6A — Flexible and curtailable load instruments by region. Program scale is given where operators publish it; several of the smaller programs report on different bases and are described qualitatively rather than compared numerically.

Sources: PJM; ERCOT; MISO; SPP; NYISO; ISO-NE; CAISO tariff and auction materials. Program scale given where the operator publishes it.

The bilateral row sets the terms first, and one contract repays close reading because it tests this section's central proposition. In July 2025 Indiana Michigan Power filed a special contract with Google covering a data center campus at Fort Wayne, pairing a clean-capacity arrangement with a demand-response product built for machine-learning workloads. Most commercial terms are confidential, but the filed structure is public: an *unlimited* number of curtailments, each capped at a maximum duration of 12 hours in summer and 15 in winter. Google has since extended the model to

TVA, Entergy Arkansas, Minnesota Power and DTE, reaching about 1 GW of contracted demand response, and reports that the approach lets new campuses connect faster than a firm-service queue position would allow.

That structure is the inverse of the product this section has been describing. This report has argued for a cap on annual hours with bounded event duration and frequency; the I&M; contract bounds duration but leaves frequency open. Both are bounded obligations — a customer can price either — and the divergence suggests the market is converging on *boundedness* as the requirement rather than on any particular parameter. That matters for tariff design: a regulator writing a flexibility product should be less concerned with matching some canonical hour count than with ensuring that every dimension of the obligation is closed and stated. What developers decline is not curtailment but discretion.

Why unlimited events can be the cheaper side of the bargain

For most industrial customers an unlimited number of curtailments would be the harsher of the two terms. That ordering reverses for a workload that is deferrable and measured in total throughput rather than in delivery time. A training run cares about the compute delivered over weeks; whether a given hour of it happens on Tuesday or Thursday changes little, provided the interruption is short enough that checkpointing absorbs it and the cooling plant and the schedule recover. What such an operator must underwrite is the *worst single event* — the outage long enough to break a delivery commitment or strand very expensive accelerators for a shift. Cap the duration and that risk is bounded. The number of events then bears on cost roughly in proportion to lost throughput, which is recoverable, rather than to risk, which is not.

The count is also bounded in practice even where the contract leaves it open, because the counterparty has no reason to call events indiscriminately. I&M; serves its PJM capacity obligation under the Fixed Resource Requirement rather than by buying in the capacity auction, so the value of a curtailment is concentrated in the handful of hours that set its peak obligation. A utility in that position calls the hours it must and leaves the rest alone. An unlimited-count term granted to a counterparty whose own economics confine it to peak hours is a smaller concession than it appears on the page.

The term is affordable to this signatory for a further reason, which does not generalise. Google operates a global fleet and has been shifting workloads between regions and across hours since well before these agreements — the utility contracts followed a demonstration with Omaha Public Power District in which machine-learning demand was reduced across three grid events. For an operator who can move the work, a curtailment at one campus is largely absorbed elsewhere, and the marginal cost of one more event approaches zero. For a single-site colocation provider running one tenant's inference workload under a service-level agreement, the same term is difficult to accept at any price, because the work has nowhere to go.

The asymmetry repeats a pattern identified earlier in this report. Section 1 noted that a $50,000/MW security deposit screens for balance-sheet depth as much as for project credibility. Flexibility terms screen for fleet depth in the same way. A curtailment product written around what a hyperscaler with a global fleet can accept is not a neutral template for the rest of the market, and a regulator who takes the Google contracts as the model may write a tariff that only three or four firms in the country can use. The bounded-hours structure this section recommends is more portable precisely because it does not assume the customer has somewhere else to run.

What the flexibility is actually worth, and to whom

Compensation is only partly documented. The I&M; contract, approved by the Indiana Utility Regulatory Commission in 2026 under Cause No. 46276, compensates Google through credits — one associated with the demand-response structure and one with a parallel Clean Capacity Arrangement, under which Google transfers long-term accredited clean capacity to the utility and bears a penalty if it fails to deliver. The magnitude of those credits, and the contract capacity in megawatts, are redacted.

The funding source is identifiable. As an FRR entity, I&M; must hold capacity against its own peak rather than buy it in the auction. Every megawatt of peak the customer removes therefore spares the utility a megawatt of generation it would otherwise procure or build — at a time when the marginal price of that capacity in PJM has cleared at the cap for three consecutive auctions. The utility's stated case is exactly this: the arrangement reduces its long-term generation requirements and financial commitments, and the savings are shared with other customers rather than accruing only to the signatory. Whether the split is fair is the question the redactions make unanswerable from outside.

The credit is probably not the largest component of the compensation. Google's own account of why it signs these agreements is that flexibility lets new campuses connect to local grids faster, bridging the gap between load growth and the longer timelines required for new generation. Where time to energization binds and the alternative runs to a queue position measured in years, an earlier energisation date carries more value than a per-megawatt payment. SPP's CHILLS makes the same trade explicit, and PJM's Non-Firm Contract Demand offers it as a tariffed product. Flexibility buys speed; the credit follows as a secondary term.

This conclusion raises a governance problem. Should bilateral utility contracts become the principal venue for flexible service — and on current evidence they outpace any tariff or market product — confidential filings will set the terms of the most important reform available. The Citizens Action Coalition intervened in the Indiana docket to argue precisely this, objecting that the redactions extended to contract capacity and to the penalty provisions that protect ratepayers, and that no public ratepayer-impact analysis accompanied the filing. The objection concerns evaluability rather than merit: the public cannot assess a bargain struck on its behalf. A product this report recommends should not arrive in a form that only its two signatories can read (Sections 3 and 7).

For context on why this matters so much: U.S. data-center demand should reach roughly 66 GW in 2027, up from 31 GW in 2025, with the data-center share of summer peak rising from about 4.1% to 8.5%. Berkeley Lab's range spans 74–132 GW of growth and 6.7–12% of U.S. electricity by 2028. Absent flexibility, that translates into an estimated 25–50 GW of incremental fossil generation by 2030.

Proposed solutions

- Non-firm / curtailable interconnection service. The core product. SPP's HILL-adjacent service and PJM's NonFirm Contract Demand both offer a faster energization date in exchange for interruptibility. FERC's Category 4 requires every RTO to address transmission service for flexible large loads.
- ERCOT Controllable Load Resource (CLR) and ramp-rate obligations. Large loads registering as CLRs become dispatchable in the ERCOT market and accept dispatch ramp rates as a condition of that participation; NPRR1325 imposes registration and dispatch obligations on large loads generally, but ramp-rate discipline follows the CLR election rather than applying to every large load.
- Flexibility credited in planning, not treated as an afterthought. GridLab/Telos' NV Energy case study found 1– 2 GW of data-center flexibility produces roughly $300–400 million NPV savings over 2025–2050 when it is embedded in integrated resource planning. Most utilities still model large loads as firm and inflexible.
- Cheaper flexibility from elsewhere on the system. Headroom does not have to come from the data center. Public Service Co. of Oklahoma procured demand-response capacity at about $32/kW-year, versus $100+/kW-year for behind-the-meter generation at a data center. Expanding conventional DR often delivers the same MW of headroom at lower cost.
- Technical flexibility levers inside the facility. Thermal storage (ice banks, chilled-water tanks) can shift 20– 40% of cooling load for 4–8 hours; DVFS modulates GPU/CPU draw; spatial shifting and 'compute peering' move workloads between regions; on-site BESS bridges short events without touching compute at all.
- Transparency as an enabler. Publishing transmission availability, risk assessments, and the specific conditions under which curtailment will be called lets developers price the risk instead of refusing it.

ASSESSMENT **analytical interpretation, not a cited fact** **confidence: LIKELY**

Neither technology nor economics blocks flexibility — contract terms and risk perception do. Operators will accept curtailment; they will not accept an undefined obligation to a risk-averse utility exercising unlimited discretion. ICF's modeling anchors on roughly 80 hours of curtailment per year for flexible loads, but notes that extreme conditions could push the tail past 300–400 hours — and the customer's counsel reads the tail, not the average. On the evidence above, procedural reforms appear to unlock more headroom than technical ones: a bounded, priced, auditable curtailment product specifying a cap on hours, a notification window, and a compensation rate. Flexibility also holds a structural advantage the other levers lack — it requires no new steel. With turbines and transformers booked to the end of the decade (Section 2), demand flexibility substitutes for equipment that cannot arrive in time, and equipment scarcity raises its value. A jurisdiction offering such a product early should capture a meaningful share of data-center investment, though tax, land, fiber, and water will shape how much and how quickly. Two observations qualify that conclusion. The strongest available evidence for the procedural reading is PJM's accreditation change, which added roughly 1,800 MW of cleared demand response between two auctions with essentially no change in the megawatts offered — capacity that appeared because the counting rules changed. Against that, operators who can move workloads between sites negotiate the flexibility terms now being signed, and a term a global fleet can afford may lie beyond a single-site provider; a tariff modelled on those contracts risks selecting for fleet depth in the same way that financial gating selects for balance-sheet depth (Section 1). The confidentiality of the bilateral terms compounds this, since the market cannot converge on a standard it cannot read. One further caution applies to the headroom figure itself: it states an adequacy quantity, computed against a balancing authority's observed seasonal peak, and network constraints fall outside it. A load can hold every flexibility term in this section and still require the upgrades of Section 3, because the hours a local constraint binds need not be the hours a system peaks.

7. Jurisdiction and the Non-RTO Grid: Who Actually Writes the Rules

The issue. The first six sections assume someone holds authority to impose the fix; that assumption remains contested. Section 201(b) of the Federal Power Act gives FERC jurisdiction over interstate transmission and wholesale sales, and expressly reserves to the states authority over retail sales, local distribution, generation, and siting. A data center serves as a retail customer and, physically, as a transmission-connected element whose behavior determines whether the bulk system holds together. Ninety years of settled doctrine did not anticipate a retail customer doubling as a bulk-power-system contingency. DOE's October 2025 §403 directive pushed hard on this line, asking FERC to standardize interconnection for loads above 20 MW connecting to the transmission system. NARUC responded bluntly: FERC asserting jurisdiction over load interconnection sits outside the boundaries of the FPA, and because states set rates across customer classes, a federal service offering aimed at a narrow class of retail customers would displace a balancing act that is solely a state decision. The National Conference of State Legislatures added the practical version of the question: if FERC takes the interconnection decision, who then holds the reliability and supply consequences?

FERC had already shown its hand on this, and the case is worth setting out because the distinction it turns on recurs throughout this section. In October 2025 the Commission rejected Tri-State Generation and Transmission’s proposed High Impact Load tariff, three votes to nil, in Docket ER25-3316. The rejection did not rest on the merits. Tri-State faced roughly 7 GW of data-center interest across its forty-odd member distribution cooperatives in Colorado, Nebraska, New Mexico and Wyoming, and a genuine cost-shifting problem that the tariff was a reasonable attempt to solve. What defeated it was the vehicle. Tri-State operates as a wholesale generation and transmission cooperative, selling power to its member cooperatives at rates that fall under FERC jurisdiction — which explains why the filing went to FERC at all. The High Impact Load tariff did not stop at the wholesale relationship. It specified the terms on which a member cooperative’s own end-use customer would be served — minimum demand obligations, contract duration, collateral and exit charges applying to the data center itself. Those are terms and conditions of retail sale, which §201(b) reserves to the states without qualification. A federal tariff cannot reach through a wholesale contract to set them, however sound the policy behind it, and the Commission has said explicitly, citing Supreme Court precedent, that it may not reach retail sales no matter how direct or dramatic the effect on wholesale rates.

The states hold substantive powers in that lane, not formal ones, which explains how firmly they defend the boundary. A state commission determines which customer classes exist and how costs are allocated among them, so creating a large-load class is an act of redistribution among residential, commercial and industrial ratepayers that only the state may perform. It determines which utility is obliged to serve a customer in a given territory and on what conditions that obligation may be qualified. It determines whether a minimum-take provision, a twelve-year term or an exit fee is just and reasonable as applied to a retail customer, and it approves the resulting rate. And it holds the consequence: if a large load defaults or never materialises, the stranded cost falls on the distribution utility’s remaining retail customers, whose rates that same commission sets. NARUC’s objection is this in one sentence — a federal service offering aimed at a narrow class of retail customers would displace a balancing act that is solely a state decision — and the NCSL’s question follows from it: a federal body that takes the interconnection decision does not thereby acquire the obligation to serve, or the ratepayers who absorb the failure.

The June 18, 2026 orders read best as FERC accepting that constraint rather than testing it. The Commission did not federalize large retail load interconnection. It claimed the transmission lane: service to Eligible Customers, the study processes that determine whether the system can supply that service reliably, network upgrade costs flowing into jurisdictional rates, generator interconnection procedures, and ancillary-service charges for load that still leans on the grid. Against that, it repeatedly preserved state authority over the specific terms of retail sales to large loads, over which entities may serve them, and over siting and construction. That baseline reaches transmission only, which likely explains why FERC chose six regional show-cause orders over a generic national rule: an order construing a particular RTO's tariff is far more durable on review than a rulemaking that has to define the jurisdictional boundary in the abstract.

The December 2025 co-location order had already drawn the same line, and more explicitly. Having found core portions of PJM’s tariff unjust and unreasonable, and having directed three new transmission services in their place, the Commission declined to address the jurisdictional question underneath: whether the interconnection of a retail load served through a co-location arrangement falls to it at all. The position of that reservation matters as much as its content. FERC reached the rates, the terms, the study obligations and the capacity accounting, then stopped at the one question that would have settled which regulator governs the arrangement as a whole. A reader should take the June 18 orders as the same choice repeated across six regions rather than as a new position. On the evidence of both, the jurisdictional boundary will move through litigation or through Congress, and not through the Commission volunteering to redraw it.

Who decides what

The federal-state line generates the litigation, but other boundaries matter just as much, and an actor holding no authority over electricity can stop a project outright. The table below names each participant, the source of its authority, what it decides, and — the column that usually explains the deadlock — what it cannot reach.

Actor	Authority derives from	Decides	Cannot reach
FERC	Federal Power Act §§201–206, 215; Natural Gas Act	Interstate transmission service and its rates; wholesale sales; RTO and ISO tariffs; generator interconnection procedures; network-upgrade cost allocation; ancillary-service charges; approval of NERC standards.	Retail sales and their terms; local distribution; generation and transmission siting; which entity is obliged to serve a customer.
NERC and the Regional Entities	FPA §215; certification by FERC, with authority delegated to six regional entities including Texas RE and WECC	Mandatory reliability standards, and which entities are registered to be bound by them. Registration criteria for computational load are due December 31, 2026.	Anything concerning rates. Any entity it has not registered — which is why registration precedes every reliability obligation in Section 4.
RTOs and ISOs (PJM, MISO, SPP, CAISO, NYISO, ISO-NE)	A FERC-accepted tariff and a stakeholder governance process; not a delegation of governmental power	System operation and dispatch; interconnection study queues and their sequencing; market design and capacity accreditation; planning cases and reliability criteria.	Nothing takes effect until FERC accepts it. No authority over retail rates, over customer classes, or over whether a facility may be built.
ERCOT and the PUCT	Texas law; ERCOT is exempt from FERC transmission jurisdiction under FPA §201(b)(1) because it is not synchronously interconnected across state lines	Protocols (NPRR, NOGRR, PGRR), interconnection process, market design, and retail rules — all within one state, which is why reforms move faster here than anywhere else.	NERC reliability standards still apply through Texas RE. Federal reliability obligations are not escaped by the jurisdictional exemption.
State public utility commissions	State statute, operating in the space §201(b) reserves	Retail rates and tariff terms; which customer classes exist and how costs are allocated among them; the obligation to serve and certificates of convenience and necessity; integrated resource plans; utility capital approvals; transmission siting in most states.	Wholesale rates; interstate transmission service; the contents of an RTO tariff. A state may not bind a neighbouring state whose grid it shares.
State legislatures	Plenary state authority	Statutory mandates for large-load tariffs, moratoria, tax abatement, siting law, and the scope of commission authority itself.	The federal lane, though state law shapes what a utility may offer within it.
Municipalities and counties	Local land-use and police powers, delegated by the state	Zoning, land use and building permits; water and wastewater service; local noise, air and lighting conditions; property-tax abatement; and, increasingly, moratoria on data-center development.	Rates, grid operation and reliability standards. Local authority cannot make a project connect — but it can prevent one from being built at all.
Municipal utilities and electric cooperatives	Municipal charter or member governance; largely exempt from FERC rate jurisdiction under FPA §201(f), and in many states from commission rate regulation	Their own retail terms, including large-load tariffs, set by city council or member board rather than by a commission.	Not exempt from reliability standards where they operate bulk-power facilities. Tri-State's rejected filing shows the limit of routing retail terms through a wholesale tariff.
US Department of Energy	FPA §202(c); §403 directives; national-interest corridor designation	Emergency orders that override economic dispatch and may keep units running; corridor designations that unlock federal siting backstop; directives asking FERC to act.	Ordinary ratemaking and interconnection. A §403 directive asks; it does not decide.
EPA and state environmental agencies	Clean Air Act and state analogues	Air permits for on-site and backup generation, including permitted run hours — often the binding constraint on behind-the-meter generation (Section 5).	Anything about electrical service or grid operation.
Federal courts	Judicial review of final agency action, principally in the courts of appeals	Whether a FERC order survives challenge — which is why the Commission preferred six region-specific show-cause orders to a generic rule that would have to define the jurisdictional boundary in the abstract.	They review; they do not design.

Table 7A — Authority over large-load integration. Read the final column first: almost every unresolved question in this report sits where one actor's competence ends and no other actor's begins.

Sources: Federal Power Act §§201–206 and 215; state enabling statutes; FERC orders as listed in the Catalog.

Three consequences follow from the shape of that table. The first is that no actor holds both the decision and its consequence. FERC and the RTOs determine whether a load may connect and on what transmission terms; the state commission and the distribution utility carry the cost if the load never arrives or fails to pay. A structure that separates the two invites each side to discount what it does not bear.

The second is that several actors hold a veto and none holds an initiative. A county can refuse a zoning permit, a state commission can refuse a tariff, FERC can reject a filing, and an air permit can cap the run hours that make on-site generation viable — but no participant in the table can direct that a facility be built, connected and served on a defined timetable. That asymmetry is why the reforms in Sections 1 to 6 take the form of process changes and price signals rather than directives.

The third concerns the two entries that sit outside the ordinary structure. Cooperatives and municipal utilities escape FERC rate jurisdiction under §201(f) and, in many states, commission rate regulation as well, leaving their large-load terms to a member board or a city council rather than to any regulator — a significant fact given how much of the load is landing in their territories. ERCOT sits outside federal transmission jurisdiction altogether, which is the reason it produces the earliest and most complete reforms in this report and the reason those reforms bind nobody else. Both are reminders that the map is not a hierarchy. It is a set of overlapping competences, most of which were settled before a single customer could constitute a system contingency.

Figure 7 — Where the line falls, and what is still being fought over

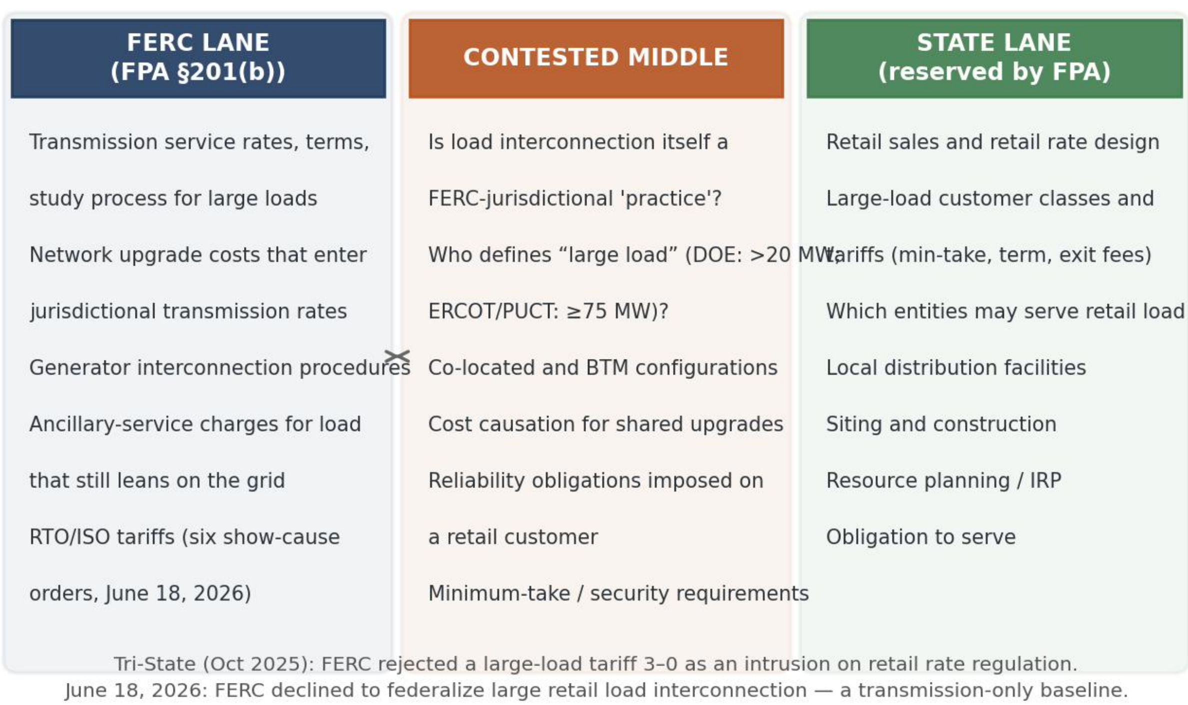


Figure 7 — The Federal Power Act's division of authority, showing the matters clearly federal, those clearly reserved to the states, and the contested band between them. The middle column matters because it contains the instruments large-load policy actually requires — service terms, cost allocation, and the conditions attached to interconnection — none of which sit wholly on either side. Readers should conclude that the difficulty is structural rather than political: no participant holds both the decision and its consequence, which is why remedies must be assembled from instruments held by different authorities (Table 7A).

The consequence: a patchwork, and the utilities nobody is orchestrating

Because the retail lane belongs to the states, fifty commissions now write the substantive commercial rules for large loads fifty times over, on fifty schedules. EEI counted twenty-three states with at least one approved large-load tariff as of May 2026 and another seven pending; SEPA counted at least 65 tariffs pending or in place across 34 states at the end of 2025, forty-six of them new that year. The terms show no sign of converging. AEP Ohio's PUCO-approved tariff covers loads at or above 25 MW. It requires payment for at least 85% of contracted capacity regardless of consumption, imposes a twelve-year minimum term and exit fees, and can add on the order of $10 million in first-year cost for a 100 MW facility. Enverus estimates that it cut connection requests roughly in half. That either shows the tariff filtering speculation or shows it exporting the load to another state, depending on who describes it. Dominion's GS-5 class in Virginia sits at the same strict end. Georgia Power's PLL-18 framework offers notably more flexible terms, and Georgia Power had 28 large-load projects representing about 11 GW under contract as of Q1 2026. Pennsylvania's 2026 model tariff took a third path, adopting a “butfor” cost-causation standard directly.

Figure 8 — Fifty regulators, one problem: the state large-load tariff patchwork

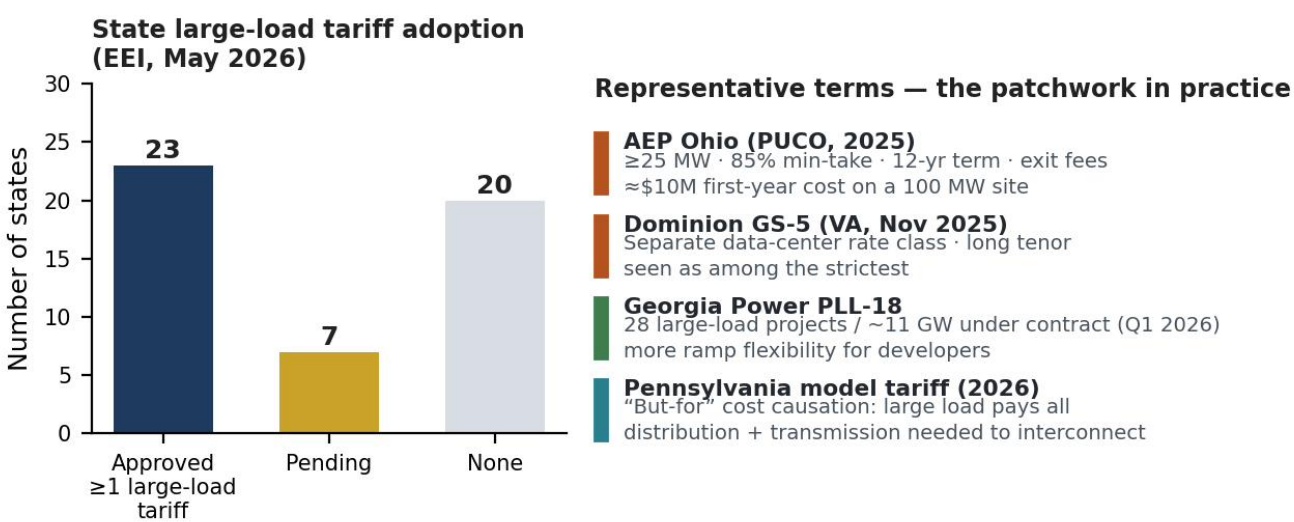


Figure 8 — State adoption of large-load tariffs, and the variation in their terms. Adoption is now widespread, but minimum-take levels, contract durations, exit fees and collateral requirements differ substantially between jurisdictions. The variation matters because these are retail terms reserved to the states (Section 7), so no federal action will harmonise them. For a developer, jurisdiction selection has consequently become a larger financial variable than most engineering decisions.

The second consequence is geographic. Sections 1 through 6 tell an RTO story, and a large share of American data-center load never lands in an RTO. The Southeast, most of the West outside CAISO, and the co-op and municipal systems have no ISO tariff to reform, no stakeholder process to run, and no show-cause order to answer. FERC's Chairman used the June 18 meeting to encourage transmission providers outside RTO/ISO regions to bring comparable reforms voluntarily under FPA §205, and the Commission expressly declined to foreclose future action reaching non-RTO regions. Encouragement carries no deadline. In the meantime the vertically integrated utility does the interconnection study, sets the tariff, plans the resources, and answers to a single state commission — which is faster and more coherent than the RTO process, and also almost entirely opaque to anyone outside it.

The jurisdictional patchwork runs inside the RTO footprint as well as outside it. CAISO does not offer traditional Order No. 888 transmission service, so FERC had to apply a different analytical framework to reach it — and still found substantially similar deficiencies. NYISO’s order turns on cost shifting through the Transmission Service Charge when speculative service requests drive local transmission planning, which is Section 3’s problem arriving through a New

York-specific mechanism. The same failure mode surfaces through six different tariff structures, and no single remedy fits all six.

A third consequence concerns the utilities that no commission regulates at all. Municipal utilities and electric cooperatives are exempt from FERC rate jurisdiction under §201(f), and in many states from public utility commission rate regulation as well. Their large-load terms are therefore set by a city council or an elected member board, in a process with no formal record, no intervenor rights and no expert staff review. This is not a marginal category. Cooperatives serve much of the low-cost rural land where campuses are now siting, and municipal systems serve several of the metropolitan areas competing hardest for them.

The consequences run in both directions. A board answerable directly to the members who are also the ratepayers can weigh a cross-subsidy question with an immediacy no commission docket achieves, and municipal utilities have been among the fastest movers on flexible service — the demonstration underlying Google's demand-response contracts was run with Omaha Public Power District. But the same structure means a minimum-take provision or an exit fee can be agreed without the contested proceeding that would test it elsewhere, and a member co-op has neither the staff nor the leverage of a state commission when negotiating with a counterparty of this size. Tri-State's attempt to solve the problem once at wholesale, and its rejection, left forty-odd distribution cooperatives to answer the same question separately.

Counties and municipalities hold a further power, and it is the bluntest in the report. Land use, zoning, building permits, water and wastewater service, and local noise and air conditions are all local decisions, and a growing number of jurisdictions have imposed moratoria on data-center development outright. No energy regulator can override them. A project may hold a queue position, a signed tariff and an approved interconnection agreement and still not be built, which means the reforms in Sections 1 through 6 govern only the projects that clear a gate none of those sections describes.

How the non-RTO grid works, and where it differs

About two-thirds of U.S. electricity load sits inside an RTO or ISO; the remaining third is served by vertically integrated utilities that own their transmission, plan it themselves, and answer to a state commission rather than to a market. Outside the organized markets no independent operator runs an auction or a single regional queue. The transmission provider is usually the local utility itself, and it decides what to build through an Integrated Resource Plan — a multi-year forecast of load and resources filed with, and approved by, the state commission — rather than through an RTO process. For a large load, that changes who sits on the other side of the table and which document governs the connection.

Dimension	RTO / ISO	Non-RTO transmission provider
Grid operation	Independent regional operator dispatches the system	Utility runs its own balancing authority, or joins an Energy Imbalance Market
Transmission planning	Centralized, region-wide planning	Utility-led planning through an Integrated Resource Plan, with regional coordination
Serving a large load	Interconnection under the RTO tariff and a common study queue	Bilateral service under the utility's Open Access Transmission Tariff (OATT) and state process
Building transmission	Independent, sometimes competitive, expansion process	Utility proposes and builds, subject to state regulatory approval
Wholesale supply	Competitive energy and capacity markets	Bilateral contracts; increasingly, day-ahead markets (SPP Markets+, CAISO EDAM)
Cost allocation	Regional market rules and FERC-approved formulas	Utility transmission tariff and state ratemaking
Who writes the rules	FERC on transmission; RTO stakeholder process	State commission on retail terms; utility under FERC Order 890 and NERC standards

Table 7B — How large-load service differs outside the organized markets. The contrast is becoming less absolute as non-RTO utilities join energy-imbalance and day-ahead markets, but transmission planning and ownership remain with the utility.

The providers are among the largest utilities in the country. The Southeast has no RTO at all: service there runs through Duke Energy, Southern Company's operating utilities, the Tennessee Valley Authority, Dominion's Carolina system, and Florida Power & Light. Most of the West outside California is served by the Bonneville and Western Area Power Administrations, NV Energy, Arizona Public Service, Salt River Project, PacifiCorp, and Idaho Power. In place of an RTO planning process, these utilities coordinate through regional reliability and planning bodies — SERTP in the Southeast; NorthernGrid, WestConnect, and the Northern Tier Transmission Group in the West; and WECC's committees across the interconnection — which coordinate reliability and expansion without operating a market.

Region	Major transmission providers	Regional planning coordination
Southeast (no RTO)	Duke Energy; Southern Company (Georgia, Alabama, Mississippi Power); Dominion (Carolinas); Florida Power & Light; Tennessee Valley Authority; Santee Cooper	SERTP
West, outside CAISO	Bonneville Power Administration; Western Area Power Administration; NV Energy; Arizona Public Service; Salt River Project; PacifiCorp; Idaho Power; Public Service Co. of New Mexico	NorthernGrid; WestConnect; Northern Tier Transmission Group; WECC (TEPPC); CREPC
Alaska & Hawaii (islanded)	Chugach Electric and Golden Valley Electric (AK); Hawaiian Electric (HI)	Standalone systems; no interstate market

Table 7C — The major transmission providers outside the organized markets, and the bodies through which they coordinate. Note that "outside ERCOT" is not the same as "outside a market": much of Texas outside ERCOT, and the non-RTO Midwest, is served by other RTOs (SPP, MISO) rather than by non-RTO utilities.

Sources: FERC and EIA ISO/RTO service territories; utility and regional-planning-body membership. The share of national load inside RTO regions (about two-thirds) is a FERC figure; the data-center-specific split is not tracked in any authoritative national dataset.

Three practical differences follow for a large load that lands outside a market. Its interconnection study is run by the transmission provider itself, under FERC Order 890 planning requirements for jurisdictional utilities and the applicable NERC reliability standards, rather than through an independent queue. Its cost allocation and any network upgrades are set by the utility's transmission tariff and the state ratemaking process, not by regional market rules. And the resource adequacy it relies on is assessed through the utility's Integrated Resource Plan rather than a capacity market — which is why an estimated one-third of the fast-growing pipeline, concentrated in exactly these territories, sits outside the market mechanisms Sections 1 through 6 describe. The line is blurring: many of these utilities now participate in Energy Imbalance Markets, and several are entering regional day-ahead markets such as SPP Markets+ or CAISO's Extended Day-Ahead Market. But those are energy-dispatch arrangements; planning, transmission ownership, and the authority to approve a large-load connection stay with the utility and its state commission.

The authority vacuum is being filled by retail politics

The jurisdictional map holds nothing static: ratepayer politics redraws it in real time, and that determines which of this report's fixes get adopted. Legislators filed more than 300 data-center bills across thirty-plus states in 2026 alone, a decisive shift from courting data centers with tax incentives to constraining them with cost-causation rules. The forms vary: separate large-load rate classes (at least eighteen states), water disclosure and permitting requirements, tax-incentive rollbacks, first-of-their-kind consumption taxes (Virginia's $0.011/kWh, effective July 1, 2026), and, in a growing minority, outright construction moratoriums — New York imposed the first statewide pause in the nation — and it arrived by **executive order signed July 14, 2026**, not by the legislation this report originally anticipated. The order pauses discretionary state environmental permits for new hyperscale facilities drawing **50 MW or more** for up to one year while the Department of Public Service prepares a Generic Environmental Impact Statement and advances the Energize NY proceeding, which would require data centers either to pay more for energy or to self-supply. Maine's governor vetoed one, and dozens of localities have imposed their own. The federal Ratepayer Protection Pledge (March 4, 2026), in which seven hyperscalers voluntarily committed to cover their own generation costs, supplies the unenforceable counterweight, and it has not slowed the state activity. In mid-July 2026 the administration moved to broaden the pledge, convening utilities, developers and the governors of leading states — and it did so in the same week New York's moratorium took effect, which illustrates most sharply that voluntary federal signaling has not arrested the state-level constraining trend. Two implications follow for the rest of the report. First, political economy governs adoption of the fixes in Sections 1 through 6: “but-for” cost assignment, minimum-take tariffs, and exit fees fall to public-utility commissions under political pressure rather than to engineers, and a well-designed tariff loses all value if a moratorium freezes the market it was written for. Second, the air- and water-permitting and local-opposition gate treated as a secondary problem in Section 8 is, in political-economy terms, often the primary one: by one tally, at least 48 datacenter projects representing roughly $156 billion of investment were blocked or stalled by local opposition in 2025, a higher hit rate than the interconnection study imposes. On that evidence, community consent rather than connection capacity increasingly binds where a gigawatt can land.

Proposed solutions

- Federal floor, state ceiling. The position several state regulators have urged: FERC sets minimum interconnection standards and performance requirements; state commissions retain retail tariff regulation on top of them. On current evidence this structure appears most likely to produce a stable equilibrium, and it mirrors how NERC standards already operate across the jurisdictional line.
- NERC registration as the jurisdictional workaround. Reliability obligations attach to registered entities on the bulk power system, not to retail customers as such. If NERC's computational-load registration criteria land where they appear headed, ride-through and modeling requirements bind a 500 MW campus in Georgia and one in PJM identically — without resolving the FPA question at all. This may be the quietest answer in the report, and among the most consequential.
- Voluntary §205 filings by non-RTO transmission providers. The path FERC invited on June 18. Watch which Southeastern and Western transmission owners take it up; the ones that do would signal an interest in federal cover for terms their state commissions might otherwise contest.
- Multi-state coordination compacts. The thirteen PJM-state governors' joint statement of principles is the working prototype: states acting in concert to allocate costs to data centers and protect residential customers, which achieves harmonization without federal preemption.
- Tariff convergence pressure from the customer side. The Data Center Coalition argued in the Tri-State docket that the utility-by-utility patchwork itself stifles development. That argument serves the Coalition's interest, though the underlying observation appears sound; industry pressure for standard terms may harmonize the tariffs faster than regulatory action would.

ASSESSMENT **analytical interpretation, not a cited fact** **confidence: EMERGING**

No institution holds clean authority over the thing that most needs governing. FERC regulates the wires and the studies but not the customer; the states regulate the customer but not the interstate system whose stability that customer now affects; NERC regulates reliability, but only over entities it first brings into its registry. Each actor therefore reaches for the instrument nearest to hand rather than the one best matched to the problem, which explains why the remedies fragment. For a developer this carries direct financial consequence rather than academic interest: jurisdiction now ranks among the largest uncontrolled variables in the capital plan. A 300 MW campus that pencils comfortably under Georgia Power's framework may fail under an 85%-minimum, twelve-year AEP Ohio structure, and a state-commission docket rather than anything on the site accounts for the difference. Exposure concentrates not in the strictest jurisdictions — the market understands and prices those — but in jurisdictions mid-transition from a legacy tariff to a cost-causation framework that has not yet hardened.

8. Off-Grid and Islandable Load: Shifting Many System Impacts Rather Than Eliminating Them

The issue. Where queue access is constrained, timelines misaligned, tariff terms onerous, and jurisdiction unsettled, the developer's rational move stops asking the grid for permission and puts the generation on-site — gas engines, turbines, or fuel cells, increasingly with storage — connecting only for backup, or not at all. It offers the fastest path to power in 2026, at a premium: on-site gas energizes a campus far sooner than an interconnection, but (per Ascend Analytics, May 2026) at a levelized cost above ERCOT's forward power price across every thermal configuration. On-site generation therefore purchases speed at a cost premium rather than reducing cost. Intuition suggests that a load leaving the grid stops being the grid's problem. The 2026 evidence suggests the opposite: the on-site generation sold as insulation supplies precisely the mechanism that makes these loads dangerous to the system. Three distinct problems follow, and none of them existed at scale two years ago.

Problem 1 — Uncommanded islanding as a bulk-system contingency

To the operator, islanding is not a benign exit from the system but an instantaneous contingency it never commanded — a large block of load leaving with no notice and no coordination. That is the shape of the first problem. A campus with on-site generation and fast static-transfer switching can disconnect from the grid in under a second when its protection senses a disturbance — and does so uncommanded, to protect the compute. NERC's 2026 State of Reliability report (released July 2026, and characterized by NERC CEO Jim Robb in the Wall Street Journal's July 3, 2026 coverage as a "five-alarm fire") documents the pattern: clusters of data centers severing their grid connection and shifting to on-site backup on their own, without operator coordination. A February 2025 event dropped roughly 1,800 MW in moments; a June 2025 event about 1,300 MW; ERCOT logged nine separate cryptocurrency load losses above 100 MW across the year. When that load leaves, system frequency rises (generation now exceeds load) and voltage rises (less power is flowing), and operators must intervene to pull both back into band. This resembles the ride-through failure mode in Section 4, with an additional consequence: on-site generation converts a load that might have tripped into one designed to disconnect.

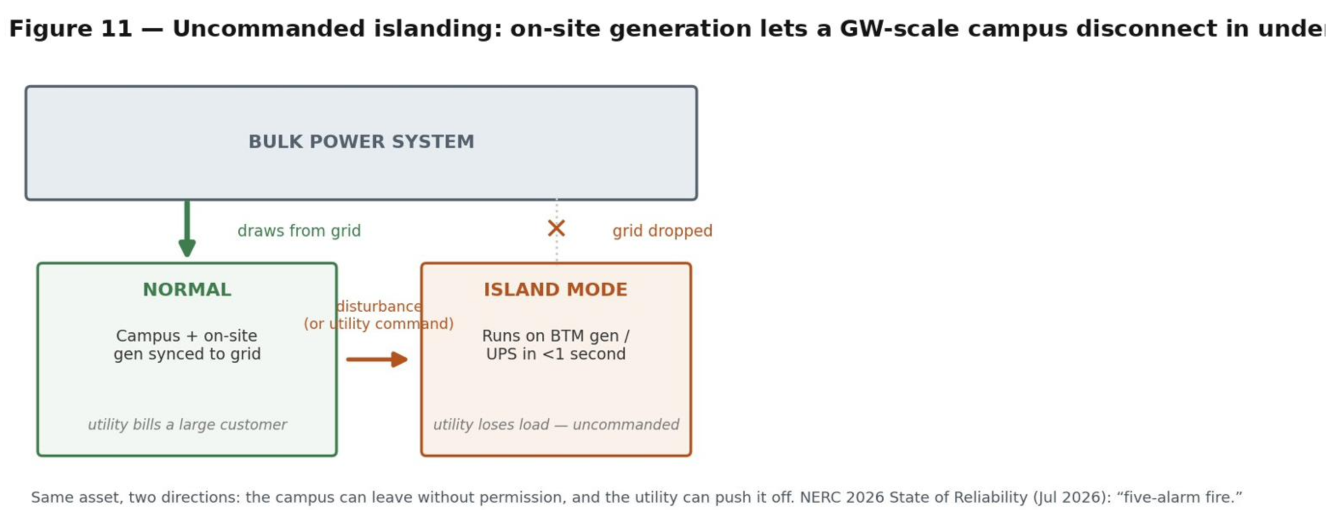


Figure 11 — The same on-site generation works in both directions. It lets the campus disconnect without operator coordination, and it gives a utility the physical basis to disconnect the campus. Either way the grid sees an instantaneous swing.

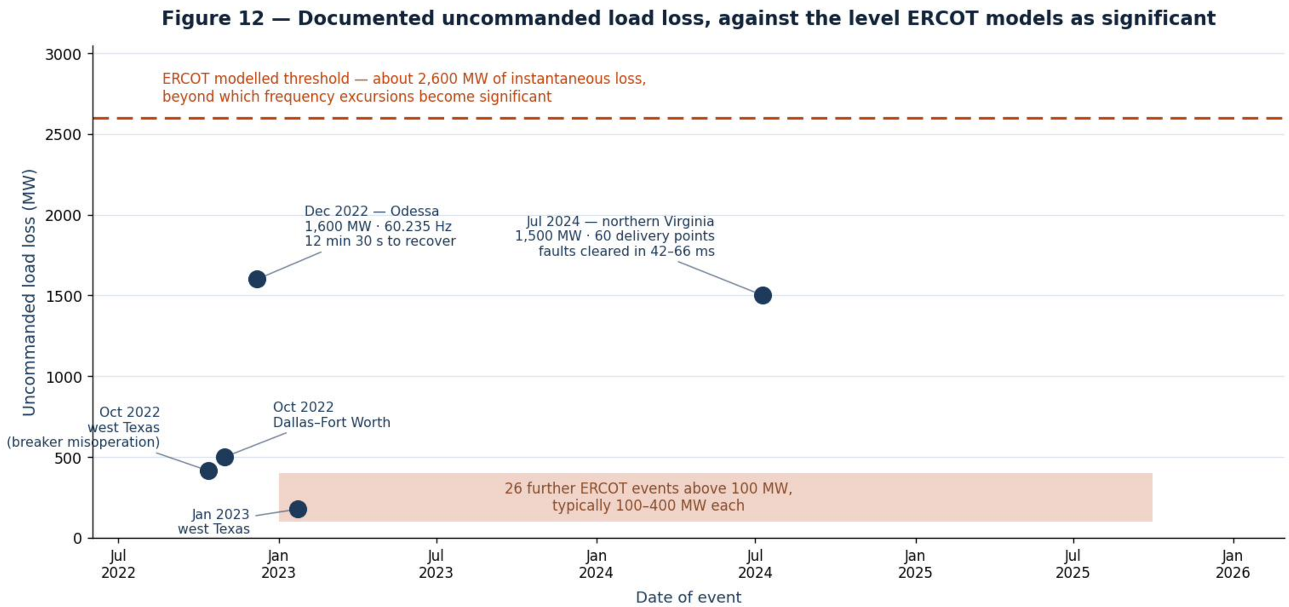


Figure 12 — Documented uncommanded load loss, against the level ERCOT models as significant. Events from the ERCOT PDCWG tabulation and NERC's incident review; the shaded band covers the 26 further ERCOT events above 100 MW recorded between January 2023 and September 2025. None of these qualify as outages in the conventional sense — no equipment failed and no storm struck; protection circuits sensed a disturbance and the loads left of their own accord. The dashed line marks about 2,600 MW of instantaneous loss beyond which ERCOT's 2030/31 study case shows frequency excursions becoming significant. Two events have already reached within a thousand megawatts of it.

Problem 2 — The stranded-cost and cross-subsidy trap

The problem here is financial rather than physical: an islandable campus can leave the utility holding infrastructure built on its behalf, with the unrecovered cost falling on everyone else. Full separation from the grid is rare in practice. Most "off-grid" campuses in fact want a grid tie — for backup when the on-site fleet is down, for startup before it is running, and for the option to sell surplus — so the utility must still build and maintain interconnection and network capacity sized for a customer that may draw almost nothing most of the year. Two kinds of not-drawing create the exposure. A campus may never materialize at the size it reserved, leaving capacity contracted against load that never arrives; or it may materialize and then island during the exact peak hours when its capacity contribution was being counted, so that the utility has planned generation and wires against a load that vanishes precisely when the system needs it most.

The exposure is not hypothetical. AEP's subsidiaries in Indiana, West Virginia and Kentucky offer the cautionary case already on the books: they acquired roughly 750 MW of generating capacity for data centers that did not materialize and were, as of late 2025, seeking to sell it back into the PJM market. Whether the load never arrives or arrives and then departs, the cost of the stranded capacity does not disappear — it is recovered from the customers who remain, who receive no service in exchange. That is the same cross-subsidy examined in Section 3, arriving through a different door: there it enters through cost allocation in the interconnection study; here it enters through the capacity a no-show or departing customer leaves behind.

Problem 3 — Air, siting, and the reliability of the “backup” itself

On-site generation does not exit regulation so much as swap one regime for another. Moving the load off the grid moves the project out of the electricity-regulation world and into the air-permitting world, where large reciprocating-engine and turbine fleets face Clean Air Act permitting, local emissions limits, and community opposition that can delay or block a campus more effectively than any interconnection study. That opposition carries real weight: by one tally it blocked or stalled at least 48 data-center projects representing roughly $156 billion of investment in 2025 (Section 7), which makes siting consent a higher-probability failure than the connection study for many projects — and it bears hardest on the gas-burning off-grid campuses this section describes. The reliability of the “backup” is the second half of the problem. A fleet of dozens of engines islanding a gigawatt of critical load has its own contingency profile, and NERC’s guidance now treats behind-the-meter resources as something planners must model explicitly rather than ignore — which cuts against the assumption that on-site generation is a self-contained solution the grid can forget about. The supply-chain constraint from Section 2 compounds all of it: on-site turbines draw on the same sold-out order book as the grid’s, so equipment scarcity, air permitting, and community opposition gate the off-grid route at once.

A distinction runs underneath these siting fights, because two accounting systems that bear on them are routinely confused. A renewable energy certificate records that one megawatt-hour of renewable generation occurred somewhere on the grid at some point in a compliance year. Retiring certificates against annual consumption supports a claim of fully renewable operation, and that claim is accurate as an accounting statement. It describes nothing about the electricity flowing into the facility in any particular hour, and nothing at all about the emissions produced on the site itself.

Air permits operate on the other basis entirely. A permit governs actual stack emissions from identified units, with limits expressed in tons per year and, commonly, in permitted run hours. No certificate offsets a permit condition, and no procurement contract changes what a reciprocating engine emits at the fence line while it runs. Local air quality is a local, hourly phenomenon; certificate matching is a national, annual one. The result is a facility that can be entirely accurate in describing itself as renewably powered while also being the reason a bank of engines runs during the hours a county measures ozone.

The same gap appears on the grid side. Certificates do not alter dispatch: the units that run in the hour the load draws are set by the operator against the system’s constraints, so a campus procuring renewable certificates in one region may still be served by a gas unit in its own. The response developing in the market is hourly rather than annual matching — time-stamped certificates, and carbon-free energy measured hour by hour against the load it actually served — which brings the accounting into contact with the physics. Regulators are separating the two questions for the same reason. A clean-energy claim is not a defence in an air-permit proceeding and does not qualify a siting objection, and this report treats them as independent constraints throughout.

Proposed solutions

- Registration regardless of grid draw. The cleanest answer: if a facility can affect bulk-system frequency or voltage, its ride-through and modeling obligations should attach whether it imports 1 GW or islands 1 GW. NERC's computational-load registration criteria (draft out for comment through May 15, 2026) are the vehicle, and they are the reason off-grid operation does not remove a facility from the scope of Section 4. **FERC's July 16, 2026 order in RD26-7-000 directs NERC to put those criteria into its Rules of Procedure by December 31, 2026**, which closes that gap on a date certain rather than on a stakeholder timetable.
- Islanding as a coordinated, not autonomous, action. NERC's May 2026 Reliability Guideline calls for interpersonal communication capability between large loads and their operators and for large loads to be subject to Operating Instructions. Applied to islandable campuses, this converts an uncommanded exit into a scheduled, visible transition the operator can plan around.
- Standby and exit-fee tariff design. Standby rates and minimum-take terms (Section 3 and Section 7) that make a grid-tied-but-islanding customer pay for the capacity held in reserve on its behalf, so the option to leave is priced rather than free.

- Behind-the-meter resources modeled explicitly. The Reliability Guideline's call for resource-adequacy models that represent firm versus flexible load and behind-the-meter generation — so that the plan carries an islandable gigawatt as the swing resource it represents rather than as a firm subtraction from demand.
- The regulatory backstop. On-site generation also gives a utility the legal basis to require a campus to island during emergencies — converting the islanding capability into a demand-response asset if the terms are written that way. ERCOT's WLPUN constraint (Section 5) offers an early version: unresolved imbalance within one minute lets ERCOT limit or suspend the arrangement.

ASSESSMENT **analytical interpretation, not a cited fact** **confidence: EMERGING**

Off-grid development answers Sections 1 through 7: when the grid moves too slowly, costs too much, and tangles legally, capital routes around it. That routing confers only partial independence. A gigawatt of islandable load at the end of a transmission line remains a bulk-system element whether or not it draws power, because each time it connects, disconnects, or oscillates it moves frequency and voltage for everyone else. The 2025 event record supports this reading, and it explains why the NERC track — registration, modeling, coordinated islanding — may matter more than the tariff work: registration alone reaches a load that has stopped asking the grid for anything. A developer treating island mode as a private engineering detail holds, in NERC's framing, an unpriced regulatory variable. Pricing should follow, and current evidence suggests fully off-grid campuses will remain the exception — attractive for the first gigawatt of speed, but carrying emissions, capital, and now reliability-compliance costs that a well-designed non-firm grid product (Section 6) can avoid.

9. The Long View: A 2018–2035 Timeline of Actions and Decisions

Why a long horizon matters. The preceding sections are a snapshot of a system in acute stress in mid-2026. But none of it began in 2026, and none of it ends there. The large-load problem is the collision of two multi-year processes — a demand curve that bent sharply upward around 2022–2023, and a regulatory apparatus that moves in three-to-ten-year rulemaking and construction cycles. Placing the two on one axis allows a judgment about whether the 2026 flurry of orders marks a turning point or merely the moment the backlog became visible. This section places the actual decisions of FERC, NERC, the six jurisdictional ISOs and RTOs — PJM, MISO, SPP, CAISO, NYISO and ISO-NE — together with ERCOT and the PUCT, the DOE and the states, on a single timeline from 2018 through 2035, and marks the proposed and projected milestones due over the next decade.

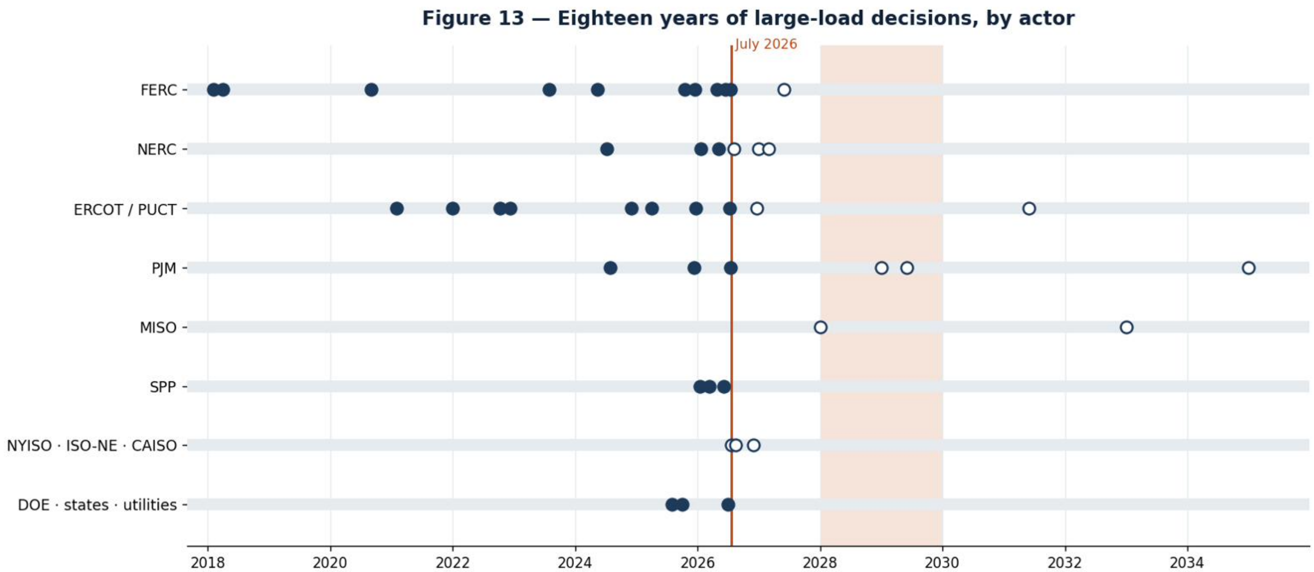


Figure 13 — Eighteen years of large-load decisions, by actor. Each row is one decision-maker; solid markers record actions that have occurred, hollow markers scheduled compliance dates, proposals, and projected onsets. The vertical rule marks July 2026 — everything to its right forecasts rather than commits. The shaded band covers 2028–2029, the years NERC's assessment marks as the onset of elevated shortfall risk in MISO and PJM, which fall before most of the remedial standards become enforceable and well before the extra-high-voltage transmission is energised.

The 2025–2027 window carries most of the action and is hard to read at full scale, so Figure 14 expands just that stretch — the same events, labelled individually, with room to separate them.

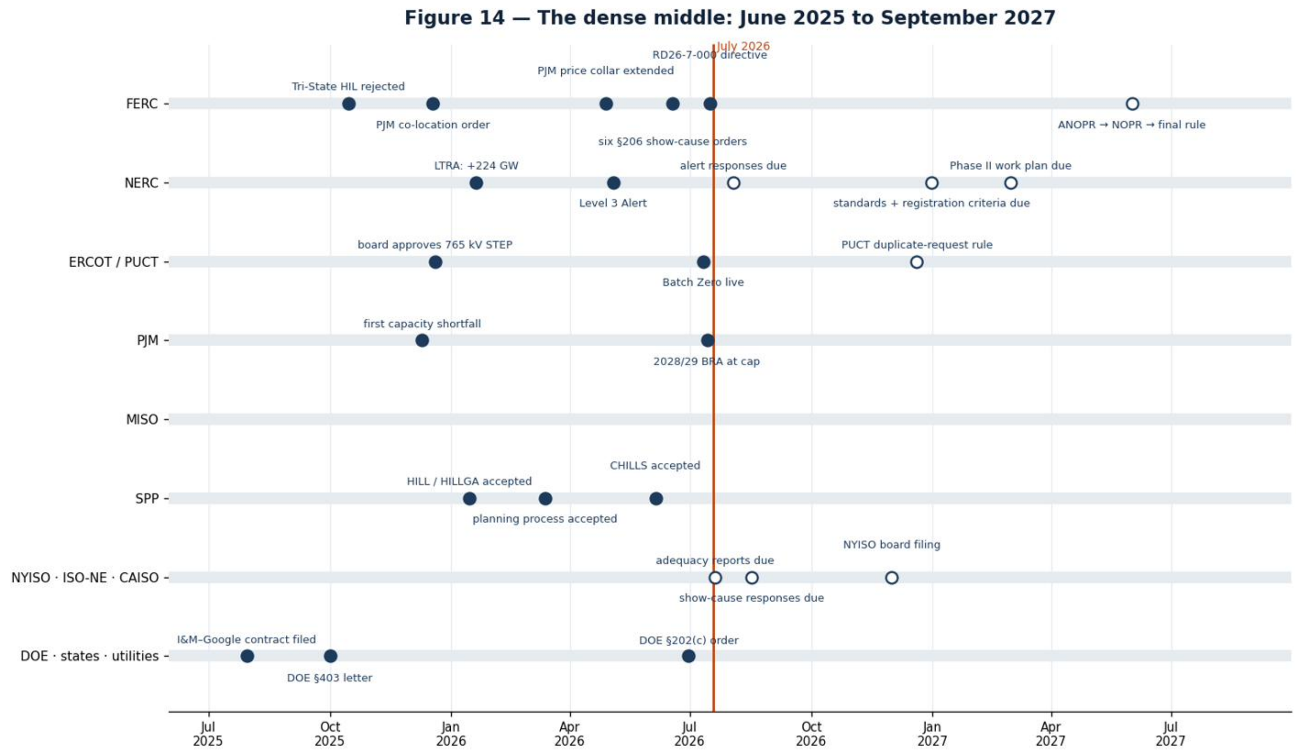


Figure 14 — The dense middle, June 2025 to September 2027. The same events at expanded scale, with every marker labelled. The concentration of completed actions to the left of the July 2026 line, set against the scheduled and proposed items to its right, states the section's argument visually. The actors front-loaded the response, and the binding compliance dates all fall after it: the August 17 show-cause responses, the December 31, 2026 NERC standards and computational-load registration criteria, the PUCT duplicate-request rule, and NYISO's own board filing.

2018–2023: the foundation laid for a different problem

The rules that govern large-load integration today were mostly written to solve the opposite problem — connecting supply, not demand. FERC's Order No. 841 (February 2018) opened wholesale markets to electric storage; Order No. 845 (April 2018) reformed the generator interconnection process; Order No. 2222 (September 2020) brought distributed energy resource aggregations into the markets. The generation-side queue reform that everyone now cites as the template for load, Order No. 2023, issued July 28, 2023 and took effect that November — mandating cluster studies, readiness deposits, and “first-ready, first-served” processing. In parallel, the shock of Winter Storm Uri (February 2021) reset ERCOT's and NERC's cold-weather and reliability agendas. The demand signal was already building underneath all of this: ERCOT stood up its Large Flexible Load Task Force in 2022 to cope with a cryptocurrency-mining boom that was the leading edge of the load wave now dominated by AI.

2024–2026: the year and a half when load became the story

The pivot is visible in the density of the timeline. FERC issued Order No. 1920 (May 13, 2024) on long-term transmission planning and Order No. 1977 on siting. NERC's July 10, 2024 Northern Virginia event — about 1,500 MW lost to a fault the load rode through badly — became the archetype the entire 2026 reliability program is built around. Then the cascade, in four registers. Markets moved first: PJM's capacity price spike in the July 2024 auction, and its first-ever capacity shortfall in the December 2025 auction. Planners followed: ERCOT's 2024 Regional Transmission Plan (December 2024) with its dual 345 kV and 765 kV builds, and NERC's Long-Term Reliability Assessment raising the ten-year peak forecast by 224 GW (January 2026). Then the federal instruments: DOE's Section 403 letter to FERC (October 2025), FERC's PJM co-location order (December 18, 2025), FERC's six show-cause orders (June 18, 2026), and DOE's Section 202(c) emergency backup-generation order (June 30, 2026). And finally the operating rules: NERC's Level 3 Alert and Reliability Guideline (May 2026), and ERCOT's Batch Zero going live (July 11, 2026). Eighteen months carry more consequential large-load actions than the prior six years combined.

2027–2035: what is scheduled, proposed, and feared

The right half of the timeline necessarily softens, but it holds real content. Several items are concrete compliance targets: PUCT's duplicate-request rule falls due by December 2026; NERC's Project 2026-02 should produce a “bridge” Reliability Standard by the end of 2026, with broader standards drafted from 2027 and a computational-load registry following. As of FERC's July 16, 2026 order in RD26-7-000 those NERC dates no longer represent a schedule NERC set for itself: the standards and the registration criteria are both due December 31, 2026, and a Phase II work plan for further standards is due March 1, 2027. On the demand-and-market side, PJM's second consecutive capacity shortfall (2028/29 BRA, July 14, 2026) is the first hard market evidence that the adequacy gap the right half of this timeline anticipates is materialising. FERC's ANOPR is widely expected to mature into a NOPR and then a final rule across 2027, with co-location and flexible-load service terms taking effect thereafter. Others are projections rather than plans, and they are where the risk concentrates: NERC's 2026 Long-Term Reliability Assessment places PJM in a high reliability-risk category beginning around 2029 and MISO from roughly 2028, as demand outruns resource additions. The long-lead physical builds anchor the far end — ERCOT's 765 kV backbone is a late-decade-into-2030s project, and PJM's and MISO's own forecasts run to summer peaks near 210 GW and demand additions of 18 GW of data-center load, respectively, by 2035. Above all the timeline shows a timing inversion: the reliability-risk windows (2028–2029) arrive before most of the remedial standards are enforceable and long before the major transmission is energized (early 2030s). The regulatory timeline trails the physical one, and on the current schedule the physical constraints arrive first. The inversion is more usefully described as four clocks running at different speeds than as a single race between “policy” and “physics.” Load is the fast clock (Section 2's eighteen-months-to-three-years); equipment, wires, and the rules themselves are the slow ones, the last a NOPR-to-enforceable-standard cycle measured in years. The report's thesis is simply that the first clock now outruns the other three. What makes the gap durable is that two exogenous forces set the speed of the slow clocks, and neither is a grid-engineering problem the sections above can solve.

Figure 15 — Four clocks, not three: load arrives fast; equipment, wires, and rules run slow — and two forces set that speed

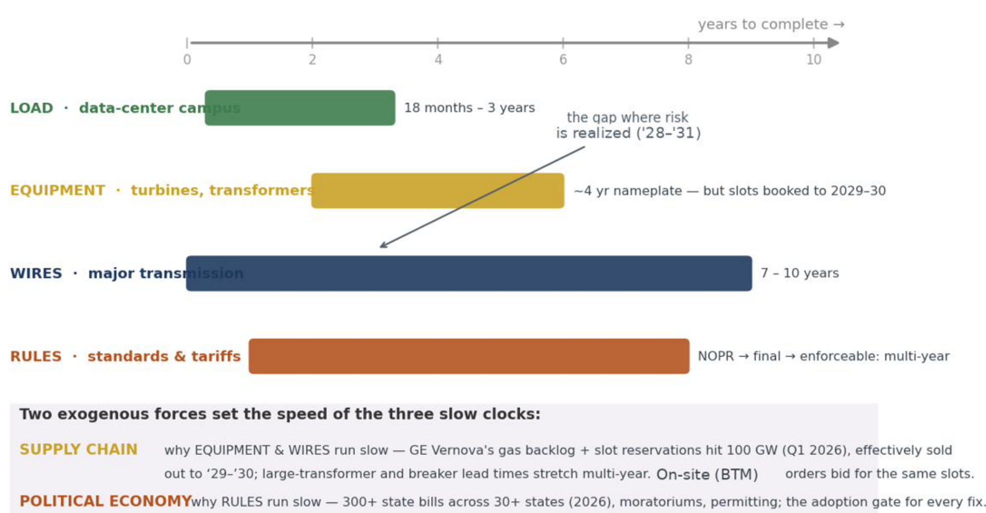


Figure 15 — The timing inversion as four clocks. Load moves fastest; equipment, wires, and rules move slowly, and supply chain (equipment and wires) and political economy (rules) set the pace of the three slower clocks. The stress lands in the gap, roughly 2028–2031.

The first force is supply chain, which sets the equipment and wires clocks. On the order-book evidence in Section 2, gas-turbine manufacturing rather than developer demand appears to be the binding constraint — Section 2 details the backlog that has booked slots to the end of the decade, which suggests the “four years for generation” figure is optimistic. Two consequences matter for the timeline. First, the on-site alternatives do not avoid the bottleneck: a behind-the-meter turbine order (Sections 5 and 8) competes for the same scarce slot as a utility's grid-serving order, so on-site generation bids against the grid rather than stepping outside it, and may widen the adequacy gap it was intended to address. Second, it explains why flexible load (Section 6) outranks every other near-term lever: it adds no equipment to the queue, so scarcity only raises its value.

The second force is political economy, which sets the rules clock. Section 7 details the wave of state legislation — 300-plus bills, separate rate classes, and a growing minority of moratoriums — that is writing the jurisdictional map in real time. It bears on the timeline directly: political economy governs adoption of nearly every fix in Sections 1–6, because “but-for” cost assignment, minimum-take tariffs, and exit fees are set by public-utility commissions under political pressure rather than by engineers, and a moratorium can freeze a market no matter how well its tariff is written. The rules clock moves slowly not because the drafting is difficult but because the politics are — and the permitting and local-opposition gate (Sections 7 and 8) can stop a project before any tariff applies.

Both forces feed back into the Section 1 thesis rather than sitting beside it. The interconnection queue reflects optionality as much as demand, and scarce turbine slots produce the same behavior, GE Vernova's 56 GW of slot reservations represent the equipment-side analogue of speculative queue entries — positions held for option value that may never convert. Divergent state tariffs and moratoriums push developers to arbitrage across jurisdictions exactly as they arbitrage across points of interconnection.

Neither supply chain nor political economy constitutes a ninth or tenth problem. The report identifies eight, and these are not additions to that list. They are two further venues in which the same speculation problem plays out, and they compound one another: a turbine backlog and community opposition jointly close the off-grid door, while ratepayer politics and long-lead transformers jointly sharpen the stranded-cost fear behind the AEP example in Section 3. The practical consequence for a reader is that neither force can be delegated. A remedy for any of the eight problems that assumes equipment arrives on schedule, or that a commission will approve the tariff supporting it, has not been assessed against the two conditions most likely to defeat it.

ASSESSMENT **analytical interpretation, not a cited fact** **confidence: EMERGING**

Two caveats govern the chart. First, everything right of the July 2026 line remains contingent: rulemakings slip, standards get remanded, auctions and forecasts change annually, and a single court decision on co-location or jurisdiction (Sections 5 and 7) could reorder the entire right half. Read those dates as each actor's current intent rather than as a schedule anyone has committed to. Second, the timeline mixes categories by design — binding rules, voluntary guidelines, market outcomes, and risk forecasts — because developers and planners navigate all of them simultaneously. Read that way, the chart offers no reassurance of timely repair; it warns instead that remedy and risk run on different clocks, and that the interval between them, roughly 2028 through 2031, will carry the operational stress documented in Sections 2 and 4.

International Experience: How Other Systems Are Managing Large-Load Growth

The United States is not the first system to have a large-load surge collide with a slower grid, and several jurisdictions that hit the wall earlier offer a preview of where U.S. policy may head. Three European cases are especially instructive because each chose a different instrument — a connection moratorium tied to self-supply, a congestion-driven flexible-connection regime, and an efficiency-gated capacity release — and each has now lived with the consequences long enough to judge. Three further cases follow them: Australia, which is writing ride-through obligations before experiencing the events that prompted them; China, the one large system directing siting rather than pricing it; and the wider Asia-Pacific, where access to a constrained grid has become a rationed good.

Ireland — the moratorium that became a self-supply mandate

Ireland is the clearest cautionary tale, because data centers reached a scale relative to the system that the U.S. as a whole has not. They consumed roughly 22% of Ireland's metered electricity in 2024, with EirGrid's median scenario projecting 31% by 2034 — Dublin is Europe's second-largest data-center cluster at about 1,150 MW. Facing supplyshortfall risk in the greater Dublin area, the regulator (CRU) imposed a de facto moratorium on new grid connections in November 2021. In December 2025 it replaced the moratorium with a Large Energy User Connection Policy that does not simply reopen the door: new data centers above 10 MVA must provide on-site flexible generation or storage sized to 100% of their grid connection, site in “unconstrained” locations away from congested nodes, and match 80% of annual demand with new Irish renewables within six years. In parallel, EirGrid moved on April 1, 2026 to impose fault-ride-through obligations on large demand facilities, with a two-year derogation for existing sites — the same ride-through discipline Section 4 describes, arriving in Ireland on essentially the same schedule as ERCOT's NOGRR282. Notably, Irish government policy explicitly rejects fully islanded, fossil-powered data centers as out of line with climate goals, closing the very off-grid pathway Section 8 examines.

The Netherlands — congestion, and the right to a flexible connection

The Dutch grid suffers the worst congestion in Western Europe, and the response has redefined what a grid connection means. TenneT has declared effectively zero headroom for large new connections across much of NoordHolland into the mid-2030s, with connection queues exceeding 5 GW and a national grid-upgrade program measured in the hundreds of billions of euros that will not fully relieve pressure before the 2030s. Rather than queue everyone, the Netherlands has shifted from a “right to connection” toward a right to a flexible connection. Four instruments carry that shift. It prices 24/7 firm capacity as a premium product, requires large loads to participate in congestion management, and offers off-peak and non-firm connection contracts — TenneT allocated 9 GW of off-peak capacity against more than 70 GW of applications. It also contracts batteries as “congestion mitigators.” This is the clearest real-world implementation of the flexible/non-firm service that Section 6 identifies as a leading U.S. lever — born of necessity rather than choice.

Singapore — the efficiency-gated capacity release

Singapore, land- and power-constrained, imposed an outright moratorium on new data centers in 2019, then reopened deliberately. A 2022 pilot allocation came first, followed by the May 2024 Green Data Centre Roadmap. The roadmap unlocks at least 300 MW of new capacity but gates it on efficiency and green-power conditions: a PUE target below 1.3 and at least 50% green power for priority access, with faster approvals and earmarked grid allocations for the best performers. Singapore's model treats grid capacity as a scarce public resource rationed by efficiency, a different philosophy from the U.S. market-clearing approach but one increasingly echoed in the efficiency and waste-heat mandates appearing in European law.

Australia — writing the ride-through rule before the event

Australia matters to this report for a reason none of the European cases supply: it is addressing the Section 4 problem prospectively, and citing the American record as its justification. In March 2026 the Australian Energy Market Commission published a draft rule creating technical access standards for large inverter-based loads, raising the threshold that triggers them from 5 MW to 30 MW and requiring covered facilities to remain connected through voltage and frequency disturbances and to restore demand in a controlled manner afterwards. Consultation closed on 7 May 2026 with a final rule expected mid-year. The Commission's stated reasoning cites the July 2024 Virginia event by name — sixty data centres removing about 1,500 MW during a single fault — alongside incidents in Ireland and Texas.

The queue evidence is equally familiar. AEMO disclosed a transmission connection pipeline for the first time in its March 2026 quarterly report: eleven projects above 5 MW totalling 5.4 GW of maximum demand, about 60% in New South Wales and 40% in Victoria, against a national fleet that consumed roughly 3.9 TWh in FY25, near 2% of grid-supplied electricity. New South Wales alone lists 44 projects in development totalling 11.4 GW, while industry expects about 1.2 GW to actually energise in Sydney by 2030 — close to a tenfold ratio between requested and expected load, and the term used locally for the difference is the same one this report uses. AEMO now models data centres as a separate demand category and describes them as active grid participants rather than as load, which is the modelling consequence of Section 4 stated in planning terms.

China — directing siting instead of pricing it

China is the only large system attempting to solve the problem by allocation rather than by price, which makes it the useful contrast case even though its instruments do not transfer. Installed data-centre capacity stood near 32 GW at the end of 2025 and is expected to reach about 40 GW by the end of 2026, on a path to more than 60 GW by 2030. Consumption should roughly double to around 289 TWh by 2030, taking the sector from about 1.2% of national electricity demand to 2.3%. That remains a smaller share than in the United States, on a system nearly three times the size, growing at 19% compound after 38% over the preceding five years.

The organising instrument is the East Data West Computing programme, which designates eight national computing hubs in the western provinces and ten clusters in the east, directing latency-tolerant workloads toward the west where wind and solar resources are concentrated and reserving the east for latency-sensitive services. Data-centre development is a named priority of the 15th Five-Year Plan covering 2026 to 2030. The conditions attached are binding rather than aspirational: new large facilities must achieve a power usage effectiveness below 1.25 and projects inside the national hubs below 1.2, against a national average target under 1.5, and the 2025 green data centre action plan requires new projects in the hubs to source at least 80% of their energy from renewables.

Two observations follow for a reader of this report. Directed siting dissolves the queue-integrity problem of Section 1 by removing the queue: a developer does not request interconnection at fifteen points to preserve optionality when the location is assigned. What it removes with it is the information a queue carries, since an administratively chosen site reveals nothing about where demand would have gone. And the mechanism most Chinese operators use to meet the renewable requirement is green electricity certificate procurement, which raises precisely the accounting question set out in Section 8 — an annual certificate does not establish that the facility ran on renewable power in any given hour.

Japan, Korea, Singapore, Malaysia and India — access as a rationed good

The wider Asia-Pacific pattern is consistent enough to state briefly. Japan, India and South Korea have begun steering data-centre development toward regions with stronger grid availability and lower-carbon supply, while Singapore, Malaysia and South Korea have turned to regulatory frameworks that ration access to constrained systems outright. Reporting from the region also indicates grid-support obligations moving toward the Australian model, including ride-through duties on facilities above 30 MW. Where geography constrains the network and no unconstrained alternative region exists, these systems converge on administrative allocation rather than queue position — the same destination as the European cases above, reached from a different direction.

Jurisdiction	Trigger	Instrument	The transferable lesson
Ireland	Data centers ~22% of national demand; Dublin supply risk	2021 connection moratorium → 2025 policy requiring 100% on-site generation, renewable matching, siting rules; 2026 fault-ride-through mandate	Self-supply and ride-through can be made connection prerequisites; islanded fossil sites can be barred
Netherlands	Severe transmission congestion; 5+ GW queues	Shift from firm “right to connect” to flexible/non-firm connections; mandatory congestion management; battery mitigators	Non-firm service (§6) works at scale when firm capacity is genuinely unavailable
Singapore	Land and power scarcity in a city-state	2019 moratorium → 2024 roadmap releasing ~300 MW gated on PUE < 1.3 and ≥ 50% green power	Capacity can be rationed by efficiency; performance becomes an entry requirement
Australia	A 5.4 GW transmission connection queue against a national fleet consuming about 2% of demand; NSW pipeline of 11.4 GW against roughly 1.2 GW industry expects to energise in Sydney by 2030.	AEMC draft rule (March 2026) creating technical access standards for large inverter-based loads, raising the threshold from 5 MW to 30 MW and requiring facilities to ride through voltage and frequency disturbances and restore demand in a controlled manner. Final rule expected mid-2026.	Ride-through obligations can be written before the events occur. AEMO reclassified data centres as active grid participants and models them as a separate demand category rather than as ordinary load.
China	Roughly 32 GW installed at end-2025 and about 40 GW expected by end-2026, with data centres at 1.2% of national demand but growing at a 38% compound rate over the past five years.	East Data West Computing: eight state-designated computing hubs in the west and ten clusters in the east, with siting directed toward renewable resources. Binding efficiency mandates — PUE below 1.25 for new large facilities and 1.2 inside the national hubs — and a requirement that new projects in the hubs source at least 80% renewable energy.	The one system directing siting rather than pricing it. Central allocation removes the queue-integrity problem by removing the queue, at the cost of the price signals that reveal where demand actually wants to be.
Japan, Korea, Singapore, Malaysia, India	Concentrated demand meeting constrained transmission in dense urban regions.	Regionally directed development toward areas with stronger grid availability and lower-carbon supply, and regulatory frameworks rationing access to constrained systems.	Where geography constrains the grid, jurisdictions converge on rationing access by administrative allocation rather than by queue position.

Table I1 — What other systems did, and what transfers. Five jurisdictions reached the same constraint earlier and answered it differently. The final column is the point: the instruments travel, the institutional conditions that made them work often do not.

Sources: CRU (Ireland); ACM and TenneT (Netherlands); IMDA and EMA (Singapore); AEMC and AEMO (Australia); Chinese government programme documents.

Every one of these systems ended up somewhere the U.S. debate now heads: making firm grid access conditional — on self-supply (Ireland), on flexibility (Netherlands), or on efficiency (Singapore) — rather than an entitlement. The U.S. advantages are its scale, its spare land, and its deep capacity markets; its disadvantage is the fragmented authority of Section 7, which makes a coherent national conditionalaccess policy far harder to write than in a single-regulator system. The lesson is less any specific rule than the direction of travel: as large-load growth outruns the grid, “connect and we will build for you” gives way everywhere to “connect on conditions,” and the systems that wrote those conditions early suffered less disruption than those that improvised late.

The comparison in numbers

The instruments described above are responses to very different magnitudes, and the table below sets the six markets on comparable terms. Two columns deserve caution. Installed capacity and consumption are reported on different bases by different authorities — metered consumption in Ireland, contracted capacity in Australia, installed gigawatts in China — so the rows should be read down rather than across. And the final column is forecast rather than record: it is included because policy in each jurisdiction answers the projection rather than the present, but every figure in it carries the fivefold uncertainty that Section 1 describes.

Market	Data-centre consumption, latest reported	Share of national electricity	Fleet and pipeline	Forecast (speculative)
World	415 TWh (2024, IEA)	About 1.5%	Roughly half of consumption in the United States, a quarter in China, 15% in Europe.	IEA projects about 945 TWh by 2030. The US and China account for close to 80% of the growth.
United States	176 TWh (2023, LBNL)	About 4.4%	Northern Virginia alone is the largest market in the world; data centres there consume roughly 26% of state electricity.	LBNL's 2025 Update puts 2030 at 649 TWh in the reference case, 578–664 TWh across sensitivities, and 11.8% of national electricity within a 9.5–15.3% band. The IEA puts the same year at 426 TWh (Figure D2).
China	Consumption roughly doubling to 2030	About 1.2%, rising toward 2.3%	~32 GW installed at end-2025; ~40 GW expected end-2026; more than 60 GW planned by 2030, sited through the East Data West Computing hubs.	289 TWh by 2030 (Rystad). Growth of about 19% a year, after 38% a year over the previous five.
Ireland	7,663 GWh (2025, CSO)	23%, up from 5% in 2015	More than 80 facilities, clustered in greater Dublin, where the local share reaches roughly 79%.	The IEA has projected around a third of national consumption; the CRU's Large Energy Users policy now conditions new connections on matching generation.
Australia	3.9 TWh (FY25)	About 2% of NEM consumption	More than 250 facilities. 11 projects totalling 5.4 GW in the transmission connection queue; New South Wales lists 44 projects at 11.4 GW against roughly 1.2 GW industry expects to energise in Sydney by 2030.	12 TWh by FY30 (about 6%) and 34.5 TWh by FY50 (about 12%) under AEMO's Step Change scenario.
Singapore and southern Malaysia	Not separately reported here	Constrained by policy rather than by demand	A regional hub operating under an efficiency-gated capacity release after the 2019 moratorium.	IEA expects South-East Asian data-centre demand to more than double by 2030.

Table I2 — Data-centre development by market. Latest reported consumption, share of national electricity, fleet and pipeline, and forecast. Forecast figures are projections by the cited bodies, not commitments.

Sources: IEA, *Energy and AI*; LBNL, 2024 US data-centre report and 2025 Update; Rystad; Ireland Central Statistics Office; AEMO; AEMC.

Read down the share column and the governing pattern appears. The instrument each jurisdiction reaches for tracks the share of its own system that data centres occupy, not the absolute size of the fleet. Ireland at 23% imposed a moratorium and then a self-supply condition. Australia at about 2% is writing technical performance standards. The United States at roughly 4% is arguing about tariffs and cost allocation. China at 1.2% is directing siting for industrial-policy reasons rather than reliability ones. The concentration figures explain why: national shares understate the problem everywhere, since 26% in Virginia and 79% in Dublin are the numbers the operators of those systems actually face. A jurisdiction acts when the local share becomes unmanageable, whatever the national average shows — which is the same conclusion Table D1 reaches for the United States, arrived at from six directions.

Cross-Cutting View: Where Each Institution Stands

The matrix runs across two tables because eight columns will not fit legibly on one page. The first covers FERC, NERC and the three regions with the most developed large-load programmes; the second covers the remaining four jurisdictional operators, all of which are now working to the same August 17 deadline. The split is presentational, not analytical — since June 18 all six regions have been answering the same set of questions.

The same eight problems, read across the institutions. The first table covers FERC, NERC, ERCOT, PJM and MISO; the second covers SPP, NYISO, ISO-NE and CAISO. No institution leads on all six. ERCOT — the one grid outside FERC's jurisdiction — leads on the operational rules and trails on cost allocation, while PJM inverts that pattern.

Issue	FERC	NERC	ERCOT	PJM	MISO
1. Queue / speculative load	Category 1 showcause; readiness screens	Registration criteria (indirect)	Batch Zero (in force 7/11/26); $50k/MW; SB6 gating	CIFP-LLA loadforecast reform	Queue cap; largeload study reform
2. Resource adequacy gap	RA reports due 7/20/26	Reserve sizing for clustered load loss	BYOG; ~23 GW added 2024–25	Backstop procurement (capped $555/MW-day); EIT	ERAS (~27 GW, 90day GIA); MTEP EPR
3. Cost allocation	Category 2; anticost-shift mechanisms	Out of scope	$50k/MW security; state-led	“But-for” assignment; loadbilled backstop	ERAS upgrade costs to the customer
4. Ride-through / dynamics	Defers to NERC	Level 3 Alert (5/4/26); Project 2026-02 standard by YE2026	NOGRR282 / NPRR1308 — ahead of NERC	Modeling and telemetry work	Load modeling and telemetry reform
5. Co-location / BTM	Dec 2025 §206 order; rehearing 6/18/26; Categories 3 & 5	Guideline-level only	WLPUN / PCLR; Form X due 7/10/26	Firm & Non-Firm Contract Demand; CIR adjustment	Show-cause response pending
6. Flexible load	Category 4	Flexibility credited in 2026 guideline	Controllable Load Resource; ramp-rate rules	Non-firm service; DR expansion	“Speed-to-reliablepower” workstream
7. Jurisdiction / non-RTO	Transmission-only baseline; §205 invitation to nonRTO TOs	Registration binds loads across the jurisdictional line	Outside FERC jurisdiction; PUCT is the sole authority	13-state governors’ principles; state tariffs (Ohio, Virginia, Pennsylvania)	Multi-state footprint; tariffs set state by state
8. Off-grid / islandable	Categories 3–4 reach BTM and flexible load service	Registration + coordinated islanding; BTM modeled explicitly	WLPUN one-minute leash; ramp-rate rules	Standby/exit-fee tariffs; stranded-cost exposure (AEP)	BTM resources in resource-adequacy modeling

Table X1 — The eight problems against the five bodies furthest along. Reading down a column gives an institution's whole agenda; reading across a row shows how differently the same problem is being answered. The blank and thin cells are as informative as the full ones.

Sources: FERC; NERC; ERCOT; PJM; MISO, from the filings listed in the Catalog.

Issue	SPP	NYISO	ISO-NE	CAISO
1. Queue / speculative load	HILL and HILLGA study processes in force since 15 Jan 2026 — the most developed gating of the six.	Load interconnection procedures above 10 MW at 115 kV or higher, or 80 MW below; own reform to a Dec 2026 board filing.	No express large-load provisions; FERC recorded a less urgent concern. ISO-NE and the New England transmission owners have said they will seek a 90-day abeyance.	No traditional Order 888 service, so the study framework differs structurally.
2. Resource adequacy gap	CHILLS conditions long-term service on curtailability rather than on new supply.	Generation-adequacy informational report due 20 Jul 2026.	Same informational report; existing capacity auction unchanged for large load.	Jul 20 report: no systemic adequacy shortfall, attributed to integrated CEC/CPUC/CAISO planning rather than to tariff terms. Adequacy runs through the CPUC programme.
3. Cost allocation	HILL allocates study and upgrade cost to the requesting load.	FERC flagged cost shifting through the Transmission Service Charge when speculative requests drive local planning.	Not separately addressed; falls within the general show-cause findings.	Not separately addressed; falls within the general show-cause findings.
4. Ride-through / dynamics	Defers to NERC.	Defers to NERC.	Defers to NERC.	Defers to NERC; inverter-based resource experience is the most extensive of the four.
5. Co-location / BTM	Addressed through the conditional-service framework rather than a separate co-location rule.	FERC found no co-location terms in the tariff — a named deficiency.	No express provisions.	No express provisions; behind-the-meter generation is prevalent but governed by state rules.
6. Flexible load	CHILLS, accepted 5 Jun 2026 — the closest instrument in force to a bounded curtailment product.	Special Case Resources, an established aggregator programme built for smaller commercial load.	Active Demand Capacity Resources, likewise aggregator-mediated.	Proxy Demand Resource and the Demand Response Auction Mechanism.
7. Jurisdiction / non-RTO	Multi-state footprint; terms set state by state below the tariff.	Single-state footprint; the NY Department of Public Service runs a parallel state initiative.	Six-state footprint with six commissions and no single state counterpart.	Largely single-state; the CPUC holds the retail lane and the state has its own large-load proceedings.
8. Off-grid / islandable	Telemetry and curtailability conditions apply to conditional service.	Show-cause response pending.	Show-cause response pending.	Show-cause response pending.

Table X2 — The remaining four jurisdictional operators. Read alongside the table above. The pattern across both is that depth of programme tracks how early a region began rather than how much load it faces: SPP, with a fraction of PJM's data-center interest, has the most complete set of instruments in force, while ISO-NE and CAISO are largely answering the show-cause orders from a standing start.

Sources: FERC show-cause orders of June 18, 2026, and the tariff materials of SPP, NYISO, ISO-NE and CAISO.

Three observations hold across the twelve columns. Ride-through is the one row where every operator except ERCOT defers to NERC, which is why the December 31, 2026 filing carries the weight it does — no region is building an alternative. Flexible load is the row with the most activity and the least standardisation: five different instruments, none interoperable, and the two that reach gigawatt-scale customers (SPP's CHILLS and PJM's Non-Firm Contract Demand) postdate the others by a decade. And cost allocation is the row where the RTOs have the least room, because the terms that decide it are retail terms held by the states (Section 7), which is why the FERC column is fuller than any regional one.

The near-term watch list that appeared here has been consolidated into ***Upcoming Dates and Events: A 12–18 Month Watchlist****, which now sits before Section 1 and extends the horizon to mid-2028.*

Synthesis: The Eight Issues as One Planning Problem

Read separately, the eight issues look like eight problems for eight committees. Read together, they present a single problem in eight forms. Every section of this report traces back to a single fact: a new class of load — large, fast, concentrated, mobile, and self-protecting — has arrived on a grid built for demand that was small, slow, diffuse, captive, and passive. Each issue marks where one of those properties collides with one of those assumptions, and the remedies stop looking fragmented once they are mapped back to the property each is actually answering.

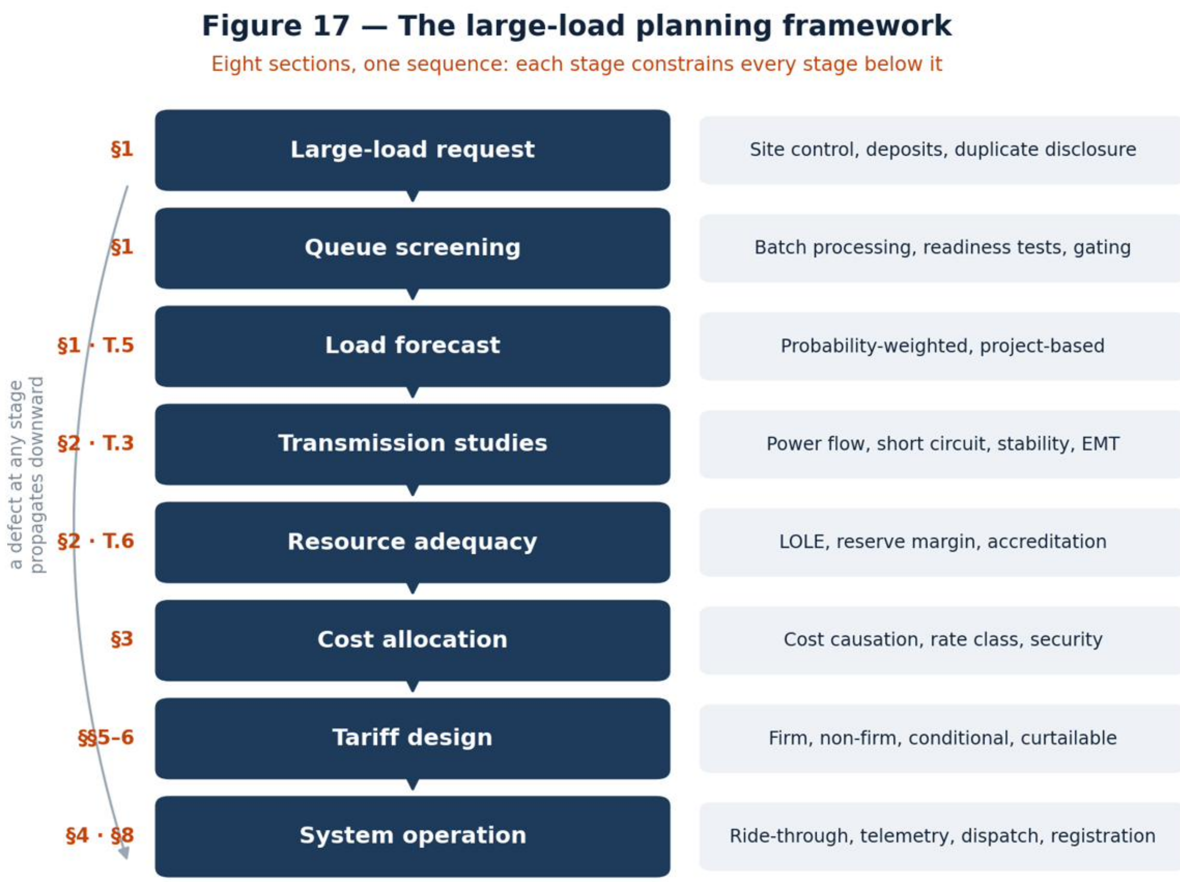


Figure 17 — The large-load planning framework. One or more sections of this report treat each stage, and each constrains every stage below it: a queue that records optionality rather than demand corrupts the forecast, which corrupts the studies, which corrupt the adequacy assessment and the cost allocation built on it. The figure states the report's organising claim: the eight issues mark stages of one planning process rather than eight independent problems.

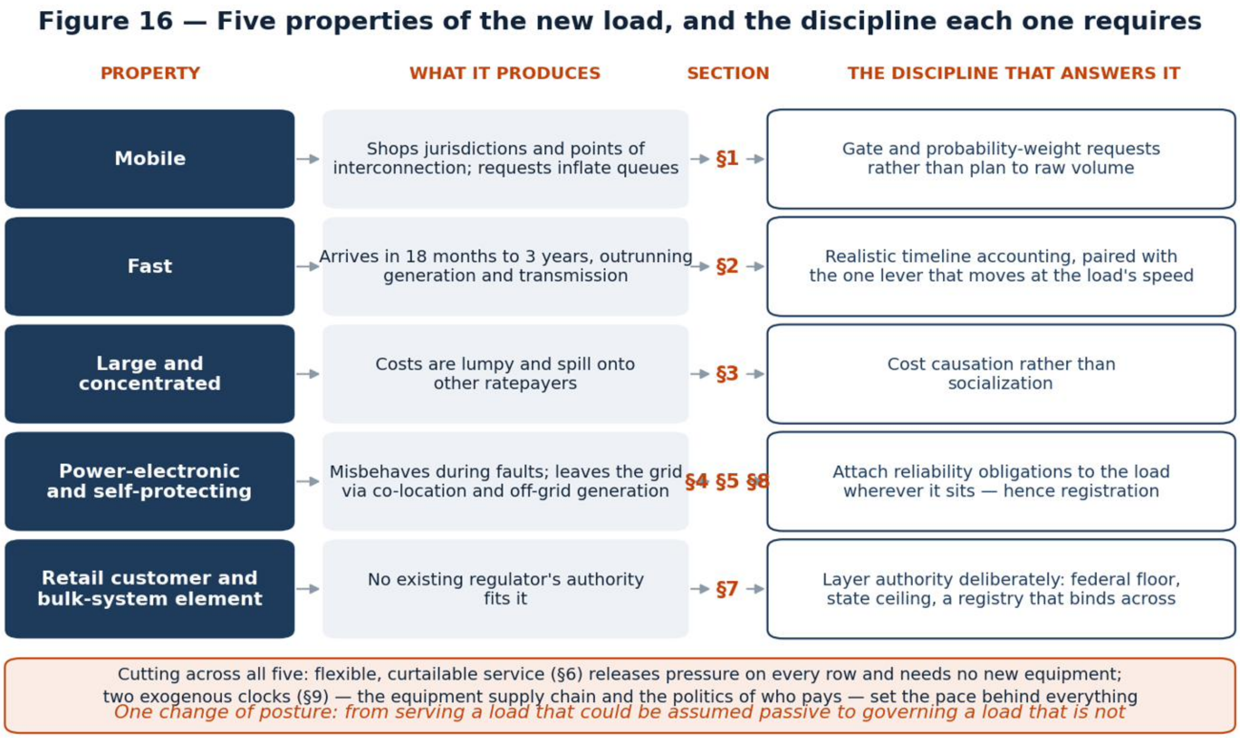


Figure 16 — Five properties of the new load, and the discipline each one requires. Each row reads left to right: a defining property of computational load, what that property produces, the section that treats it, and the discipline that answers it. The banner beneath covers the two elements that cut across every row — flexible service, which releases pressure everywhere and needs no new equipment, and the two exogenous clocks that set the pace behind all of it.

The framework reads across five rows. Because the load moves, it shops jurisdictions and points of interconnection, which inflates the queue (§1) and invites regulatory arbitrage (§7); the discipline gates and probability-weights requests rather than planning to raw volume. Because it arrives fast, it outruns generation and transmission (§2); the discipline pairs realistic timeline accounting with the one lever moving at the load's own speed. Its size and geographic concentration make its costs lumpy and spill them onto other ratepayers (§3); cost-causation rather than socialization supplies the discipline. Power electronics and self-protection make it misbehave during faults (§4) and tries to leave the grid altogether through co-location (§5) and off-grid generation (§8); the discipline that reaches all three attaches reliability obligations to the load wherever it sits — which is why NERC registration, rather than any tariff, carries much of the weight. And because the load acts at once as a retail customer and as a bulk-system element, no existing regulator's authority fits it (§7); the discipline layers authority deliberately — a federal reliability floor, state retail ceiling, and a NERC registry that binds across the line.

Two things sit outside the rows because they cut across all of them. Flexible, curtailable service (§6) releases pressure across the entire system: it alone moves fast enough to match the load, and it alone needs no new equipment, which is why it recurs as a leading answer to the timeline, cost, and adequacy problems at once. And two exogenous clocks (§9) set the pace behind everything — the equipment supply chain and the politics of who pays — neither of which a grid engineer can fix, both of which intensify the speculation at the root.

The eight sections separate for exposition and not in operation, and four connections carry most of the weight. The first runs from queue integrity to everything downstream. Capacity prices, transmission plans and reserve margins all compute against forecast load, so a queue that records optionality rather than demand corrupts each of them simultaneously. Sections 2 and 3 cannot be solved while Section 1 remains unsolved. The gating measures now in force — financial security, duplicate-request disclosure, probability weighting — precede the cost-allocation debate rather than running beside it.

The second connects speed to firmness. A developer unable to obtain firm service on an acceptable timetable does not wait; it buys a different product. That impulse produces co-location (Section 5), off-grid generation (Section 8) and curtailable service (Section 6) as three responses to one constraint, and it explains why they are converging on similar terms. Each trades some claim on the system for an earlier energisation date. The distinction that matters turns on whether the operator can see the trade: a curtailment contract shows, and an islanding campus does not.

The third runs between physical behaviour and legal status. A load that disconnects during a fault (Section 4), or that leans on the grid when its host generator trips (Section 5), is a bulk-system element behaving as one — but obligations attach to registered entities, not to physical significance. This is why the December 31, 2026 registration filing recurs throughout the report as a precondition rather than a formality. Until an operator can identify the facility, it can neither bind it to a standard nor investigate what it did.

The fourth runs jurisdictional, and it constrains the other three. Every remedy above sits on one side of the line §201(b) draws — interconnection and transmission service on the federal side, retail terms and cost allocation on the state side — while the problem sits across it (Section 7). No participant holds both the decision and its consequence, so each remedy has to be assembled from instruments held by different authorities. That supplies the structural reason this report describes eight problems and no single fix: not because the engineering lacks clarity, but because the statute divides the authority to act on it by design.

Central Findings

After examining engineering studies, regulatory proceedings, operational events, and planning reforms across North America, five conclusions emerge.

Finding 1 — Large-load integration is one planning problem expressed at multiple stages of the interconnection process. The queue, the forecast, the studies, the adequacy assessment, and the cost allocation built on them form a single chain, so a distortion at the front — optionality recorded as firm demand — propagates through every stage below it. The eight sections therefore resist eight separate fixes.

Finding 2 — Time, not energy, is the limiting resource. Generation, transmission, and the load arrive on different clocks, and the load now moves fastest of the three. The binding question has shifted from whether capacity can be built to whether it arrives before the load does.

Finding 3 — Engineering assumptions have become regulatory assumptions. How a facility rides through a fault, how far it can curtail, and whether it islands no longer concern only a protection engineer; those behaviors increasingly determine the terms of tariffs, interconnection agreements, and NERC registration. The physics of the load now shapes the rules that govern it.

Finding 4 — Institutional authority is fragmented in the same places as technical responsibility. The load sits astride the §201(b) line — a bulk-system element on the federal side, a retail customer on the state side — so no single regulator holds both a remedy and its consequence. Every fix must be assembled from instruments that different authorities control.

Finding 5 — Demand flexibility moves on the same deployment timescale as the problem, where the supply-side responses do not. Interruptible and curtailable service needs no new turbines or transmission and runs at the load's own speed, which is why it recurs across the report as a leading answer to the timeline, cost, and adequacy problems at once. The trade-off lies in who directs the reduction, and it determines how much of the capability a planner can credit. A load that schedules its own reductions incurs little cost and earns no planning credit, because no obligation attaches to them. A reduction the operator can call earns full credit, and the load prices it against the cost of an interruption it does not choose. Section 6 sets out the terms that convert the first into the second, and explains why the difficulty lies in that conversion rather than in the physical capability.

Why these findings matter. The load has outrun the cycle meant to absorb it. Interconnection queues, resource plans, and reliability standards all run on multi-year clocks, and roughly a decade of demand has compressed into a few of them. The instruments taking shape in response — registration criteria, curtailable-service terms, cost-allocation formulas, ride-through obligations — will set the terms on which the next generation of load connects, and they will resist unwinding once they take force. The next five to ten years will therefore decide whether the North American framework adapts through the instruments it already holds or whether the accumulated mismatch forces a deeper redesign. That choice falls now, in proceedings already open, not at some later reckoning.

Read together, the five findings point past any single technical fix. A faster queue, a ride-through standard, a cost-causation tariff, a registration rule — each answers one finding and none resolves the set, because one difficulty runs beneath them all: the engineer, the market, the planner, and the regulator each see a different load. The engineer sees a fault-ride-through case, the market a price-responsive customer, the planner a firm block of demand, the regulator a retail account — and the load answers to all four at once, which its predecessors never did.

Reliable integration will therefore depend less on discovering a single solution than on aligning those four views around one shared account of the load, and writing the studies, incentives, plans, and rules against that account rather than against four partial pictures of it. A grid that agrees on what it connects can integrate it. That agreement, more than any single reform, marks where the evidence in this report finally points.

Technical Annex

The main report is written to be read without an engineering background. This annex serves readers who need the layer beneath it: how planners forecast future demand, characterize a load, decide through interconnection and transmission studies what to build, judge resource adequacy, model the system, analyze scenarios, assess reliability, trace distribution impacts, and fit the planning paradigms together. Each part cross-references the section it supports. Nothing here is required to follow the argument; all of it is required to act on it.

T.1 Data-center electrical architecture (supports the Primer and Section 4)

Why this section exists. Every reliability argument in this report turns on what happens inside the facility between the utility feed and the processor, because that chain contains the protection devices that decide whether the load stays connected. A reader who treats a data center as a block of demand cannot evaluate a ride-through requirement, since the requirement applies to equipment inside the block. The architecture below is also what makes flexibility possible: the same uninterruptible supply that lets a facility disconnect cleanly lets it ride through a curtailment instruction, so Sections 4 and 6 describe opposite uses of one capability.

From the grid's point of view a data center presents not as a building full of computers but as a large, inverter-interfaced power-electronic load. Utility power arrives at medium-voltage switchgear at the facility point of interconnection, passes through uninterruptible power supply (UPS) systems that rectify AC to DC and invert it back — usually with battery storage on the DC bus — then flows through power distribution units to the racks, where server power supplies convert it once more. Every server therefore sits behind at least two stages of power-electronic conversion, so the facility presents to the grid as an inverter-dominated load with almost no rotating inertia to resist frequency change.

Figure T1 — Data-center electrical architecture: where grid power becomes compute, and where ride-through lives

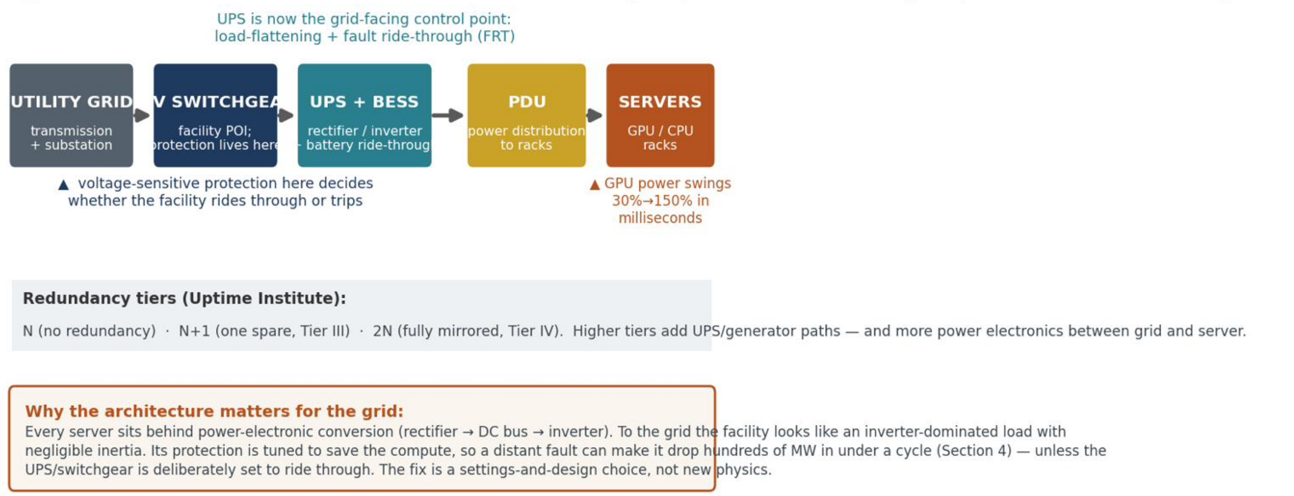


Figure T1 — The electrical chain from utility feed to processor. Two points in that chain determine how the facility behaves during a grid disturbance: the protection settings at the facility switchgear, and the uninterruptible supply that increasingly doubles as the grid-facing control point. The owner specifies both, not the system operator, so facility behaviour during a fault stays largely invisible in planning models. The reader should conclude that ride-through obligations and registration are the same problem approached from two directions (Section 4).

Each stage of that chain removes information the grid would otherwise have. The medium-voltage switchgear sets the protective thresholds at which the whole site separates, and those settings are chosen by the owner rather than the operator. The transformer's winding configuration determines whether an unbalanced fault reaches the equipment as an unbalanced sag or is partially tempered by a delta-wye arrangement — which is why NERC's January 2026 reviews identified winding configuration as a root cause. The uninterruptible supply decides whether a disturbance produces a full transfer, a partial one, or none. By the time power reaches the rack, the facility's response to a grid event has already been determined by three pieces of equipment the operator neither specifies nor sees.

The load beyond the rack is not uniform either, and the split matters for both reliability and flexibility. Roughly half to two-thirds of facility demand goes to IT load — servers, storage, networking — which the uninterruptible supply protects. The remainder is mechanical: chillers, computer-room air handlers, pumps and cooling towers. Mechanical plant is frequently *not* on the protected bus, because carrying it there would require far more battery and generator capacity. That asymmetry produces two consequences. Cooling can trip on a voltage sag while the IT load rides through, so a facility may shed in stages rather than at once. And thermal inertia in the cooling system is itself a flexibility resource — pre-cooling a hall before a curtailment window lets compute continue while mechanical load drops, which is the mechanism behind several of the demand-response arrangements in Section 6.

Two developments are changing the picture faster than the standards can follow. Rack power density has risen from roughly 5–10 kW in a conventional hall to 100 kW and beyond for accelerator racks, with 600 kW designs in prospect, which is what drives the shift to liquid cooling and concentrates a very large load into a small electrical footprint. And the battery in the uninterruptible supply is increasingly capable of acting as a grid resource rather than merely a buffer: the same hardware that lets a site island can, if the controls permit and the operator can see it, provide fast frequency response or hold load through a sag instead of dropping it. Whether that capability is used or merely present comes down to settings, which returns to the theme of Section 5 — capability the operator cannot see is capability the system cannot count.

Two architectural facts drive the reliability problem. First, redundancy: availability tiers (Uptime Institute's Tier I– IV, expressed electrically as N, N+1, or 2N) add parallel UPS and generator paths for resilience — which means more power electronics, and more protection logic tuned to protect the compute, sitting between the grid and the load. Second, the UPS has become the grid-facing control point. Traditional “UPS-last” designs treated backup power as a downstream facilities component; modern AI-oriented designs place a medium-voltage UPS with battery storage as a high-bandwidth control point that can, in principle, flatten load and provide fault ride-through. That same equipment, set conservatively, also disconnects the facility in under a cycle during a distant fault (Section 4). The fix for uncommanded tripping turns largely on UPS design and protection settings — a configuration choice, not new hardware physics.

Planning implication. Because the behaviour that matters is set below the meter and specified by the owner, the binding constraint becomes visibility rather than cost. The operator cannot require, credit, or even verify what it cannot see, so the architecture described here turns into an operations and registration problem the moment the facility energises (Sections 4 and 5).

T.2 Industrial load characteristics: why computational load differs in kind (supports Sections 4 and 6)

Why this section exists. Planning practice rests on a century of observed load behaviour — diversity, weather sensitivity, predictable ramps — and each of those regularities is an empirical fact about the load mix rather than a law. This section sets out which of them computational load breaks and which it preserves, because the answer determines which planning tools remain valid. Where the load resembles conventional industrial demand, existing methods transfer; where its coincidence factor approaches one and its output is insensitive to weather, they do not, and the forecasting and adequacy problems in T.5 and T.6 follow directly from that difference.

Large industrial loads are not new; aluminum smelters, steel mills, and chlor-alkali plants have drawn hundreds of megawatts for decades. The novelty lies in the electrical signature. Traditional heavy industry draws a relatively steady load with a high load factor and, crucially, often tolerates interruption and changes slowly. AI training clusters invert every one of those properties. Thousands of GPUs ramp in lockstep at mini-batch boundaries, producing quasi-periodic, multi-megawatt power swings over seconds — and instantaneous swings from roughly 30% to 150% of nominal within milliseconds at the rack level. Aggregated to gigawatt scale, published estimates put the resulting system-level ramp rates above 1,000 MW per second, and field measurements have documented forced oscillations (a 14.7 Hz mode in one study) that can couple with weakly damped grid modes. The contrast is worth making precise, because it explains why existing planning methods become less accurate:

Characteristic	Traditional heavy industry	AI training data center	Crypto mining
Load factor	High, steady	High but volatile (bursty)	High; can idle instantly
Ramp behavior	Slow, gradual	Multi-MW swings in seconds; >1,000 MW/s aggregated	Near-instant on/off
Inertia contribution	Rotating motors add some	Negligible (inverter-fed)	Negligible
Fault behavior	Rides through; motor loads	Voltage-sensitive trip via UPS/protection	Trips or self-curtails
Flexibility	Limited; process-bound	Real but SLA- and checkpointconstrained	High (economics-driven)
Power quality	Harmonics from drives	Harmonics + fast transients + oscillations	Harmonics; switching

Table T2A — Computational load against the industrial load the system was built around. The rows that matter are ramp rate and controllability: a smelter is large but slow and contractually interruptible, while a training cluster is large, fast, and governed by a scheduler the operator cannot see.

Sources: NERC; ERCOT; EPRI; and the peer-reviewed and preprint literature listed in the Technical Annex references.

The practical implication: a planner who models an AI campus with the assumptions appropriate to a steel mill will understate both its instability risk (Section 4) and, paradoxically, its flexibility potential (Section 6). Computational load proves simultaneously more dangerous and more controllable than the industrial load it superficially resembles.

Three properties deserve separate statement, because each breaks a different planning assumption. **Coincidence.** Aggregate load has always been smaller than the sum of individual peaks, because customers peak at different times, and the whole apparatus of diversity factors rests on that. A fleet of facilities running near maximum around the clock has a coincidence factor approaching one, so its contribution to system peak approaches its full connected load — and, being built to common reference designs with common protection settings, it also responds to a disturbance in unison. Diversity fails on both the demand side and the fault-response side at once.

Weather insensitivity. Conventional load forecasting reduces largely to weather modelling: heating and cooling drive the peaks, and the relationship between temperature and demand forms the strongest signal in the data. Computational load barely responds to weather on the IT side, though its cooling load does. The consequence is not that it is easier to forecast but that the established method does not apply — and, more seriously, that it adds to demand in extreme temperatures without the offsetting behaviour that other load classes show, arriving at full strength in precisely the hours when the supply side is most degraded (T.6).

Ramp rate. A steel mill's arc furnace cycles on a schedule an operator can anticipate. A training cluster changes power by a large fraction of its rating in seconds when a job starts, checkpoints or fails, and synchronised workloads across thousands of accelerators make those swings coherent rather than averaging out. Reported behaviour runs to tens of percent of nominal rack power within tens of milliseconds, aggregating to hundreds of megawatts per second at campus scale. No North American jurisdiction presently imposes a generally applicable ramp-rate limit on an individual large computational load, though individual service agreements may contain one, which is the gap Section 4 identifies; the market procures the load-following resource that must absorb the swing without anyone specifying how large the swing may be.

T.3 Interconnection and transmission studies: can the existing grid take the load? (supports Sections 1 and 5)

Why this section exists, and how the studies are used. An interconnection study is the gate through which every project in Section 1’s queue must pass, and its sequence explains both the duration of the queue and where it can be shortened. The studies run in a fixed order because each depends on the one before: a feasibility screen establishes whether an obvious constraint exists, a system impact study quantifies the constraint and identifies upgrades, and a facilities study prices and schedules them. Each stage assumes a set of prior projects already connected, so a withdrawal upstream forces restudy downstream — the mechanism behind cluster processing and the readiness deposits that Order No. 2023 introduced on the generation side and that large-load reform is now copying.

For the reader, the operative point is what these studies do *not* decide. They establish whether the network can carry the load and what it costs to make it do so. They do not establish whether enough generation exists to serve the load over a year (T.6), whether the load will behave acceptably during a disturbance (T.4 and T.8), or whether anyone should bear the cost (Section 3). A project can clear every study and still present each of those three problems, which is the structural reason this report treats interconnection and adequacy as separate sections rather than stages of one process.

When a large load applies to connect, the operator must answer a deceptively simple question: can the existing transmission system serve it reliably, and if not, what must be built? A staged process governed by NERC's transmission planning standard TPL-001 (currently TPL-001-5.1) answers it, requiring the system to withstand a defined spectrum of contingencies without cascading or uncontrolled load loss. The studies run in a fairly consistent order, each a gate the project must pass:

- Feasibility / screening. A first-pass look at whether the point of interconnection has obvious capacity, and at rough upgrade scope. In batch regimes (ERCOT's Batch Zero) the operator pools and ranks requests against shared transmission limits rather than studying them one at a time.
- Steady-state power-flow (load-flow) study. Solves the network under normal and contingency conditions to check for thermal overloads on lines and transformers and for voltage violations (typically outside 0.95–1.05 pu). This study most often produces the upgrade list.
- Short-circuit (fault-duty) study. Compares fault current before and after the load (and any associated generation) against breaker interrupting ratings per TPL-001-5.1; if the addition pushes a bus over its breaker rating, equipment must be replaced.
- Transient-stability (RMS dynamic) study. Exercises the load's dynamic model against the network to evaluate frequency response, voltage recovery, ramping, energization inrush, load tripping/reconnection, and fault ride-through against the operator's disturbance-performance requirements.
- EMT screen at weak-grid points. Where local system strength runs low (screened by short-circuit ratio and its weighted/composite variants and by critical clearing time), a full-waveform electromagnetic-transient study becomes mandatory to catch control-interaction instabilities that RMS tools miss. SPP's HILL process makes this gate explicit. What comes out the other end is an upgrade list and its cost. Typical remedies escalate in expense and lead time: reconductoring an existing line with higher-capacity conductor, adding or upgrading transformers, installing reactive support (capacitor banks, STATCOMs, or SVCs) to hold voltage, and — the expensive, slow end — building new lines or substations. The contingency framework is the heart of the analysis: planners test the loss of single elements (N-1) and combinations (N-1-1), sizing the system against the largest credible loss. The large-load era breaks two assumptions inside it. First, load historically behaved as a passive, roughly predictable quantity — not a contingency that can vanish in a cycle; NERC's 2026 guidance now asks planners to treat the sudden loss of a clustered large load as a studied contingency comparable to losing a large generator (BAL-002's Most Severe Single Contingency). Second, because data centers cluster, the local grid can be too weak for stable inverter-dominated operation — the precise condition the EMT screen exists to catch. Who pays for the resulting upgrades is the cost-allocation fight of Section 3; whether the study even sees the failure mode is the modeling problem of T.4.

Two features of this sequence explain most of the delay that Section 1 describes. Each study is performed against a base case that assumes a specific set of prior projects already connected, so a withdrawal upstream changes the answer downstream and forces restudy — the restudy cascade that serial queues produce and that cluster processing exists to prevent. And the studies are sequential by necessity rather than by convention: a stability study cannot be run before the power flow that defines its operating point, and an EMT screen is meaningless until short-circuit analysis has established where grid strength is low. Reforms that shorten the queue therefore work by reducing the number of times the sequence is run, not by running it faster.

What the studies produce is a list of network upgrades with costs and construction lead times, and that output is the hinge between the engineering and everything else in this report. The upgrade list determines the interconnection cost that Section 3 argues about allocating. The construction lead times — four years for a large transformer, longer for a new line — set the energisation date that Section 2 finds the load outrunning. And the studies establish which network conditions bind, which is what makes conditional and curtailable service possible at all: a load willing to reduce when a constraint binds can often be served without the upgrade — the structure of SPP's CHILLS and of European flexible connection agreements (Section 6).

CASE · SPP CHILLS — CONDITIONAL SERVICE

SPP's Conditional High Impact Large Load Service shows the interconnection gate and the control room negotiating directly. Rather than wait for a designated resource to come online or for network upgrades to finish, a large load can take long-term but non-firm energy and transmission — served now, on the condition that SPP may curtail it when capacity or transmission runs short. The assessment runs in about 90 days, and each conditional load carries separate telemetry and billing so the operator can see and bill it in real time. FERC accepted the tariff revisions effective July 1, 2026 (Docket ER26-1323), subject to condition. The trade is explicit: the load buys speed by accepting curtailability, and the studies of this section define which hours and which constraints trigger it — a planning result enforced as an operating obligation.

CASE · IRELAND — SELF-SUPPLY AS THE CONDITION OF ACCESS

Ireland reached the same gate from the opposite side. With data centres at roughly a fifth of national electricity demand and about half of the Dublin region's, the Commission for Regulation of Utilities issued a Large Energy Users Connection Policy in December 2025 that conditions a new data-centre connection on the load supplying its own firm capacity. A facility above 10 MVA must provide dispatchable on-site or proximate generation or storage matching its maximum import capacity and participating in the wholesale market, on a six-year path to compliance and with a renewable-supply match of about 80% of annual demand; the self-supply requirement replaces an earlier mandatory-curtailment regime. Location still governs — the system operators assess each site against local constraints and short-circuit strength. Where SPP conditions access on curtailability, Ireland conditions it on the load carrying its own generation — the co-location logic of Section 5 written into connection policy.

T.4 Power-system modeling: the tool decides what the study can see (supports Section 4)

A study can only match its load model, and the record here weighs heavily against the legacy practice. For decades, load in stability studies was represented statically — as a fixed or voltage-proportional P–Q injection (the ZIP or constantpower model). A static model cannot represent a load that disconnects itself during a voltage sag, so a study built on one will report that a facility rides through a fault it would in reality trip on. That gap amounts to more than rounding: it separates seeing the dominant failure mode from missing it.

Before the load model itself, it is worth setting out what each class of study is for, because the terms recur throughout this report and the distinctions decide which questions a given analysis can answer. The four study types below run in sequence during an interconnection review, and the three load models beneath them determine what each of those studies sees.

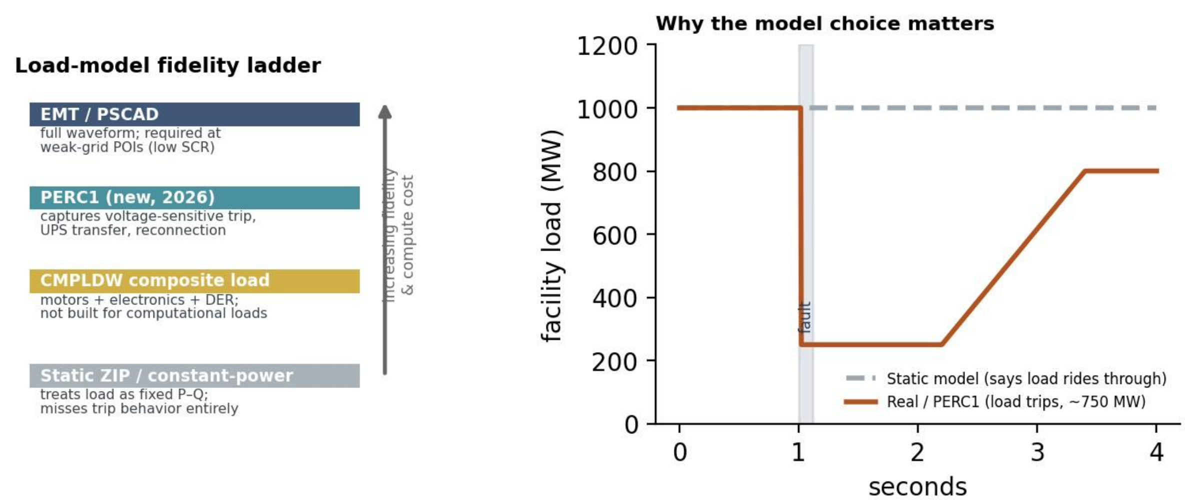
Model or tool	What it represents	What it is used for	What it cannot see
Power flow (steady state)	The network at one instant: voltages, angles, and flows for a fixed dispatch and load.	The first screen in any interconnection study — whether adding the load overloads a line or drives a bus outside its voltage band, under normal and contingency conditions.	Time. It has no dynamics, so it cannot show what happens during or after a fault, only the condition before and the condition after.
Short-circuit study	Fault current magnitude and distribution for faults at each bus.	Sizing breakers and coordinating protection; establishing whether existing equipment can interrupt the available fault current once new sources are added.	Load behaviour. It answers what the network delivers into a fault, not how connected equipment responds to one.
Positive-sequence dynamic (RMS) simulation	Machine and control dynamics at fundamental frequency over seconds to tens of seconds.	The standard transient stability study: whether generators stay synchronised and voltage and frequency recover after a contingency.	Waveform-level behaviour. It approximates power-electronic control loops and current limits, which is why it is unreliable for inverter-dominated interfaces.
Electromagnetic transient (EMT) simulation, PSCAD class	Full instantaneous waveforms, including converter switching, control loops and protection logic, at microsecond resolution.	Required where power-electronic content is high or grid strength is low — and, under NERC's May 2026 alert, at large-load points of interconnection. It is the only class of study that reproduces the trip behaviour in Section 4.	Scale. It is computationally expensive and run on a reduced network, so it answers local questions rather than system-wide ones.
Static load model (ZIP or constant power)	Load as a fixed or voltage-proportional real and reactive power draw.	Legacy practice in stability studies, adequate while load was predominantly motors and resistive heating.	Self-disconnection. It cannot represent a load that trips on a voltage sag, so a study using it reports ride-through where the facility would in fact leave.
WECC composite load model (CMPLDW)	A mix of motor classes, static load, electronic load and distributed generation behind a representative feeder.	The industry workhorse for distribution-connected load in stability studies.	Computational load. Built to capture motor-driven feeders; NERC has stated explicitly that it does not represent the parameters that make large data-center load risky.
PERC1	Purpose-built computational load: voltage-sensitive trip threshold and duration, reconnection voltage and delay, IT versus non-IT load fraction, and UPS transfer logic.	Required by NERC's Level 3 alert for planners studying these facilities, or a model of equivalent or better capability. It is the precondition for enforcing a ride-through standard — an obligation cannot be required if it cannot first be simulated.	Facility-specific settings it is not given. The model is only as good as the parameters an operator can compel the owner to provide (Section 4).
PERC2 and the Data Center Load Modeling Technical Reference	Successor specification and the accompanying parameter guidance.	In development; intended to standardise how these facilities are represented across regions rather than leaving each planner to its own assumptions.	Not yet issued at the date of this report.

Table T4A — What each model is for. The final column is the operative one: every model in this table abstracts the system to make a specific question tractable, and each stays silent on questions outside that purpose. The failures described in Section 4 were not modelling errors so much as the use of a model outside the range of questions it was built to answer.

Sources: NERC Level 3 Essential Action Alert (May 2026) and accompanying reliability guideline; WECC composite load model documentation.

Two practical points follow. The models accumulate rather than replace one another: a planner runs power flow to find the binding constraints, short-circuit to check the equipment can clear a fault there, positive-sequence dynamics to confirm the system recovers, and EMT only at the interfaces where the previous step cannot be trusted. And the load model feeds the last two rather than standing as a study in its own right — which is why substituting PERC1 for a static representation changes the answer without changing the study, and why the modelling reform is the precondition for every reliability requirement that follows it.

Figure T2 — Power-system modeling: the tool determines whether a study even sees the failure mode



A static model reports a facility that rides through a fault it would actually trip on. Capturing the real behavior needs CMPLDW at minimum, PERC1 for computational loads, and EMT at weak-grid points — the modeling reform NERC's Level 3 Alert (May 2026) requires. Trace is illustrative.

Figure T2 — The fidelity ladder and why it matters. The same disturbance produces opposite conclusions depending on the load model: a static model shows ride-through; a computational-load model shows the trip that actually happens. The same ladder runs through the simulation method itself: a static representation, then positive-sequence (RMS) dynamics, three-phase RMS, full electromagnetic-transient (EMT) simulation, and finally hardware-in-the-loop testing — each rung resolving faster phenomena at higher computational cost, and each warranted only where the rung below it cannot answer the question.

The industry's workhorse dynamic model, the WECC composite load model (CMPLDW), marked a large step up — representing a mix of motor types, static load, power electronics, and distributed resources — but its developers built it to capture motor-driven distribution feeders, not gigawatt computational loads, and NERC has been explicit that it does not capture the parameters that make these loads risky. In response came a new model, PERC1, purpose-built to represent the voltage-sensitive trip threshold (the voltage and duration at which the load disconnects), the reconnection voltage and time delay, the split between IT and non-IT load fraction, and the UPS transfer logic. NERC's Level 3 Alert (May 2026) requires planners to use PERC1 or a model with equivalent-or-better capability, and to add full-waveform EMT (PSCAD-class) modeling at weak-grid interconnections or where power-electronic content is high. A Data Center Load Modeling Technical Reference and a PERC2 specification are following. The modeling reform sets, in a real sense, the precondition for every other reliability fix: you cannot require ride-through performance you cannot first simulate. Two further modeling frontiers matter for planning. First, inverter-based representation: because the facility interfaces the grid through power electronics with negligible inertia, planners must model it the way they now model inverter-based generation — with explicit control loops, current limits, and fault-current behavior that positive-sequence tools approximate poorly, which is why EMT is increasingly mandatory at the point of interconnection. Second, flexibility representation: if a large load is to be credited as curtailable (Section 6) or accredited for adequacy (T.6), the model must separate firm from flexible load components, represent behind-the-meter generation and storage, and reflect operating windows such as AI-training schedules. A model that carries a flexible gigawatt as a firm subtraction from demand will misstate both the reliability risk and the reserve value — the reason NERC's guideline asks explicitly for firm-versus-flexible decomposition.

How the study is actually performed. The preceding pages say what each model represents and why the load model decides the answer. They do not say how an engineer chooses between them, parameterises them, or knows when a study has stopped being trustworthy. Those judgments carry most of the practical risk in this subject, because a study that is run correctly against the wrong assumption returns a clean result and a false assurance. What follows is the working guidance, stated as practice rather than as requirement — the mandatory elements appear in the NERC Level 3 Alert described above.

- When RMS is sufficient. Positive-sequence RMS (phasor) simulation assumes a balanced system at fundamental frequency and represents the network by its impedance, a good approximation for phenomena slower than roughly five to ten hertz. That covers most of what a planner does: angular and voltage stability against normally cleared faults, governor and exciter response, inter-area oscillations, and the large screening runs across many contingencies that no EMT tool could complete in the time available. RMS remains the right first instrument for a large load whose behaviour is well characterised and whose trip settings are represented in the load model. It stops being sufficient at the point where the answer depends on what happens inside the cycle.
- When EMT is required. Four conditions, any one of which is sufficient. A weak interconnection, conventionally a short-circuit ratio below about three, where the converter's control loops and the network interact rather than acting in sequence. High power-electronic content at the point of interconnection — inverter-based generation, storage, HVDC, or the rectifier front end of the load itself — where control interactions and sub-synchronous oscillations live above the phasor model's ceiling. Any question whose answer depends on sub-cycle behaviour: fault ride-through at the equipment's own trip threshold, protection coordination, unbalanced faults, harmonic interaction, or the transfer of a UPS. And any case where the RMS study returns a marginal result, since marginal in phasor terms provides no evidence of margin. The Level 3 Alert makes the first two mandatory rather than discretionary at weak-grid interfaces and where power-electronic content is high.
- Which model, and how to parameterise it. For conventional motor-driven feeders the composite load model remains the right tool, parameterised from feeder composition — the motor fractions by class, the static and power-electronic fractions, the distributed-resource share, and an equivalent feeder impedance — drawn from end-use data rather than from library defaults, which are regional averages and belong to someone else's system. Where computational load is material the composite model is the wrong instrument at any parameterisation, because the quantities that decide the outcome sit outside it: NERC's Level 3 Alert directs planners to PERC1 or an equivalent, whose four governing parameters — the voltage and duration at which the load disconnects, the voltage and delay at which it returns, the IT to non-IT split, and the UPS transfer logic — can only come from the customer. That is the practical significance of the registration and data-provision requirements in Section 4: without them the model has no inputs, and the study reverts to an assumption.
- Validating an inverter or UPS model. Validation means benchmarking the model against measurement, not confirming that it runs. The practical sequence is: a flat start, confirming the model initialises and holds steady with no disturbance applied — a model that drifts here will produce plausible nonsense later; response tests against voltage and frequency steps and against a fault-and-recovery profile, compared with factory or commissioning records; and, where measurement exists, replay of a recorded system event through the model, with the discrepancy characterised rather than tuned away. For power-electronic equipment the RMS and EMT representations must also be benchmarked against each other on a common disturbance: where they disagree, the EMT result governs and the RMS model needs the work. Vendor-supplied models should be treated as unvalidated until the settings in them match the settings commissioned on site, which is the point at which most of the disagreement is found.
- What invalidates a study. Six assumptions do most of the damage, and all of them are defaults rather than decisions. A static or ZIP load representation, which cannot trip and therefore cannot reproduce the event the study is meant to examine. A load model with no trip logic or no reconnection logic, which converts an uncommanded loss of demand

into a smooth voltage recovery. Aggregation at the transmission bus, which averages away the distribution-level behaviour that decides whether the load rides through. Positive-sequence representation of an unbalanced event, which understates the phase-to-ground voltage the equipment actually sees. Generic inverter models left at library defaults, which encode a manufacturer's example rather than the installed settings. And a single peak-load case, which is the wrong condition for a computational campus whose worst hour for the system may be a light-load night. A study carrying any of these is not conservative; it stays silent on the failure mode.

Planning implication. Because planners can evaluate only what their models represent, an inappropriate study method understates both the upgrade list and the operational risk — and does so silently, returning a clean result for the wrong question. The modelling reform of Section 4 therefore does not merely refine the analysis; it makes the reliability requirements that depend on it possible at all.

Sources: NERC Level 3 Essential Action Alert (May 2026) and accompanying reliability guideline; WECC composite load model documentation and model validation guidance; NERC modelling notifications and inverter-based resource performance guidelines; IEEE 2800 model-validation provisions. Stated as engineering practice, not as a filed position.

T.5 Load forecasting: methods, load types, and uncertainty (supports Section 1)

Why this section exists, and what the methods are for. Every number in Sections 1 through 3 rests on a forecast, and forecasts of large load are produced by a different method from the econometric extrapolation that served a stable system. The conventional approach regresses historical consumption on weather and economic drivers, which works when next year's load is a modest perturbation of this year's. It fails when a single customer can add several percent to a zone, because the historical series contains no analogue. The replacement is a bottom-up, project-based method: enumerate the specific requests, assign each a probability of proceeding, apply a load factor and ramp schedule, and sum. The probability weighting separates a forecast from a queue total, and the assumptions listed below explain how two forecasters using the same method arrive at answers differing by a factor of three.

Section 1 introduced the three forecasting families (econometric top-down, bottom-up project-weighting, and probabilistic scenario). This part adds the two things a forecaster must get right beneath the method: how different large loads are characterized, and how uncertainty is carried through. Characterizing the load matters because "data center" is not one thing. Each type presents a different shape to the grid, and a forecast that lumps them together will misjudge both peak and flexibility:

- AI training clusters. Very high load factor, run flat-out around the clock, but with second-scale synchronized swings; near-unity coincidence with system peak. Some scheduling flexibility, constrained by checkpoint overhead.
- AI inference / cloud. High, steadier load tied to user demand; more diurnal shape; tighter service-level agreements limit curtailment.
- Traditional colocation. Moderate, diverse tenant mix; the closest to legacy commercial load; more predictable.
- Cryptocurrency mining. High load factor but highly price-responsive — can idle in seconds when power prices spike or the grid tightens, making it the most flexible of these load classes by nature.

Beneath the method sit five assumptions that move the answer as much as the request-weighting does. A careful reader should interrogate each before trusting any large-load forecast. *Coincidence and diversity:* AI training carries a coincidence factor near one, so assuming legacy diversity understates its peak contribution. *Load factor and ramp schedule:* how fast a campus reaches full draw, and how flat it then runs. Developers supply both through attestation letters, such as ERCOT's "Officer Letter Loads," and the forecaster must decide how far to trust them. *Spatial allocation:* because loads cluster, a forecast reasonable statewide can badly understate one zone — PJM adjusted 14 of 15 zones in 2026. *Weather and economic normalization:* the econometric base rests on weather and macro inputs such as Moody's Analytics data. *Probability weighting:* feasibility-stage against contracted load, among the largest swing factors and the least standardized of the five. Uncertainty suffers frequent mishandling. Because these dials are set independently and rarely disclosed together, point forecasts across the industry diverge by a factor of three. The mature response is not a single number but a distribution: a reference case bracketed by high- and low-growth scenarios, ideally probabilistically weighted, with the spread itself reported as information. NERC's March 2026 gaps assessment put it plainly — existing load-forecasting methods may be insufficient for loads that bid into multiple areas at once, overstating expected demand. A forecast measures nothing; it stacks assumptions. Reading one honestly means asking which dials were turned, how far, and how wide the resulting band is. That posture leads directly into scenario analysis (T.7).

CASE · PJM — WHEN THE SAME REQUEST IS COUNTED TWICE

PJM's 2025 long-term forecast shows the speculative-load problem propagating into planning. The forecast carries peak growth of about 32 GW between 2024 and 2030, of which roughly 30 GW — the overwhelming majority — is data-center load. Because that number drives both capacity procurement and transmission planning, its integrity matters twice over. The difficulty is double-counting: a developer hedging against local constraints can file the same project into several utilities' study queues, so aggregated utility forecasts (one analysis put the PJM-footprint total near 55 GW by 2030) can exceed any real demand. PJM's response has been procedural rather than physical — a standardized large-load adjustment-request process meant to gauge how likely each request is to materialize, and when — which is the forecasting discipline this section describes, applied at regional scale.

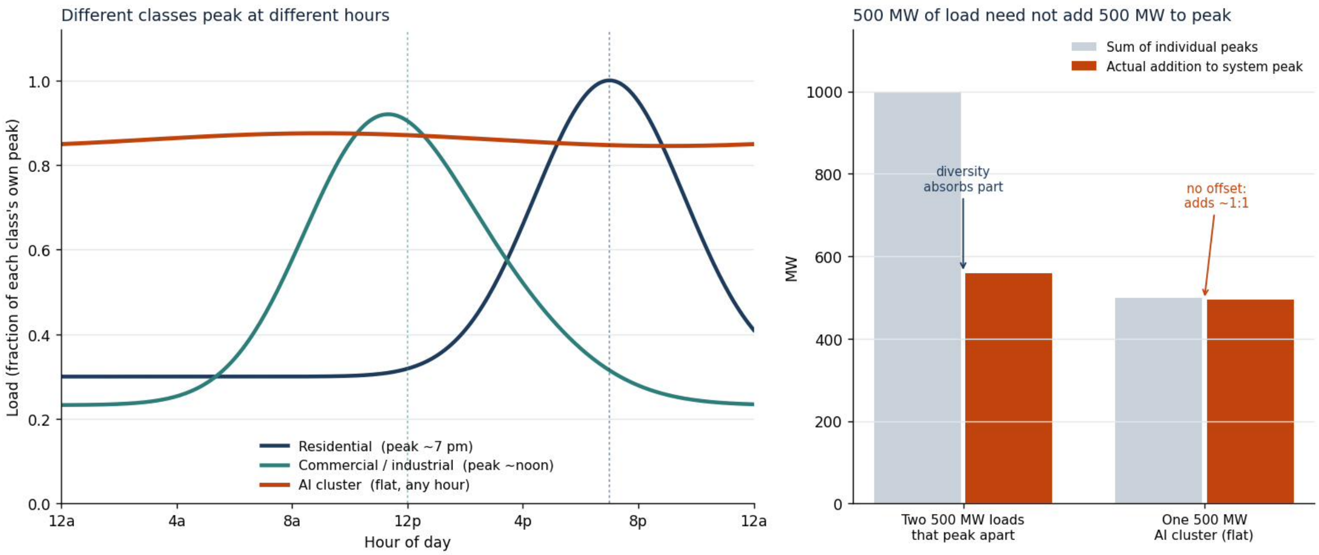


Figure T7 — Coincident versus non-coincident peak. Residential load peaks in the evening and commercial-industrial load near midday, so their combined peak falls below the sum of the two — diversity absorbs part of it. A flat, high-load-factor AI cluster has no such offset: it sits near full value in every hour, including the system peak, so it adds to peak almost one-for-one. Shapes are schematic, each normalised to its own peak.

T.6 Resource adequacy planning: will there be enough generation and reserves? (supports Section 2)

Interconnection studies (T.3) ask whether the wires can carry the load. Resource adequacy planning asks a different question: across a whole year of weather, outages, and demand variation, will there be enough accredited generation and reserves to keep the lights on to an agreed standard of reliability? The answer is expressed in a small set of probabilistic metrics that a reader of any RTO assessment will encounter.

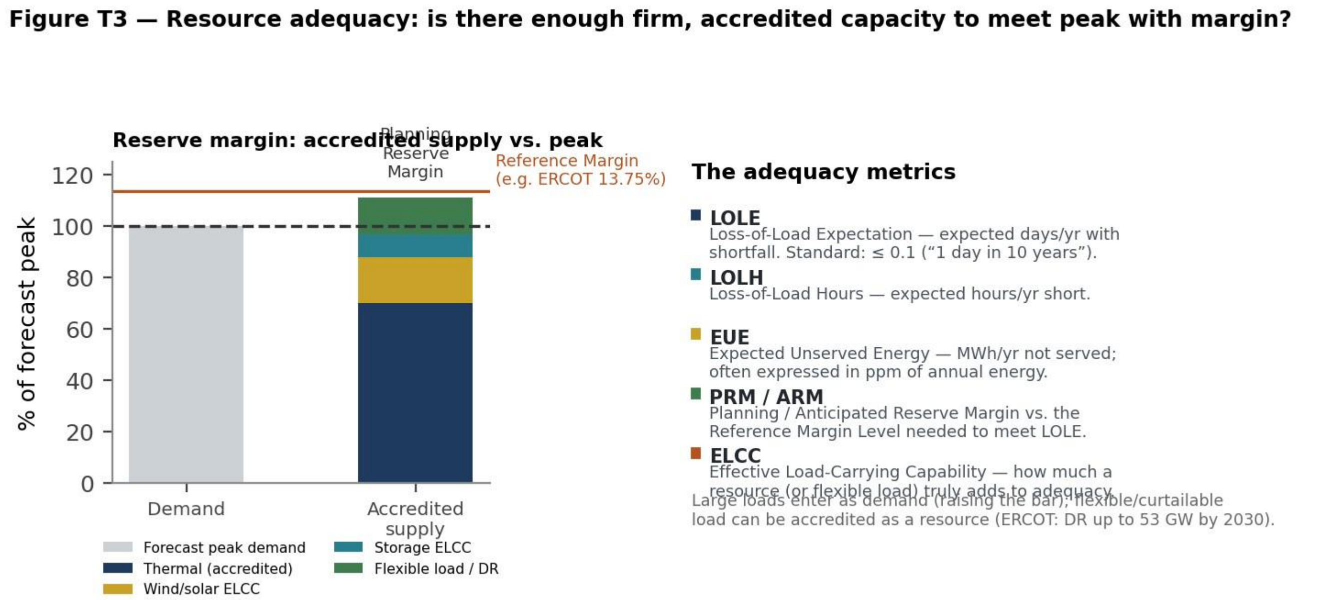


Figure T3 — The two halves of adequacy: is accredited supply above peak with adequate margin (left), and do the probabilistic metrics stay within their thresholds (right)? Large loads raise the bar; flexible loads can be accredited to help meet it.

The reliability standard is conventionally a loss-of-load expectation (LOLE) no worse than 0.1 days per year — the familiar "one day in ten years." Around it sit companion metrics: loss-of-load hours (LOLH), expected unserved energy (EUE, often expressed in parts-per-million of annual energy), and the planning or anticipated reserve margin (PRM/ARM) compared against the reference margin level needed to satisfy LOLE. These are computed with probabilistic assessments (ProbA) — Monte Carlo simulations over many combinations of weather years, loadforecast uncertainty, and generator forced outages (SERC, for instance, runs thousands of simulations per hour across four decades of weather). The 2026 assessments indicate the strain: PJM's anticipated reserve margin has fallen sharply, and NERC places several regions at elevated or high risk beginning around 2028–2029. Large loads change the calculation in two directions. As demand, they raise the bar — more accredited capacity is needed to hold the same LOLE. Genuine flexibility, however, lets them earn accreditation as a resource: ERCOT's 2025 assessment credits demand response rising toward 53 GW by 2030 and notes that because new large loads can be curtailed during emergencies, their rapid growth has substantially less effect on reserve margins than a firm forecast would imply. The accreditation method matters enormously — ERCOT's switch from historical peakcapacity factors to probabilistically-derived effective load-carrying capability (ELCC) for wind and solar materially cut their credited contribution. The frontier question, now in the literature, asks how to accredit co-located load-plus-generation-plus-storage systems, where naive approaches (summing individual ELCCs) produce large errors. Getting accreditation right determines whether a flexible gigawatt counts as a problem or part of the solution.

The distinction between the two questions is worth stating carefully, because they are routinely conflated in public argument. An interconnection study is deterministic and local: it asks whether a specified network, under a specified condition, holds. Adequacy planning is probabilistic and system-wide: it asks how often, across thousands of possible years, supply falls short anywhere. A system can pass every interconnection study and still be inadequate, and it can be adequate in aggregate while a particular corner of the network cannot be served. The two failure modes have different remedies — wires for the first, capacity or demand reduction for the second — the reason Sections 1 and 2 treat them separately.

Metric	What it measures	Conventional threshold	What it misses
LOLE — loss-of-load expectation	The expected number of days per year on which available supply falls short of demand at any point in the day.	0.1 days per year, the familiar "one day in ten years."	Duration and depth. A one-hour shortfall of 50 MW and a twelve-hour shortfall of 5 GW each count as one event.
LOLH — loss-of-load hours	The expected number of hours per year of shortfall.	No universal standard; used alongside LOLE to expose duration.	Magnitude. It counts hours, not megawatts unserved.
EUE — expected unserved energy	Expected megawatt-hours of demand that cannot be served, often normalised as parts per million of annual energy.	No universal standard; the metric most sensitive to depth.	Timing and concentration. The same annual EUE may fall in one severe week or spread thinly.
PRM and ARM — planning and anticipated reserve margin	Accredited capacity above forecast peak, as a percentage.	Compared against a reference margin level set so the system meets 0.1 LOLE.	Everything the accreditation step already decided. A margin is only as meaningful as the ELCC values behind it (Table 2A).
ELCC — effective load carrying capability	How much perfectly firm capacity a resource is equivalent to, given when it is available relative to when shortfall risk occurs.	Recalculated annually and by class; not a fixed property of a technology.	Interactions at scale. A resource's marginal contribution falls as more of the same type is added.

Table T6A — The adequacy metrics and what each one hides. No single metric is sufficient: LOLE counts events without weighing them, LOLH counts hours without weighing megawatts, and expected unserved energy weighs megawatt-hours without locating them in time. Assessments therefore quote all of them, and a reader should treat a single figure quoted alone as incomplete rather than as a summary.

Sources: NERC *Long-Term Reliability Assessment* and reliability assessment guidelines; PJM ELCC study documentation.

How the numbers are produced matters as much as what they measure. A probabilistic assessment draws a weather year from the historical record, applies a load forecast with its own uncertainty band, draws forced outages for each generating unit from its failure history, and then dispatches the fleet against demand hour by hour — repeating the whole exercise thousands of times to build a distribution of outcomes. SERC, for instance, runs thousands of simulations per hour across four decades of weather data. The output is not a prediction of a particular year but a frequency: the proportion of simulated years in which supply fell short. The one-day-in-ten-years standard is therefore a statement about the tail of that distribution, not a promise about any actual decade.

Large loads enter this calculation in three distinct ways, and only the first is widely discussed. As demand they raise the bar, since more accredited capacity is required to hold the same LOLE. Less obviously, they change the *shape* of demand: a conventional load fleet is diverse, so its aggregate peak is lower than the sum of individual peaks, whereas a fleet of facilities running near maximum around the clock has a coincidence factor approaching one and contributes almost its full nameplate to every hour, including the risk hours. And they interact with the weather draws that drive the simulation differently from ordinary load: a data center does not reduce its demand in a heat wave, so it adds to peak in precisely the hours when the supply side is most degraded.

The third path runs the other way. A load that can be curtailed on instruction is eligible to be accredited as a resource rather than counted only as demand, which moves it from the wrong side of the adequacy equation to the right one. This is the formal mechanism behind Section 6, and it explains why the accreditation reform described there mattered more than any physical change: PJM's ELCC revision moved roughly 1,800 MW of demand response into the cleared supply stack without a megawatt of new equipment. Adequacy planning supplies the arithmetic in which that reclassification takes effect.

CASE · NV ENERGY — FLEXIBILITY AS AN ADEQUACY RESOURCE

The GridLab–Telos study of NV Energy puts a number on that third path. The study treats data-center flexibility as a resource rather than as firm demand, crediting it toward the planning reserve margin through a demand-side effective load carrying capability. On that basis about 1 GW of flexible data-center load defers on the order of 665 to 865 MW of new firm capacity and saves roughly \$300–400 million through 2050. The saving depends on building the flexibility into resource planning from the start rather than adding it afterwards. The mechanism is exactly the reclassification described here: a load that can curtail moves from the demand side of the adequacy equation to the supply side, and the saving is the firm capacity no longer needed to hold the same reliability standard.

Three limits of the discipline should be held in view. Adequacy assessment assumes energy can reach the demand, so a system adequate on paper may still fail on deliverability — a transmission question handled in T.3, not here. It assumes fuel is available when dispatched, an assumption that winter events have repeatedly tested. And it rests entirely on the accreditation values discussed in Table 2A: change the ELCC assumptions and the reserve margin moves without any change in the physical fleet. When an assessment reports a deteriorating margin, the first question is whether the fleet changed, the load changed, or the counting changed.

T.7 Scenario analysis: planning when the forecast itself is uncertain (supports Sections 1 and 2)

Why this section exists, and how scenarios are used. Scenario analysis does not improve a forecast; it makes decisions without one. It separates the choices robust across futures from those depending on one future arriving, so the first can proceed now and the second waits for a decision point. Applied to this report's subject, a transmission line useful under both high and low load growth differs in kind from a generating unit whose economics require the high case, and the discipline consists in identifying which is which before committing capital. Scenarios also let a planner avoid the trap of the central estimate: a plan optimised for the reference case frequently proves the worst under both tails.

When the central forecast can be off by a factor of three, planning to a single number is hard to defend. Scenario analysis plans across a range instead — building a small set of internally consistent futures, testing the system in each, and looking for the decisions that hold up across all of them. For large loads, three dimensions dominate the scenario space:

- Growth rate. Reference, high, and low demand trajectories — typically differing by whether speculative load materializes. A reference case near ~2.8%/yr can fan to ~5%/yr (high) or ~1.2%/yr (low), a spread of roughly 110 GW nationally by 2035.
- Geographic concentration. The same total load produces very different system impacts depending on where it lands. A case that clusters gigawatts in one weak zone stresses local transmission and system strength far more than one that disperses them, so zone-level and nodal scenarios matter more for data centers than for diffuse load.
- Operational assumptions. Flexibility (how many hours of curtailment loads will accept), ramp behavior, coincidence with system peak, and behind-the-meter generation dispatch. These swing both the peak and the reserve value of the same nameplate load.

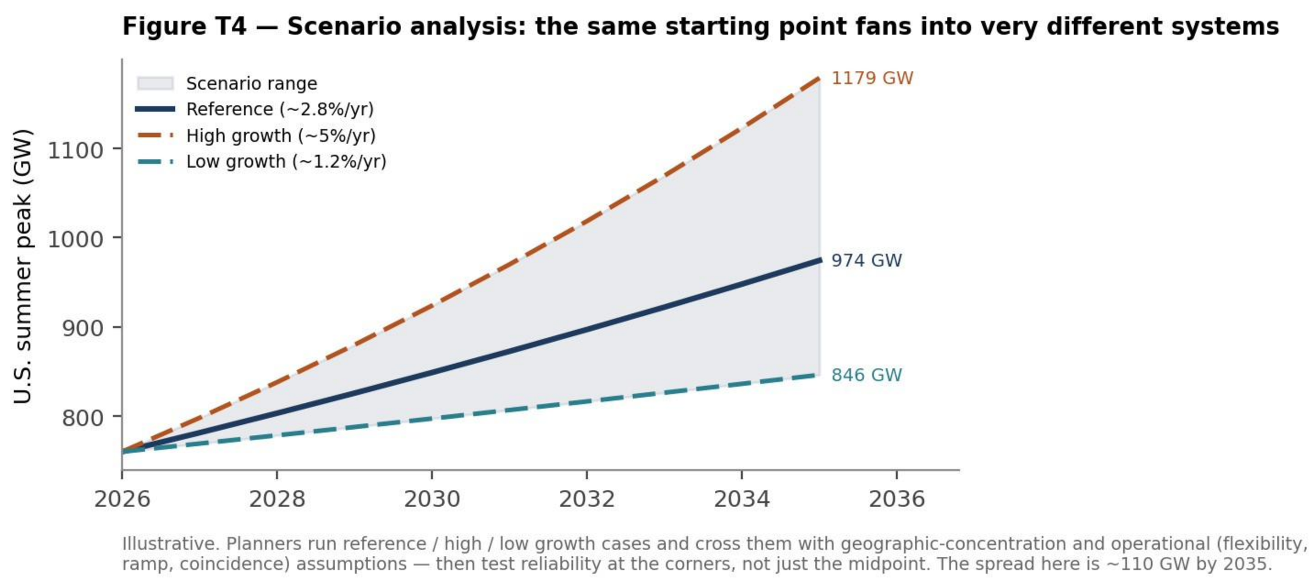


Figure T4 — Scenario analysis in one picture: a single starting point fans into materially different systems depending on assumptions about load growth, resource cost and policy. The spread matters because a plan optimised for the central case frequently proves the worst under both tails. The implication for a planner is that scenarios are not a way of improving a forecast but a way of identifying which commitments are robust across futures and which depend on one arriving — so reliability should be tested at the corners of this space, not only along the reference line.

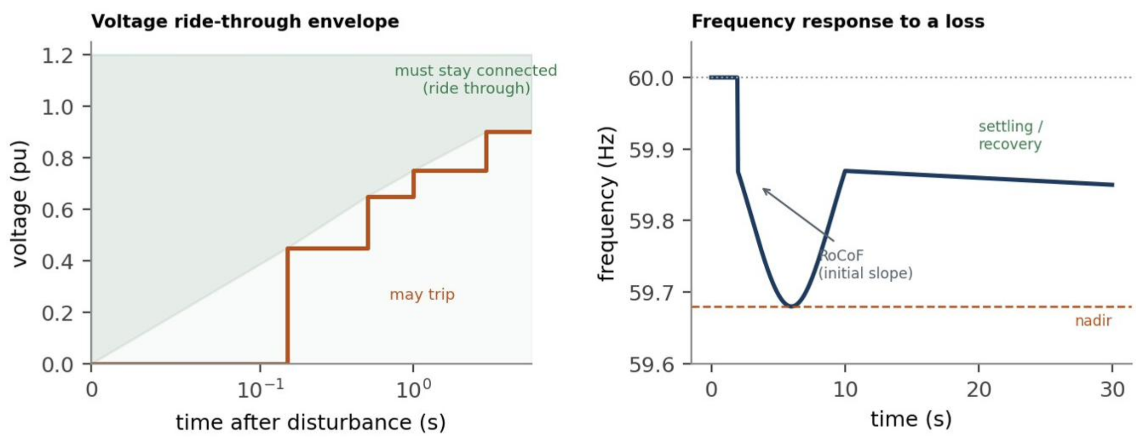
The purpose is not to predict which scenario occurs but to find robust decisions and no-regret moves — upgrades and reforms that pay off across the range — and to identify the scenarios in which the system fails, so those can be monitored and hedged. Flexible-load service (Section 6) attracts precisely because it comes close to no-regret: it helps in the high-growth case and costs little in the low. NERC's May 2026 guideline pushes exactly this way, asking for resource-adequacy studies evaluated across many weather, load, and outage combinations on a network-aware footprint rather than against a single deterministic peak.

T.8 Reliability assessment: the metrics a study must satisfy (supports Sections 4 and 2)

Why this section exists, and what the criteria do. The metrics here are the pass-fail conditions that convert a simulation into a decision. A stability study produces voltage and frequency traces; the criteria in this section state which traces are acceptable, and they supply the operative content of any performance obligation. This matters for Section 4 in a specific way: a ride-through requirement is nothing more than a stated voltage-duration envelope that equipment must remain connected through, so the argument about whether such requirements are reasonable is an argument about where to draw the boundary of that envelope. The contingency categories below serve the same function for the network, defining which events a planner must design against and which are treated as beyond the standard.

Reliability assessment is where all the preceding analysis is scored: given a future load condition, does the system stay within its steady-state and dynamic limits under the required contingencies? The checks fall into two families. Steady-state assessment confirms that, in the base case and under each contingency, no element exceeds its thermal rating and every bus voltage stays within limits (typically 0.95–1.05 per unit). Dynamic assessment exposes large loads. It evaluates the system's time-domain response to disturbances against three families of metric. Transient, or angular, stability asks whether machines stay in synchronism. Frequency response covers the rate of change of frequency (RoCoF) immediately after a loss, the frequency nadir, and the settling value; all three worsen as inverter-interfaced load displaces inertia. Post-fault voltage recovery covers fault-induced delayed voltage recovery (FIDVR), where stalled motors or tripping loads hold voltage down after the fault clears. Small-signal (oscillatory) stability draws increasing concern because synchronized AI workloads can act as a forcing function that excites weakly damped inter-area modes (T.2).

Figure T5 — Reliability assessment: the envelopes a study checks — voltage ride-through and frequency response



Steady-state checks (thermal, voltage 0.95–1.05 pu) plus dynamic checks: transient stability, RoCoF, frequency nadir, and post-fault voltage recovery (FIDVR). A large load that trips — or fails to ride through — shows up as a violation in exactly these plots.

Figure T5 — The two envelopes a reliability study checks. Left: the voltage ride-through envelope a large load must stay connected through. Right: the frequency response to a sudden loss, with the RoCoF, nadir, and recovery a study evaluates.

Protection coordination. The envelopes above describe when a load should stay connected; protection settings decide when it will actually trip, and the two do not coincide. A large computational load sits behind layered protection — utility relays at the point of interconnection, the facility's own main and feeder breakers, and the fast electronic protection inside every UPS and power supply — which must be coordinated so that a fault clears at the smallest affected zone rather than disconnecting the whole campus. Three settings dominate the recent ride-through record. Undervoltage and under-frequency trip thresholds, often left at a vendor default, can fire well inside the envelope that PRC-024 and the newer PRC-029 require loads to hold. Anti-islanding protection, meant to disconnect on-site generation when the grid is lost, can misread a deep transmission voltage sag as an island and trip. And breaker-and-relay timing graded for a conventional load — a downstream device clearing before an upstream one — can invert when converter front ends inject or withhold fault current in ways the coordination study never assumed. The consequence is that ride-through is a protection problem as much as a control problem: a load can carry a compliant controller and still trip on a mis-set relay, and a coordination study built on static, motor-dominated load will not catch it (Technical Annex T.3).

The contingency set is defined by TPL-001, which enumerates event categories from the everyday (P0, no contingency; P1, loss of a single element) through more severe multiple-element events (up to P7). For most categories the standard permits no non-consequential load loss — which explains why an uncommanded large-load trip during a normally-cleared fault (Section 4) raises a reliability-standard problem rather than an inconvenience: the criteria say that load loss should not occur. The through-line of this annex is that a reliability assessment is only as trustworthy as the load model feeding it (T.4), the contingencies it considers (T.3), and the scenarios it spans (T.7). Get those right and the metrics in this figure tell you whether the future grid holds; get them wrong — as a static load model does — and the study will certify a system that will not.

CASE · JULY 2024 — 1,500 MW LOST TO PROTECTION SETTINGS

The event behind this section’s protection discussion is NERC’s review of a July 10, 2024 disturbance in the Eastern Interconnection. A failed arrestor on a 230 kV line, with auto-reclosing set for three attempts at each end, produced six faults in 82 seconds, each cleared normally in 42 to 66 milliseconds, with voltage falling to 0.25–0.40 per unit in the affected area. No utility equipment shed load. Instead about 1,500 MW of data-center load disconnected itself on the customer side as UPS systems transferred to backup on the voltage dips. The loss pushed frequency to 60.047 Hz and voltage to 1.07 per unit, and operators removed capacitor banks to recover. The lesson is the one this annex returns to: the load rode through nothing, because its own protection was set to trip well inside the envelope the grid needed it to hold. The event fell in PJM/Dominion territory rather than ERCOT; NERC’s May 2026 Level 3 Alert cites comparable losses in the Texas interconnection as well.

T.9 Distribution-system impacts: the layer the bulk-power debate skips (supports the Primer and Section 3)

Why this section exists. The whole report is written about the bulk power system because that is where gigawatt campuses connect, and this section covers the consequence of that focus. Two effects run in opposite directions. A facility connecting at transmission voltage bypasses the distribution system, contributing little to its cost while still influencing the wholesale prices that distribution customers pay — a cost-allocation question rather than an engineering one. Meanwhile the faster-growing population of smaller sites lands directly on feeders designed for diverse residential and commercial demand, where the planning tools, the protection settings and the hosting-capacity assumptions were all calibrated on a load that behaved differently. The bulk-system reforms in Sections 1 through 6 reach none of these smaller facilities.

Almost everything in this report concerns the bulk power system — transmission, wholesale markets, and systemwide reliability. That focus holds for the gigawatt campuses that dominate the headlines, because a load that large connects directly at transmission or sub-transmission voltage and largely bypasses the distribution system altogether. But that very fact creates two distinct distribution-level stories worth naming.

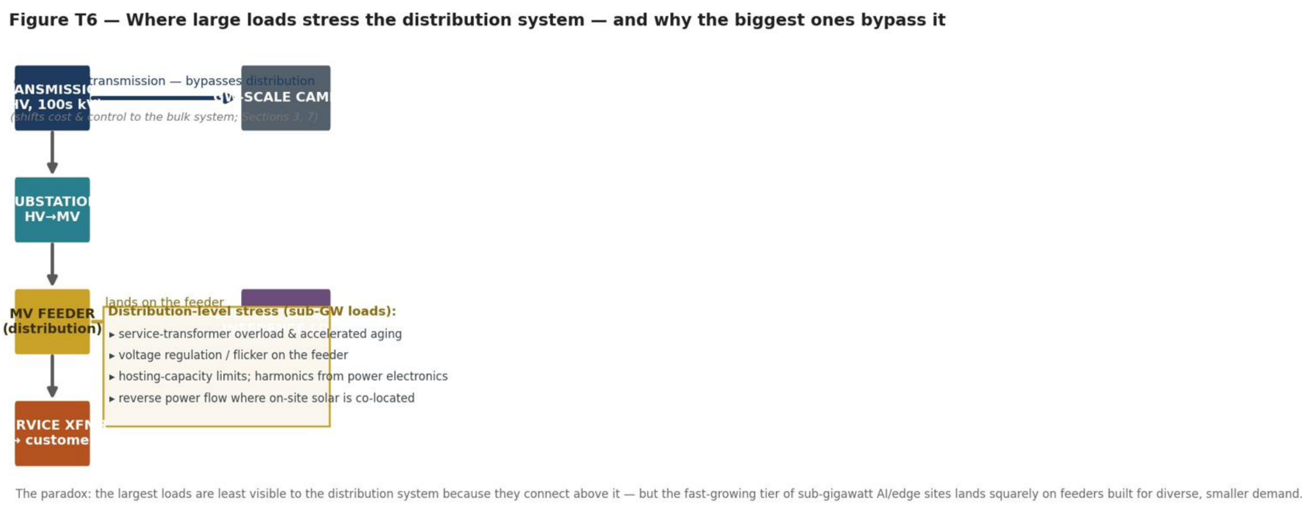


Figure T6 — The largest loads remain least visible to the distribution system because they connect above it; the fast-growing tier of sub-gigawatt AI, edge, and inference sites lands squarely on feeders built for smaller, diverse demand.

The first story concerns what the bypass itself does. When a hyperscale campus interconnects at transmission, it contributes little to distribution-system cost yet still shapes the wholesale prices and capacity charges that flow through to distribution customers' bills — which is part of why the cost-allocation fight in Section 3 plays out at the wholesale and state-tariff level rather than in distribution rate cases. The load is physically absent from the feeder but financially present on the bill. The second story is the fast-growing tier that does not bypass distribution: sub-gigawatt colocation, “edge” facilities, and AI-inference sites in the tens-to-low-hundreds of megawatts, which increasingly land on medium-voltage feeders and distribution substations that were planned for diverse, smaller commercial demand. There the classic distribution problems reappear at new intensity. Service transformers and feeders overload and age faster. Voltage regulation and flicker suffer as large, fast-varying loads swing; hosting-capacity limits and harmonic distortion injected by the same power-electronic front ends discussed in T.1; and, where on-site solar or storage is co-located, reverse power flow that distribution protection was never designed for. Utilities are responding with distribution-level hosting-capacity analysis, interconnection screens, and — as at the transmission level — flexible or curtailable connection terms. The distribution–transmission coordination gap poses the subtler risk: a load small enough to count as a distribution matter can still, in aggregate with its neighbors, present a transmission-scale contingency, and the two planning worlds do not always talk to each other. This is the same firm-versus-flexible and modeling discipline as the rest of the annex, applied one voltage level down.

For a nonspecialist the useful correction is that a gigawatt of new load is not automatically a transmission problem. Which element binds first depends on where and how the load connects. A campus that interconnects at extra-high voltage may never touch a distribution feeder, while a cluster of inference sites at medium voltage can exhaust a substation transformer, a feeder breaker’s interrupting rating, or the station bus long before any transmission limit appears. The screening order runs from the device outward — transformer and breaker ratings, then feeder thermal and voltage-regulation limits, then the substation bus — and only then to the bulk-power constraints the rest of this report treats. The debate defaults to transmission because that is where the largest campuses connect; the fast-growing tier does not, and its first constraint usually sits one voltage level down.

T.10 Planning paradigms and tools: deterministic vs. probabilistic, and how they fit together (supports Sections 1, 2, and 9)

Why this section exists. The preceding sections describe tools that answer different questions in incompatible languages, and a reader encountering them separately can reasonably conclude they conflict. They do not. A deterministic study asks whether a specified condition is survivable and returns yes or no; a probabilistic assessment asks how often adverse conditions arise and returns a frequency. Neither substitutes for the other: a probabilistic result cannot tell an engineer whether a particular bus stays within its voltage band, and a deterministic study cannot tell a planner whether the condition it examined is likely enough to warrant the cost of preventing it. The practical relationship is sequential — probabilistic methods identify which conditions merit deterministic examination, and deterministic studies establish what must be built if those conditions occur.

The report introduces forecasting (T.5), adequacy (T.6), and scenarios (T.7) separately; this part gives the nonspecialist the single frame that connects them. Grid planning has been undergoing a slow migration from a deterministic to a probabilistic paradigm, and the large-load problem is accelerating it. Deterministic planning tests the system against a fixed set of defined conditions — a forecast peak, a specified list of contingencies (the N-1/N-1-1 events of TPL-001), a design-day weather assumption — and asks whether it holds in each. It stays transparent and auditable, exactly what a reliability standard needs — the reason the contingency framework in T.3 and T.8 runs deterministic at its core. It says nothing about how likely each condition runs, and combinations no single scenario captured can blindside it. Probabilistic planning instead runs the system across a distribution of conditions — many weather years, forced-outage draws, and load realizations — and reports outcomes as likelihoods: the loss-of-load expectation, hours, and unserved-energy metrics of T.6 are inherently probabilistic. The two are complements, not rivals: deterministic analysis sets the pass/fail reliability floor, probabilistic analysis sizes the margin and prices the risk. Large, uncertain, flexible loads push planners toward the probabilistic end because a single deterministic peak cannot represent a load that might materialize or might curtail. Three tool families do the work, each answering a different question:

- Power-flow and stability tools (deterministic). Answer “does the network hold under this condition?” — the steady-state, short-circuit, transient, and EMT studies of T.3–T.4. Used for interconnection and reliability compliance.
- Production cost models (PCM). Chronological simulations of the whole market — hour-by-hour (or finer) unit commitment and economic dispatch across a year — answering “what does the system cost to run, where does congestion bind, and what are the emissions and prices?” Planners use such tools to test whether a large load raises congestion or curtails renewables, and how flexible load changes dispatch. They sit between deterministic and probabilistic: one PCM run stays deterministic, but planners sweep many to explore the range.
- Resource-adequacy models (probabilistic). Monte Carlo simulations (the ProbA of T.6) that answer “is there enough accredited capacity, and how often will we fall short?” — producing LOLE/LOLH/EUE and the ELCC accreditations that decide whether a flexible load counts as a resource. A well-run planning process uses all three and the scenario discipline of T.7 on top: deterministic studies to certify reliability, production cost modeling to understand economics and congestion, and probabilistic adequacy modeling to size reserves — each run across a range of futures rather than a single forecast. The large-load era's methodological lesson is simply that the deterministic single-number habit is no longer sufficient on its own; the uncertainty is now too large to leave unquantified.

Horizon	The decision it answers	Tool or model	Where
Seconds to minutes	Hold frequency and voltage instant to instant, and ride through disturbances	Automatic generation control; operator action	T.11
Minutes to an hour	The cheapest safe dispatch of the units already committed	Security-constrained economic dispatch	T.11
Hours to days	Which units to start, stop, and hold in reserve	Unit commitment; production-cost models	T.10
One to five years	Whether accredited capacity meets the reliability standard	Resource-adequacy (probabilistic) assessment	T.6
Five to twenty years	What to build and retire, and which lines to add	Integrated resource plans; transmission planning	T.5, T.7

Table T10A — The same grid, planned on five clocks. Each horizon carries its own tool and its own question; much of the large-load debate confuses the two — answering a years-scale adequacy question with a seconds-scale operations instinct, or the reverse. The tools complement rather than compete, and the annex sections in the final column treat each in turn.

T.11 Operational impacts: what large loads do to the control room (supports Sections 4 and 6)

Why this section exists. The preceding annex sections describe how large loads are studied, forecast, modelled and planned for. None describes what happens once the load is energised and an operator has to run the system around it. The distinction matters because the planning tools work in hours and years while the control room works in seconds and minutes, and the three characteristics that make computational load hard to plan for — magnitude, speed, and opacity — bear on operations more immediately. Four places where that shows up are sketched below. This is a map of where the operational discussion is heading in NERC and ISO forums rather than a settled account; the standards work described in Section 4 will determine most of it.

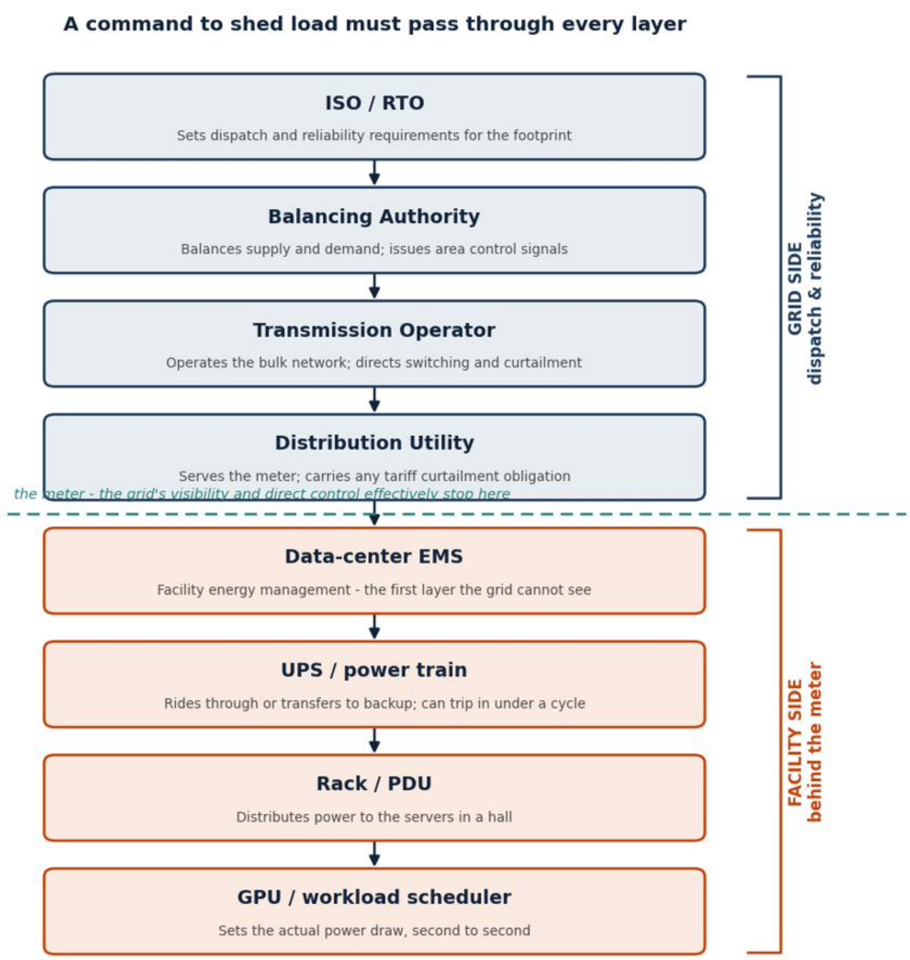


Figure T8 — The control hierarchy: who actually commands a large load. A curtailment instruction must pass through four grid-side layers and four inside the facility. The grid's visibility and direct control effectively end at the meter; below it, the load's second-to-second draw is set by equipment and a workload scheduler the operator cannot see — which is why a load's flexibility is usable only when the settings below the meter are configured to honour it.

- Situational awareness and the EMS. An operator acts on what the energy management system shows, and computational load ranks as the least visible large object on the system: it sits behind a customer meter, its internal state is not telemetered, and its next move is set by a workload scheduler the operator cannot see. State estimation and contingency analysis run against a model carrying the load as a static MW value, so the telemetry and registration requirements of Section 4 pose an operations problem as much as a planning one. Until the control room can see a campus's ramp capability and its ride-through settings, the three items below are being managed from an incomplete picture.

- Ramping and dispatch. Economic dispatch solves every five minutes against a load forecast built from historically smooth behaviour. Training workloads step between power states in seconds and can move tens or hundreds of megawatts inside a single dispatch interval, so the ramp reaches the market as forecast error and is covered by the fastest and most expensive resources on the stack. The mitigation is the flexibility of Section 6 read from the other end: a load that shapes its own ramp serves the dispatch better than one that must be chased.
- Regulation and AGC. Automatic generation control corrects the second-to-second imbalance that dispatch leaves behind, acting on area control error — the combined measure of frequency deviation and net interchange error that records whether a balancing authority is carrying its own load. Regulating reserve is procured against the expected variability of net load. Where a campus swings faster than the regulating fleet responds, ACE degrades and the balancing authority's control performance measures (BAL-001, BAL-003) move in the wrong direction — a compliance exposure before it becomes a reliability one.
- Reserve deployment and the loss-of-load contingency. Contingency reserve is sized against the largest single contingency, historically a generator or a tie line. A gigawatt campus that trips off — or rides through and then returns — presents a contingency of comparable size at the other end of the balance, one that can repeat within minutes rather than once a day. The operational question is therefore not only how much reserve is carried, but whether reserve deployed for a load-side event can be restored before the next one, which is the control-room form of the ride-through argument in Section 4.

Sources: NERC reliability standards BAL-001 and BAL-003; the large-load registration and modelling work of Project 2026-01 and RD26-7-000; NERC and ISO workshop material on data-center operating characteristics. This section is an engineering synthesis rather than a report of filed positions.

Latest Updates

This section logs every change to the report since publication. Entries are newest first. Every item is dated, carries the same confidence tag used throughout, and names the sections it revises — and each one has already been worked into the body at the point of use, so the report reads as a single current document rather than a base text plus errata. This page indexes the changes rather than housing them.

Revision history

Version	Date	What changed	Sections revised
v1.0	Jul 13, 2026	Original publication — the eight issues, timeline, technical annex.	—
v1.1	**Jul 17, 2026**	FERC Docket RD26-7-000 makes the NERC standards + registration schedule binding; PJM posts a second consecutive capacity shortfall (2028/29 BRA); New York moratorium takes effect by executive order; White House moves to broaden the Ratepayer Protection Pledge. Context note added on *Trump v. Slaughter*. Corrections: ERCOT's NOGRR282 / NPRR1308 ride-through obligations restated as mandatory within ERCOT rather than absent, and the unaddressed ramp-rate gap added to §4.	Key Takeaways; §2; §3; §4; §7; §8; §9; Watch List; Catalog; References
v1.2	**Jul 21, 2026**	July 20 resource-adequacy informational reports come due. CAISO's (EL26-71) read into §2 — no systemic California adequacy gap, with the reasons it may not generalise — against MISO's (EL26-70), which proposes an out-of-queue study process and non-firm service. ISO-NE signals a 90-day abeyance and CAISO's calendar assumes one, qualifying the common August 17 deadline. Procedural calendar added to the Federal Frame; PJM backstop procurement window dated and the 2028/29 shortfall stated at 6,831 MW.	Key Takeaways; Report at a Glance; Federal Frame; §2; §7; Watch List; Latest Updates
v1.21	**Jul 21, 2026**	§3: the Ratepayer Protection Pledge binds hyperscalers rather than colocation developers or utilities, and a commitment to pay is limited by whether the tariff bills incrementally.	§3; References
v1.22	**Jul 21, 2026**	Reviewer pass. CAISO's filing reduced to a comment and restated to what it says; duplicate PJM backstop watch-list entry consolidated; fourth confidence tag (Speculative) added; compound modifiers standardised; T.11 added on operational impacts; T.4 extended with study practice — RMS versus EMT, load-model parameterisation, inverter-model validation, and the assumptions that invalidate a study. Table K1 added to Key Takeaways: the interconnection sequence end to end, with the failure mode and the governing section at each step.	Key Takeaways; §2; Watch List; Figure D2; Technical Annex
v1.23	**Jul 21, 2026**	Editorial: reference and catalog counts now computed at build time rather than stated by hand; the guiding principle and the note on judgment set as callout boxes on the About page.	About This Report
v1.24	**Jul 26, 2026**	Technical-annex expansion and a register pass. Glossary gains grid-forming and grid-following inverters, RMS/positive-sequence simulation, protection coordination, hosting capacity, coincident/non-coincident peak, diversity factor, and capacity accreditation; N-2 folded into the N-1 entry. T.8 gains a protection-coordination paragraph; T.9 a note on which network element binds first; T.4’s simulation-method ladder folds into the Figure T2 caption; new Table T10A maps planning horizons to tools. A copula-reduction pass runs across T.1–T.11. No conclusion changes.	Technical Annex; Glossary; Latest Updates
v1.25	**Jul 26, 2026**	Five case studies added as boxed notes, each anchored to a primary proceeding: SPP CHILLS conditional service and Ireland’s CRU self-supply connection policy (T.3), PJM’s speculative large-load forecast (T.5), the GridLab–Telos NV Energy flexibility study (T.6), and NERC’s review of the July 2024 voltage-sensitive load loss (T.8). Two figures added: T7 coincident versus non-coincident peak (T.5) and T8 the control hierarchy (T.11). Editorial: About-page title corrected; guiding-principle premise reworded for clarity; a dangling reference fixed; K1’s “Stability” step relabelled “Transient stability”; a Synthesis/Central Findings row added to the structure table; glued-word typos repaired. No conclusion changes.	About This Report; Key Takeaways; Technical Annex; Latest Updates

Version	Date	What changed	Sections revised
v1.26	**Jul 26, 2026**	Register and scope pass. The executive opening no longer says the flat-demand rules “no longer hold” (they still function) but that they “are no longer sufficient” for today’s large-load growth, and “essentially each major reliability institution” becomes “the major North American reliability institutions.” The siting-constraint claim is narrowed to “many hyperscale data-center developers.” Two “Planning implication” closers added, to T.1 and T.4, tying the engineering back to the main report. A glued-word typo repaired. No conclusion changes.	Executive Summary; Watch List; Technical Annex; Latest Updates
v1.27	**Jul 31, 2026**	§5: the Susquehanna and Crane record added from primary sources — the ER24-2172 rejection and its actual holding, the front-of-the-meter restructuring of both nuclear transactions, the November 2023 fallback event, and the June 1, 2026 transfer of 760 MW of Capacity Interconnection Rights from Eddystone to Crane. New Table 5B compares the two arrangements. The Figure 9 caption notes the network-service-plus-bilateral-contract configuration the figure omits. §4: the December 2022 west Texas event isolated in a note — delayed clearing, mixed load, and a name shared with NERC's Odessa IBR reports — with July 2024 northern Virginia stated as the ride-through archetype. §6: the Duke curtailment rates restated as shares of energy rather than of time (85/177/366 hours, 76/98/126 GW), and the study's own network-constraint limitation recorded. New subsection on who directs the curtailment; Finding 5 carries the trade-off.	Key Takeaways; Central Findings; §4; §5; §6; Catalog; References; Latest Updates
v1.28	**Aug 1, 2026**	The three items carried forward from v1.27, each verified against a primary document: Crane added to §2 as the case where deliverability rather than construction binds; the $140M Susquehanna cost-shift exchange added to §3, with Susquehanna’s forgone-revenue answer; and the December 2025 order’s express declination on jurisdiction over retail loads served through co-location added to §7. Correction in §5: the order directs a choice among four transmission services, three of them new, including an interim non-firm service the report had omitted. Editorial: sentence-length pass on the longest 14 sentences (worst 141 words), and a density pass on Synthesis.	Executive Summary; Key Takeaways; Synthesis; §2; §3; §5; §7; §9; International; Annex; References
v1.29	**Aug 1, 2026**	§2 gains a proposed-solutions bullet the Crane paragraph had left implicit — reuse of an existing interconnection position, grounded on Order No. 845 surplus interconnection service, with the Eddystone-to-Crane waiver as the worked case and two limits stated: the instrument reallocates deliverability rather than creating it, and surplus service ends with the agreement it depends on. Editorial: the §4 note opened with four pronouns in one sentence and now does not; Finding 5 drops “only” and “the one lever”, states the direction-of-control trade-off in a sentence, and sends the explanation to §6.	Central Findings; §2; §4; References
v1.30	**Aug 1, 2026**	Register pass on text added in v1.27–v1.29. Eleven corrections, none touching a finding. Removed: antithesis and chiasmus constructed for cadence in Finding 5 and §6; one literary allusion; commercial metaphor where the subject was a regulatory outcome rather than a transaction; and four vague predicates. The register standard is the one stated in the README — institutional assessment rather than journalism — and the corrected passages now state the finding directly.	Central Findings; §2; §4; §5; §6

Version	Date	What changed	Sections revised
v1.31	**Aug 1, 2026**	Reviewer pass. Sourcing: the MISO 43% figure attributed to FERC's 2025 State of the Markets report (March 19, 2026) at first use, with the 24% national rate for scale and a statement that every occurrence draws on that one dataset; the PJM clearing-price series set out auction by auction, with the compressed 2025 calendar explained; the ERCOT 2.2 GW and ~9 GW figures distinguished as recent and cumulative in the Executive Summary; and the three EIA growth rates identified as coming from three products measuring three windows. Structure: the three forecasting families split into three paragraphs, and the Executive Summary's capacity-market paragraph divided. Register: two over-compressed sentences expanded. Typography: body paragraph spacing raised from 5 to 7 points against 13.2-point leading, after diagnosis showed the Executive Summary's three ideas already sat in three paragraphs and read as one only because the break was too faint.	Executive Summary; Federal Frame; §1; §3; Primer; References; render_pdf.py

Table L1 — Revision history. One row per revision, with the sections touched. The blocks below expand each entry; the body already carries every change.

Sources: FERC; NERC; ERCOT; PJM; DOE, as listed in the Catalog of Orders, Rules, and Directives Cited.

v1.2 — developments of July 20–21, 2026

The July 20 resource-adequacy reports came due, and two of the six were recovered in time for this revision — one from the filing itself, one from reporting of it. Taken together they are the first evidence of what FERC's regional route produces in practice, and the two answers point in different directions. CAISO reports no systemic generation-adequacy shortfall in California and attributes it to an integrated forecasting, procurement and planning framework; §2 carries that as a comment rather than a finding, since the region carries a fraction of the demand pressure and its procurement mandate is a feature of California's regulatory structure rather than a tariff term available to an RTO. MISO answers the same question in the opposite direction, and consistently with its position in the record: an expedited study process for large loads run outside the interconnection queue as the ERAS fast lane winds down, a non-firm transmission service option, and a published target of approval within 120 days once studies are complete. The region reporting no adequacy gap proposes least; the region with the fastest data-center growth proposes most. §2 now carries both. Two further developments are procedural. CAISO's published calendar ends in a November 16 filing and states that it assumes an abeyance, so two of the six regions now look likely to answer in mid-November rather than on August 17. And PJM's backstop procurement acquired dates — a window of September 10 to October 9, 2026 in the June 30 design paper, with cost allocated to the load additions that created the need — while the 2028/29 shortfall is now stated precisely at 6,831 MW against a 14.7% reserve margin. **Neither item reverses a conclusion of the report.** One narrows the geography of a finding; one qualifies the uniformity of the federal deadline, in the direction Section 7 already argued.

Development	Date	Effect on this report	Tag
RTO/ISO resource-adequacy informational reports come due (EL26-67 – EL26-72)	Jul 20, 2026	The 30-day filings required by the June 18 orders came due. CAISO's is read in §2; the other five had not reached public analysis at this revision. Confirms the timeline in §2 and adds the first on-the-record regional answer to the question of whether the queue can be served.	SETTLED
CAISO informational report (EL26-71) — no systemic California adequacy gap	Jul 20, 2026	Reports no systemic generation-adequacy shortfall and attributes it to California's integrated forecasting, procurement and planning framework; commits further work to its own stakeholder initiative rather than arguing that none is needed. Recorded as a comment in §2: the region carries a fraction of the demand pressure, so the contrast with the auction regions is institutional rather than a counter-example.	SETTLED
MISO informational report (EL26-70) — out-of-queue study process and non-firm service	Jul 20, 2026	The opposite answer to CAISO's, from the region with the fastest data-center growth: an expedited study process for large loads outside the interconnection queue as ERAS winds down, a non-firm transmission service option, and a 120-day approval target. Read from reporting of the filing and MISO's own programme materials; the filing itself was not retrieved.	**LIKELY**
PJM reliability backstop procurement acquires dates	Jun 30 / Jul 14, 2026	Design paper sets a procurement window of Sep 10 – Oct 9, 2026 with cost allocated to the load additions that created the need; the 2028/29 shortfall is 6,831 MW against a 14.7% reserve margin. Puts dates on an instrument §2 and §3 had described without them.	SETTLED
ISO-NE and New England transmission owners announce they will seek a 90-day abeyance	Jun 29, 2026	Qualifies the Federal Frame and Section 7: the show-cause orders were simultaneous but the responses will not be. An abeyance requested by Aug 3 and granted would move ISO-NE's substantive filing to roughly mid-November. FERC has said it will not grant abeyance as a matter of course, so the request tests how firm the common deadline is.	**LIKELY**

Table L2 — Developments of July 20–21, 2026, and what each does to the report.

Sources: FERC show-cause orders of June 18, 2026 (EL26-67 through EL26-72); ISO New England compliance notice, June 29, 2026.

v1.21 — the reach of the voluntary commitments

One addition to Section 3, prompted by a question the report had not answered directly: what the hyperscalers' own cost commitments actually reach. The Ratepayer Protection Pledge binds seven consuming companies, not the colocation developers who own much of the capacity in the queue and not the utilities that divide the costs — and even a willing payer is billed by a tariff that, under the Commission's long-standing transmission pricing policy, may recover the upgrade from everyone. **Neither point changes a conclusion.** Both sharpen why Section 3 treats the pro forma cost-recovery agreement, rather than any voluntary undertaking, as the deciding instrument.

Sources: The White House; Google; Harvard Electricity Law Initiative; Brookings.

v1.22 — editorial pass and a new annex section

An editorial pass on reviewer comments, with one structural addition. The account of CAISO's July 20 filing was cut to a comment and restated: it reports no systemic adequacy shortfall and attributes that to California's integrated forecasting, procurement and planning framework, but it does not argue that no further reform is needed, and it commits the additional work to its own stakeholder initiative — a distinction the previous wording blurred. MISO's filing now leads the passage, which is the right order of importance. A duplicated PJM backstop entry on the near-term watch list was consolidated and two rows put back into date order. The confidence key gains a fourth tag, **Speculative**, for genuine forecasting: Emerging had been carrying proposals, forecasts and opinions at once, and the long-horizon watch-list items have been retagged accordingly. Figure D2's caption now ends on a bolded Finding line. And the technical annex gains T.11 on operational impacts — situational awareness and the EMS, ramping and dispatch, regulation and AGC, reserve deployment and the load-side contingency — the one subject area the report described in planning terms without ever reaching the control room. **No conclusion changes.**

v1.27 — the two arrangements behind Section 5

Section 5 set out the co-location rules without naming the transactions that produced them. This revision adds that record from the primary documents. Three findings change how the section reads. The Susquehanna amended ISA was rejected for tracking PJM's generally applicable guidance rather than for any substantive defect, and expressly without prejudice, so the December 2025 order answers a question the Commission had declined to reach case by case. Both nuclear arrangements large enough to test co-location — Talen and Amazon at Susquehanna, Constellation and Microsoft at Crane — now take front-of-the-meter network service, and both settled on that structure before the replacement rate existed. And the Capacity Interconnection Rights adjustment that the December order treats as a write-down operates elsewhere as a transferable asset: the June 1, 2026 waiver moved 760 MW from Eddystone to Crane and advanced the restart by about three years. A November 2023 fallback event at Susquehanna, in which the co-located load drew from the grid for several hours during a unit outage, is added to the discussion of host unavailability as the only operating record the debate has.

Three further corrections followed on review. In Section 4, the December 2022 west Texas event now sits in a note of its own. It is the largest in the record and the least representative of it: the fault escalated to a three-phase fault caused by a breaker failure, so the clearing was not normal, and the load was mixed, with substantial losses from oil and gas facilities whose drives trip on undervoltage for reasons unrelated to the ITIC curve. The name compounded the difficulty, since NERC published its Odessa Disturbance reports about solar inverters rather than load; this report now calls the load event by its region. The July 2024 northern Virginia event is stated as the archetype the ride-through rules address. In Section 6, the Duke curtailment rates are restated on the basis the study actually uses — shares of annual energy, not of running time — which puts the hours in which some curtailment falls at 85, 177 and 366 rather than 22, 44 and 88, mostly as partial reductions, and the headroom at 76, 98 and 126 GW. The study's own first-listed limitation, that it does not model network constraints, is now recorded wherever the figure appears. And a new subsection distinguishes curtailment an operator calls from deferral the load schedules for itself, which is the trade-off Finding 5 now carries.

v1.1 — developments of July 13–17, 2026

Four actions landed in five days, and all four point the same way the report already did: the reliability track (registration plus standards) now carries a federal directive and a hard deadline, while the political-economy track is hardening faster and less predictably than the tariff track. **Nothing in this update reverses a conclusion of the original report.** One item changes a confidence tag; one corrects a factual detail; two reinforce existing assessments.

Development	Date	Effect on this report	Tag
FERC Docket RD26-7-000 — mandatory Reliability Standards and registration for computational loads	Jul 16, 2026	Raises the standing of a central reliability reform. NERC must file standards AND Rules-of-Procedure registration criteria by Dec 31, 2026, plus a Phase II work plan by Mar 1, 2027. The obligation and its dates are now federal mandate, not NERC's internal plan.	SETTLED
PJM 2028/2029 Base Residual Auction	Jul 14, 2026	Second consecutive shortfall (~6.8 GW, widened from ~6.5 GW); cleared at the $325/MW-day cap; only ~525 MW of new generation cleared. Confirms §2's timeline gap and keeps §3's cost fuse lit.	SETTLED
New York statewide hyperscale moratorium takes effect	Jul 14, 2026	Executive order pausing discretionary state environmental permits for facilities ≥50 MW for up to one year, pending a GEIS. Corrects §7, which had it as pending legislation.	SETTLED
White House moves to broaden the Ratepayer Protection Pledge	Mid-Jul 2026	Convening utilities, developers, and governors around a wider voluntary pledge. Reinforces §7's read that voluntary federal signaling is not slowing state constraint.	**EMERGING**
Trump v. Slaughter (context; predates v1.0)	Jun 29, 2026	Supreme Court overruled *Humphrey's Executor*; FERC's organic statute carries the same for-cause removal language. Does not alter FERC's authority — alters the independence backdrop behind §7.	SETTLED

Table L3 — Developments of July 13–17, 2026, and what each does to the report.

Sources: FERC; NERC; ERCOT; PJM; DOE.

What the update changes, in one paragraph

RD26-7-000 matters most, because it acts directly on a central reform in the current reliability program. The Synthesis identifies NERC registration, rather than any tariff, as the discipline that reaches a large load wherever it sits, including when it islands off-grid (§4, §8). Until July 16 that reform rested on NERC's own accelerated plan and was tagged *Likely*. It is now a federal directive with two dates attached, which moves the existence and timing of the obligation to *Settled* while leaving the registry threshold and the standard's substance genuinely open. The second PJM shortfall, meanwhile, is the first hard market evidence that the 2028–2031 stress window in §9 is materialising rather than remaining a forecast — and the New York moratorium, arriving by executive order in the same week as a broadened federal pledge, is §7's community-consent thesis playing out in real time.

Filing the next update

Revisions within this publication cycle are numbered v1.21, v1.22 and so on, so that the second decimal marks an update to the current edition rather than a new one. Each adds one row to the revision-history table and one short block below it, using the same four fields: **Development** (the instrument, with docket or citation), **Date**, **Effect on this report** (which conclusion it strengthens, weakens, or corrects, stated plainly — including “no change”), and a **confidence tag**. The body edit should be made at the same time and marked in the revision-history row, so that this page always describes changes that have actually been integrated. Where an update contradicts a prior assessment, the original judgment should be left visible and dated rather than quietly overwritten — the value of a report in a fast-moving docket is partly the record of what looked true when.

Data in this version is current through July 21, 2026. The July 20 resource-adequacy reports and the July 24 ERCOT submissions have now come due; the August 3 abeyance requests and the August 17 show-cause responses fall within the next month, and either may change the picture materially.

Catalog of Orders, Rules, and Directives Cited

Every formal instrument named in this report, grouped by issuer, with a one-line explanation and the date it took effect or was issued. The entry notes any item that remains a proposal or stays in development. This serves as a reference index rather than a legal citation — consult the underlying document (see Sources) for operative text. FERC — numbered rules (industry-wide) Instrument Each issuer group below opens with a retrieval note giving the authoritative repository and, where the instrument is docketed, the identifier needed to find it. Links point to the searchable repository rather than to individual documents, because document-level URLs at several of these bodies are session-generated and do not persist.

Order No. 888 / 890 / 1000 1996 / 2007 / 2011

The open-access transmission lineage: non-discriminatory transmission service (888), coordinated transparent planning (890), and regional/interregional planning with tariffed cost allocation (1000). Cited as the framework the 2026 large-load orders build on.

Order No. 841 Feb 2018

Opened RTO/ISO wholesale markets to electric-storage participation.

Order No. 845 Apr 2018

Reformed the generator interconnection process (informational studies, provisional agreements).

Order No. 2222 Sept 2020

Enabled distributed energy resource (DER) aggregations to participate in wholesale markets.

Order No. 2023 / 2023-A Jul 28, 2023 / Mar 21, 2024

Overhauled the generator interconnection queue: mandatory cluster studies, readiness deposits, first-ready-first-served. The explicit template for large-load queue reform.

Order No. 1920 / 1920-A/B May 13, 2024

Requires long-term (20-year) regional transmission planning and sets cost-allocation methods; effective Aug 12, 2024, first compliance filings June 2025.

Order No. 1977 May 13, 2024

Updates how FERC exercises its limited backstop transmission-siting authority.

FERC — case-specific orders and dockets (large loads)

Retrieval. Every item below is filed under a docket number. Enter it in FERC's eLibrary docket search at elibrary.ferc.gov/eLibrary/search to obtain the order, the filings that preceded it, and any rehearing or compliance filings that followed. Final rules and news releases are also indexed at ferc.gov.

Docket RM26-4-000 (ANOPR) Oct 2025

Advance Notice of Proposed Rulemaking, opened in response to the DOE §403 directive, on interconnecting large loads to the transmission system. Expected to mature into a NOPR and final rule across 2027 (proposed).

§206 order on PJM colocated load Dec 18, 2025

Found parts of PJM's tariff unjust/unreasonable; directed Firm and Non-Firm Contract Demand service and CIR adjustments. Compliance order Apr 16, 2026; rehearing Jun 18, 2026.

Susquehanna amended ISA rejected (ER24-2172) Nov 1, 2024

Rejected 2–1 the amendment raising co-located load at Susquehanna from 300 MW to 480 MW, holding that PJM had not shown the non-conforming provisions necessary. Rejected without prejudice; the substantive co-location questions were left open. The proceeding that produced the EL25-49 show cause of Feb 20, 2025.

Crane CIR transfer waiver Jun 1, 2026

Granted Constellation a one-time limited waiver moving 760 MW of Capacity Interconnection Rights from Eddystone Units 3 and 4 to the Crane Clean Energy Center, over the market monitor's protest, allowing full deliverability ahead of transmission upgrades dated to December 2030.

Six §206 show-cause orders (EL26-67 et seq.) Jun 18, 2026

Ordered every RTO/ISO to justify or reform its large-load and co-location tariff provisions across five reform categories; responses due Aug 17, 2026.

Tri-State HIL tariff order (ER25-3316) **Oct 2025**

Rejected a co-op's High Impact Load tariff 3–0 as an intrusion on state retail-rate jurisdiction — a central jurisdictional precedent.

SPP HILL approval **Jan 2026**

Approved SPP's High Impact Large Load process, which jointly studies a large load and its proximate generation.

DOE and multi-state directives

Retrieval. Section 202(c) emergency orders and §403 directives are published at energy.gov; where a directive asks FERC to act, the resulting proceeding carries its own docket in eLibrary.

DOE §403 directive (letter to FERC) **Oct 2025**

Invoked DOE Organization Act authority to direct FERC to consider national rules for interconnecting large loads (>20 MW) to the transmission system; triggered RM26-4.

DOE §202(c) emergency **Jun 30, 2026**

Directed backup-generation resources (including at data centers) to be order 13-state governors' cost principles available during a PJM heat emergency; effective through Jul 3, 2026. 2026

NERC — alerts, guidelines, and standards

Retrieval. Alerts, reliability guidelines, standards under development and the Long-Term Reliability Assessment are published at nerc.com. Standards projects are indexed by project number — Project 2026-02 for the large-load work described here.

Long-Term Reliability Assessment (LTRA) **Jan 2026**

Raised the ten-year summer-peak forecast by 224 GW (~69%), attributing most to data centers; flags PJM high-risk from ~2029, MISO from ~2028.

Level 3 Essential Action Alert **May 4, 2026**

NERC's highest-urgency alert; seven essential actions on large-load ride-through, modeling, and coordination. Responses due Aug 3, 2026 (non-binding).

Reliability Guideline: Emerging Large Loads **May 2026**

Voluntary best-practice template: treat clustered load loss as a reserve-sizing contingency; require communication and operating instructions; model behindthe-meter resources.

Gaps white paper **Mar 2026**

Found existing standards and practices inadequate for large-load integration; recommended registration and standards changes.

Project 2026-02 (Computational Loads) **SAR Apr 1, 2026**

Standards-development project; a “bridge” Reliability Standard targeted for year-end 2026, broader standards from 2027, plus a computational-load registry.

2026 State of Reliability report **Jul 2026**

Documented the 2025 customer-initiated load-loss events (~1,800 MW Feb, ~1,300 MW Jun); source of the “five-alarm fire” framing.

Existing standards cited **in force**

PRC-024-4 and PRC-029-1 (generator/IBR ride-through, the model for a load standard); BAL-002 (reserve/MSSC sizing); TOP-001, IRO-001, PER-005 (operating instructions, personnel); EOP-011-4 (emergency operations).

ERCOT and PUCT (Texas)

Retrieval. Protocol revisions (NPRR, NOGRR, PGRR) with their comments, impact analyses and board approvals are at ercot.com under market rules; working-group presentations cited in Sections 4 and 6 are filed under the relevant committee. PUCT rulemakings and project filings are at puc.texas.gov.

PGRR145 / NPRR1325 (Batch Zero) **Effective Jul 11, 2026**

Replaces serial large-load studies with a system-wide batch study (≥75 MW); sorts loads into base/studied/not-included with a transmission-limited allocation cap. Also carries WLPUN and PCLR co-location designations and ramp-rate/registration rules.

NOGRR282 / NPRR1308 **2026**

ERCOT large-load ride-through obligations and dynamic-modeling requirements — ahead of the NERC standard.

PUCT Project 58480 (16 TAC §25.370) **Effective Mar 1, 2026**

After 2026, admits only large loads with a qualifying executed interconnection agreement into the ERCOT load forecast.

PUCT Project 58481 (16 TAC §25.194) **Proposed; due 2026**

Implements SB6: 75 MW threshold, site-control and disclosure, ~$50k/MW financial security, and duplicate-request disclosure (due Dec 2026).

SB6 / PURA §37.0561 **2025 (Texas law)**

The statute directing the PUCT large-load rules above.

NPRR1180 / PGRR107 **2025**

Revised ERCOT forecast criteria to admit TSP-forecast load not yet backed by an agreement, under a “quantifiable evidence” standard.

2024 Regional Transmission Plan; Permian Basin Plan **Dec 2024; PUCTapproved Oct 2024**

First ERCOT RTP with dual 345 kV and 765 kV builds; the 765 kV backbone is a late-decade/2030s project.

PJM, MISO, and SPP

Retrieval. Tariff language, auction results and stakeholder materials are published at pjm.com, misoenergy.org and spp.org. Because each of these instruments takes effect only on FERC acceptance, the authoritative version is the accepted filing in eLibrary rather than the operator’s own posting.

PJM Load Adjustment Request framework **Jul 2025**

Vetting process that filters speculative/duplicative large-load requests from the forecast; drove the Jan 2026 near-term downward revision.

PJM Board CIFP-LLA Decisional Letter **Jan 16, 2026**

Board direction on the Critical Issue Fast Path for large-load additions, including BYOG-style expedited tracks.

PJM reliability backstop procurement **Approved Jun 30, 2026**

One-time capacity procurement (average cost capped at $555/MW-day) billed to large loads to cover the projected shortfall.

Firm / Non-Firm Contract Demand service **From Dec 2025 order**

New PJM transmission-service products letting co-located load cap or interrupt its grid withdrawals by contract.

MISO ERAS / MTEP EPR **2025–2026**

Expedited Resource Addition Study (~90-day GIA for reliability-critical resources) and the Expedited Project Review for large-load-driven transmission upgrades.

State large-load tariffs (retail, state-jurisdiction)

Retrieval. These are state proceedings and are not in eLibrary. Each is retrieved from the docket system of the commission concerned — for the Indiana contract discussed in Section 6, Cause No. 46276 at iurc.portal.in.gov. Commercial terms in special contracts are frequently redacted, which is itself a finding of Section 6.

AEP Ohio data-center tariff (PUCO) **2025**

≥25 MW class; 85% minimum take; 12-year term; exit fees. ~$10M first-year cost on a 100 MW site; cut requests roughly in half.

Dominion GS-5 (VA SCC PUR-2025-00058) **Nov 2025**

Separate Virginia data-center rate class; among the strictest terms.

Georgia Power PLL-18 **in effect**

More accommodating framework; ~11 GW of large load under contract as of Q1 2026.

Pennsylvania PUC model tariff **2026**

Adopts a “but-for” cost-causation standard: the large load pays all distribution and transmission needed to interconnect.

Appendix A — Scope, Methodology, and Limitations

This report is a synthesis of the public record on large-load grid integration as it stood in early July 2026. It draws on primary regulatory instruments (FERC orders and dockets, NERC alerts and guidelines, ERCOT protocol revisions, PUCT rulemakings, and state commission decisions) and on secondary analysis from law firms, market monitors, research institutions, and trade press. Where a figure originates in secondary reporting rather than the underlying filing — several of the capacity-cost and demand-forecast numbers do — it is presented as reported, and a reader relying on it for a transaction should confirm it against the source document. Five limitations are worth stating plainly. First, this is a fast-moving area: at least six of the proceedings discussed have filing deadlines inside the next sixty days, and the RTO show-cause responses due August 17, 2026 will materially change the picture. Second, the ERCOT queue and forecast figures are the subject of active dispute even within ERCOT, and the report treats them as indicative of magnitude rather than as settled quantities. Third, several figures illustrate schematically rather than measure — Figures 4 (event reconstruction), D1 and D2 (illustrative national demand and composition), and T2, T4, and T5 (illustrative traces and envelopes) are labeled as such and should not be read as data. Fourth, the Assessment boxes are the author's analytical judgment, not neutral summaries of the record; they are labeled and confidence-tagged, and the concluding note preserves the open questions rather than resolving them. Fifth, and most important for anyone quoting a number from this report: load-growth and queue figures in this field frequently mix metrics that are not comparable — five-year versus ten-year horizons, FERC Form 714 filings versus augmented “headline” totals, raw queue volume versus probability-weighted forecasts, and different vintages of the same series. This report attempts to use consistent series and to flag the distinction at the point of use (see Figure D3 and Section 1, where the widely-quoted “128 GW” 2024 figure is separated from the 67 GW FERC-714 value it is often conflated with). A reader carrying any single number out of context should first confirm which metric, horizon, and vintage it represents; a prior version of this report itself conflated the two Grid Strategies series, which is precisely the error this note exists to prevent. The report does not cover several adjacent topics that a fuller treatment would include: water and environmental permitting, community and land-use opposition, the security (NERC CIP) posture of gigawatt-scale loads, and interregional transmission expansion under Order No. 1920. These are noted here so their absence is understood as a scoping choice rather than an oversight. Readers who want the engineering layer beneath the main argument — electrical architecture, load behavior, planning studies, modeling, and forecasting assumptions — will find it in the Technical Annex.

Three registers run through this report, and the distinction matters for how much weight each statement carries. **Cited fact** rests on a primary instrument, or on reporting of one, and is sourced in the reference block for the section where it appears. **Interpretation** sits in the Assessment boxes, each carrying an explicit label and a confidence tag. Between them lies **synthesis**: statements in the body that connect several cited facts into a finding — for example, that queue integrity precedes cost allocation rather than running parallel to it. Synthesis of that kind represents the author's reading rather than a sourced claim, and the text marks it with conditional language (“appears”, “suggests”, “on current evidence”) wherever it moves beyond what the citation establishes. A reader checking any single statement should be able to place it in one of the three.

Appendix B — Glossary of Terms

Acronyms and specialized terms used in this report, in the sense intended here.

Ancillary services
Grid support services (reserves, frequency and voltage support) that keep supply and demand balanced instant to instant; large loads may be charged for them even when co-located.

ANOPR / NOPR
Advance Notice / Notice of Proposed Rulemaking — the early and formal stages of a FERC rule, preceding a final rule.

ARM
Anticipated Reserve Margin — the cushion of expected supply over peak demand; a falling ARM signals tightening adequacy (Section 2, p. 27).

BAL-002 / MSSC
NERC balancing standard and the “Most Severe Single Contingency” it sizes reserves against — the benchmark now being applied to clustered large-load loss.

BESS
Battery Energy Storage System.

BTM
Behind-the-Meter — generation or storage on the customer side of the revenue meter (Sections 5 and 8, pp. 41, 61).

BYOG
Bring Your Own Generation — an expedited pathway conditioned on the load supplying incremental generation.

Capacity accreditation
The step that converts a resource’s nameplate rating into the firm capacity credited for adequacy, usually through ELCC; a load able to curtail may be accredited as a resource rather than counted only as demand (Technical Annex T.6, p. 93).

Capacity market
A forward market (notably PJM's) that pays generators years ahead to be available at peak; its clearing price is a channel by which load growth reaches retail bills (Section 3, p. 32).

CIR
Capacity Interconnection Rights — a generator's accredited right to sell capacity; reduced when the unit serves on-site load (Section 5, p. 41).

CLR
Controllable Load Resource — an ERCOT large load that is dispatchable in the market and can be curtailed (Section 6, p. 46).

CMPLDW / PERC1
The WECC composite load model (motors + electronics + DER) and PERC1, the new computational-load model that captures voltage-sensitive trip, UPS transfer, and reconnection (Technical Annex T.4, p. 87).

COD
Commercial Operation Date — when a project begins commercial service.

Coincident / non-coincident peak
A load’s demand at the instant of system peak versus its own maximum whenever that falls; the gap explains why adding a megawatt of load need not raise system peak by a megawatt (Technical Annex T.5, p. 91).

Co-located load
Load sited directly at a generator, drawing power without full reliance on the transmission grid (Section 5, p. 41).

Curtailment-enabled headroom
The additional load an existing grid can serve if that load agrees to reduce for a limited number of hours per year — on the order of 20 to 90, depending on the curtailment level assumed (Section 6, p. 46).

Diversity factor

The ratio of the summed individual peak demands to the combined peak of the same set; it falls as loads coincide, which is why high-load-factor computational campuses erode the diversity utilities have long relied on (Technical Annex T.5, p. 91).

DR

Demand Response — paying or instructing customers to cut usage at peak; an alternative source of headroom to on-site generation.

DRRS

Dispatchable Reliability Reserve Service — a new ERCOT ancillary service created under Texas law.

ELCC

Effective Load-Carrying Capability — a probabilistic method for crediting how much a resource (or flexible load) actually contributes to resource adequacy (Technical Annex T.6, p. 93).

EMT

Electromagnetic Transient — a high-fidelity, full-waveform simulation class needed to model power-electronic load behavior at the point of interconnection, especially at weak-grid points.

FIDVR

Fault-Induced Delayed Voltage Recovery — voltage staying depressed after a fault clears; a dynamic-assessment failure mode (Technical Annex T.8, p. 97).

LOLE / LOLH / EUE

Loss-of-Load Expectation (days/yr short; standard ≤ 0.1, the “1-in-10” criterion), Loss-of-Load Hours, and Expected Unserved Energy — the probabilistic resource-adequacy metrics (Technical Annex T.6, p. 93).

ProbA

Probabilistic (resource-adequacy) Assessment — Monte Carlo simulation over weather, load, and outage combinations used to compute LOLE/LOLH/EUE.

PRM / RML

Planning Reserve Margin and the Reference Margin Level it must meet to satisfy the LOLE standard (Technical Annex T.6, p. 93).

RoCoF

Rate of Change of Frequency — how fast frequency moves after a sudden loss; worsens as inverter-interfaced load displaces inertia (Technical Annex T.8, p. 97).

N-1 / N-1-1

Contingency criteria: the system must withstand the loss of any single element (N-1), or a single element followed by another (N-1-1), without cascading; a stricter N-2, the simultaneous loss of two elements, is applied selectively to the most critical facilities — the core of TPL-001 planning (Technical Annex T.3, p. 85).

SCR / WSCR / CSCR

Short-Circuit Ratio and its weighted and composite variants — measures of local grid strength; low values flag weak-grid points where inverter-dominated loads may be unstable and EMT study is required.

TPL-001

NERC's transmission-system planning standard (currently TPL-001-5.1), defining the contingencies the system must withstand.

ZIP / static load model

A load represented as fixed or voltage-proportional constant power — the legacy model that cannot capture a load tripping during a fault (Technical Annex T.4, p. 87).

ERAS / EPR

MISO's Expedited Resource Addition Study (fast-track generator interconnection) and Expedited Project Review (fast-track transmission upgrades).

Firm service

Transmission or interconnection service the grid commits to provide at all times — the opposite of non-firm/curtailable service.

FPA §201(b) / §205 / §206

Federal Power Act sections defining FERC jurisdiction (201(b)), utility-initiated rate filings (205), and Commission-initiated “just and reasonable” review (206).

Grid-following (GFL) inverter

A converter that measures grid voltage and injects current in step with it, relying on an already-stiff grid to hold its reference; the dominant design in existing inverter-based resources and in load-side power electronics (Technical Annex T.4, p. 87).

Grid-forming (GFM) inverter

A converter that regulates its own voltage and phase angle, presenting to the network as a voltage source and lending system strength; increasingly required at weak points of interconnection (Technical Annex T.4, p. 87).

HILL

High Impact Large Load — SPP's process (and the name of the rejected Tri-State tariff) for jointly studying a large load and its proximate generation.

Hosting capacity

The load or generation a given feeder or substation can accept before a thermal, voltage, or protection limit forces an upgrade (Technical Annex T.9, p. 99).

Interconnection queue

The ordered list of pending requests to connect to the grid; large-load queues have swelled with speculative and duplicate entries (Section 1, p. 23).

IRP

Integrated Resource Plan — a state-regulated utility's long-range supply plan.

Islanding

Disconnecting from the grid and running on on-site generation; when uncommanded and at GW scale it is a bulk-system disturbance (Section 8, p. 61).

ISO / RTO

Independent System Operator / Regional Transmission Organization — operators of organized wholesale markets (PJM, MISO, SPP, CAISO, ISO-NE, NYISO; ERCOT is an ISO outside FERC jurisdiction).

LAR

Load Adjustment Request — PJM's framework for vetting large-load additions before they enter the forecast (Section 1, p. 23).

LSE

Load-Serving Entity — the entity responsible for supplying an end-use customer.

LTRA / LTLF

NERC's Long-Term Reliability Assessment and PJM's Long-Term Load Forecast — the principal ten-plus-year outlooks cited here.

NITS

Network Integration Transmission Service — the default full-service (firm) transmission product.

Non-firm service

Interruptible transmission or interconnection service, generally faster and cheaper than firm service (Sections 6 and 8, pp. 46, 61).

PGRR / NPRR / NOGRR

ERCOT Planning Guide / Nodal Protocol / Nodal Operating Guide Revision Requests — the protocol-change instruments (e.g., PGRR145, NPRR1325).

Protection coordination

The timing sequence by which relays and breakers isolate a fault at the smallest affected zone; large inverter loads complicate it through fast undervoltage trips, anti-islanding settings, and altered fault current (Technical Annex T.3, T.8).

POI

Point of Interconnection — where a load or generator connects to the grid.

PCM

Production Cost Model — a chronological unit-commitment and economic-dispatch simulation of the whole market, used to study congestion, cost, and dispatch (Technical Annex T.10, p. 101).

PUE

Power Usage Effectiveness — total facility energy divided by IT energy; a data-center efficiency metric (lower is better; best-in-class ≈ 1.3) used as a gate in some jurisdictions (International Experience).

PRC-024 / PRC-029

NERC standards requiring generators (024) and inverter-based resources (029) to ride through voltage and frequency excursions — the model for a future load standard.

PUCT / PURA / SB6

Public Utility Commission of Texas; Public Utility Regulatory Act; Senate Bill 6, the 2025 Texas large-load law.

Ride-through

The ability of a device to remain connected and operating through a grid disturbance rather than tripping offline (Section 4, p. 35).

RMS / positive-sequence simulation

Phasor-domain modelling at fundamental frequency, assuming a balanced system and representing the network by its impedance; the standard transient-stability tool, and an approximation that breaks down at inverter-dominated interfaces where EMT becomes necessary (Technical Annex T.4, p. 87).

STATCOM / SVC

Static synchronous compensator / static VAR compensator — fast voltage-support devices at the POI.

UPS / DVFS

Uninterruptible Power Supply; Dynamic Voltage and Frequency Scaling — facility-side systems relevant to ride-through and flexibility.

WLPUN / PCLR

ERCOT's Wholesale Load Point Un-Netted designation and Pre-Commercial Load Resource election, formalizing co-located and pre-energization arrangements.

Key Numbers

The figures most frequently cited in this report, with the date each was reported and the body that published it. Values drawn directly from official records are given as published; modelled or derived values are shown as ranges. Where two authorities measure the same quantity on different bases, both are listed rather than reconciled.

Metric	Value	As of	Source
ERCOT large-load interconnection queue	438 GW	Jun 2026	ERCOT
ERCOT large load approved to energize	~9 GW	Apr 2026	ERCOT LLWG
ERCOT observed simultaneous large-load peak	~3.7 GW	Mar 2026	ERCOT LLWG
ERCOT all-time system peak	85.5 GW	Aug 2023	ERCOT
PJM Base Residual Auction clearing price, 2028/29	$325/MW-day (at cap)	Jul 2026	PJM
PJM capacity cost attributable to data centres	40–45% of $47.2B over three auctions	Mar 2026	Monitoring Analytics (IMM)
Measured capacity-driven bill increase	~$10–20/month in affected PJM zones	2025–26	IEEFA; DC OPC
Largest uncommanded load loss, ERCOT	1,560–1,600 MW; 60.235 Hz; 12 min 30 s to recover	Dec 2022	ERCOT PDCWG
Largest uncommanded load loss, Eastern Interconnection	~1,500 MW across 60 delivery points	Jul 2024	NERC
ERCOT instantaneous-loss study threshold	~2,600 MW	Sep 2025	ERCOT LLWG
Documented ERCOT events above 100 MW	26	Jan 2023 – Sep 2025	ERCOT
US retail electricity sales	3,975 TWh	2024	EIA EPA Table 2.5
US installed capacity, net summer	1,230 GW at nameplate; 955 GW excluding wind and solar	2024	EIA EPA Tables 4.2.A/B
US data-centre consumption	176 TWh (LBNL); 183 TWh (IEA)	2023; 2024	LBNL; IEA
US data-centre consumption, 2030 projection	649 TWh reference, 578–664 range (LBNL); 426 TWh (IEA)	2030	LBNL 2025 Update; IEA
Modelled curtailment-enabled headroom	76–98 GW at 0.25–0.5% of annual consumption	2025	Nicholas Institute, Duke
Demand response cleared in PJM capacity market	5,531 → 7,299 MW after ELCC reform	2026/27 → 2027/28	PJM
FERC show-cause orders to RTOs and ISOs	Six; responses due Aug 17, 2026	Jun 18, 2026	FERC (EL26-67 et seq.)
NERC standards and registration criteria due	Dec 31, 2026	Directed Jul 16, 2026	FERC (RD26-7-000)

Table N1 — Key numbers. Each entry is discussed, sourced and qualified in the section indicated in the References. This page is a locator, not a substitute for the treatment in the text: several of these figures carry material caveats — notably the ERCOT queue, which includes duplicate and speculative requests, and the 2030 consumption projections, on which two careful bodies differ by roughly fifty percent.

Sources: As listed in the final column of each row; each figure is discussed and qualified in the section indicated.

References

Update sources (v1.1, July 13–17, 2026)

- FERC, *Reliability Standard(s) Pertaining to Computational Load Integration*, Docket No. RD26-7-000, 196 FERC ¶ 61,031 (July 16, 2026); FERC July 2026 open-meeting agenda, Item E-1. Coverage: Power Magazine; National Law Review; Daily Energy Insider.
- PJM Interconnection, Base Residual Auction reports for the 2025/2026 (July 30, 2024), 2026/2027 (July 22, 2025), 2027/2028 (December 17, 2025) and 2028/2029 (July 14, 2026) delivery years — the RTO clearing-price series $28.92, $269.92, $329.17, $333.44 and $325/MW-day, and the compressed auction calendar that placed two auctions in 2025 as PJM returned to a three-year-forward cycle. PJM Interconnection, 2028/2029 Base Residual Auction Report and news release, July 14, 2026 (138,318 MW UCAP; $325/MW-day; ~6.8 GW shortfall; ~525 MW new generation). Coverage: Utility Dive; PJM Inside Lines.
- Office of Governor Kathy Hochul, executive order establishing the first statewide hyperscale data-center moratorium, July 14, 2026; NY DPS Energize NY proceeding and GEIS. Coverage: Reuters; The Washington Post; CNBC.
- Ratepayer Protection Pledge, Presidential Proclamation 11014 (March 4, 2026); mid-July 2026 White House convening reported by Reuters and the Associated Press.
- *Trump v. Slaughter*, No. 25-332 (U.S. June 29, 2026), overruling *Humphrey's Executor*; *Trump v. Cook* (Federal Reserve carve-out). Analysis: Akin; Troutman Pepper Washington Energy Report; Congressional Research Service LSB11448.

Update sources (v1.2, July 20–21, 2026)

- FERC, six §206 show-cause orders, June 18, 2026 (EL26-67 PJM; EL26-68 SPP; EL26-69 NYISO; EL26-70 MISO; EL26-71 CAISO; EL26-72 ISO-NE), 195 FERC ¶ 61,211–61,216, for the procedural calendar: interventions July 9; generation-adequacy informational reports July 20; abeyance requests August 3; responses and briefs August 17; stakeholder comments on or about September 16. FERC fact sheet and staff presentation, June 18, 2026. Coverage: Day Pitney; McGuireWoods; Baker Botts; Willkie; RMI.
- ISO New England, compliance notice on the FERC large-load show-cause order, June 29, 2026, stating that the ISO and the New England transmission owners intend to request a 90-day abeyance by the August 3 deadline. ISO Newswire.
- Ratepayer Protection Pledge, The White House, March 4, 2026 (seven signatories; commitments to fund new generation, pay for all new power delivery infrastructure upgrades, negotiate separate rate structures, and pay for contracted capacity whether used or not; no enforcement or audit provisions). Google, “Supporting the White House Ratepayer Protection Pledge” (March 4, 2026) and the Capacity Commitment Framework as applied in the Ameren Missouri / New Florence announcement (May 20, 2026). On the limits: Ari Peskoe, Harvard Electricity Law Initiative, on the 1994 transmission pricing policy and embedded-cost recovery of data-center-driven upgrades (April 2026); Brookings, on the absence of enforcement (July 2026); reporting that colocation operators were not signatories and that a wider pledge including utilities and developers is being convened.
- California ISO, *Informational Report*, Docket No. EL26-71-000, July 20, 2026 — integrated CEC/CPUC/CAISO planning framework and the 2022 memorandum of understanding; ~36 GW of generation and storage added January 2020–March 2026 (>16 GW storage); 6.8 GW added in 2025; CPUC procurement obligation of 24.8 GW (2021–2032) with >20 GW under contract to 2029; 2026 summer assessment surplus above 2.5 GW against a 1-in-10 LOLE standard; CEC data-center forecast of 1.8 GW by 2030 and 4.9 GW by 2040; 2025 WEIM peak of 130.1 GW, down 3.8% year on year, per the CAISO Department of Market Monitoring annual report; expedited Large Loads stakeholder calendar ending in a November 16, 2026 filing that assumes an abeyance. caiso.com
- PJM Interconnection, news release on the 2028/2029 Base Residual Auction, July 14, 2026 (138,318 MW UCAP; 10,864 MW FRR; 149,182 MW total; 6,831 MW short; 14.7% reserve margin; special backstop procurement to be sought for September); PJM CIFP Reliability Backstop Procurement proposal paper, June 30, 2026 (procurement window September 10 – October 9, 2026; cost allocated by forecast load additions). pjm.com

- FERC, *2025 State of the Markets* (staff report, March 19, 2026) — source for the data-center growth figures used throughout: more than 50 GW of data-center capacity operating in the United States at the end of 2025, 24% compound annual growth nationally since 2020, and 43% in MISO, the fastest of any region. Every occurrence of the 43% figure in this report draws on this dataset. MISO large-load additions programme materials (three-part framework; 120-day approval target; ERAS, EPR and zero-injection interconnection agreements; new study process in development), misoenergy.org. MISO's July 20 filing in EL26-70 is reported by RTO Insider, “MISO to Design Dedicated Queue Process for Large Loads and Devoted Generation” (July 20, 2026) — new expedited study process outside the queue and a non-firm transmission service option; the filing itself had not been retrieved when this revision closed.

Organized by the section each source principally supports, following a consolidated list of the primary regulatory documents that anchor the report. This report uses a grouped-citation model rather than inline footnotes: to trace any quote, fact, or figure, find the section it appears in and consult that section's reference block below, plus the consolidated primary-documents list for any order, rule, or standard. Primary instruments are cited by name, docket, or number so a reader can retrieve them from the issuing agency's docket; secondary analyses and reporting include the retrievable link where one was used. Every headline figure in the report — the 438 GW queue, the capacity-cost shares, the 1,500 MW event, the load-growth series, the adequacy metrics — is sourced here to a primary filing or the reporting of one. This chapter is the complete provenance of the report. A note on citation form: dates given are issuance or effective dates as reported at the time of writing (mid-July 2026). Where a figure comes from secondary reporting of a primary filing, or where a number is disputed or metric-dependent (see the 128-vs-67 GW discussion at Figure D3), that is flagged at the point of use. Several proceedings were live as this report was prepared; readers should check the docket for subsequent orders, rehearings, and compliance filings. Primary regulatory documents (consolidated)

- FERC — Order No. 841 (Feb 2018); Order No. 845 (Apr 2018); Order No. 2222 (Sept 2020); Order No. 2023 (Jul 28, 2023) and 2023-A (Mar 21, 2024); Order No. 1920 / 1920-A / 1920-B (May 13, 2024) and Order No. 1977; §206 order on PJM co-located load (Dec 18, 2025), compliance order (Apr 16, 2026), rehearing (Jun 18, 2026); six §206 show-cause orders (Jun 18, 2026, Dockets EL26-67 et seq.; 195 FERC ¶ 61,211–61,216); order rejecting Tri-State's HIL tariff (Docket ER25-3316, Oct 2025); ANOPR (Docket RM26-4000, Oct 2025). FERC explainers and fact sheets at ferc.gov.
- NERC — Long-Term Reliability Assessment (Jan 2026); Level 3 Essential Action Alert (May 4, 2026); Reliability Guideline: Risk Mitigation for Emerging Large Loads (May 2026); gaps white paper (Mar 2026); Project 2026-02 Computational Loads (SAR Apr 1, 2026); Rules of Procedure computational-load registry proposal (comment through May 15, 2026); 2026 State of Reliability report (Jul 2026); Large Loads FAQ and Action Plan. nerc.com.
- ERCOT / PUCT — PGRR145 / NPRR1325 (Batch Zero, eff. Jul 11, 2026); NOGRR282 / NPRR1308; NPRR1180 / PGRR107; Market Notice M-B062326-01; 2024 Regional Transmission Plan (Dec 2024) and Permian Basin Reliability Plan; PUCT Project 58480 (16 TAC §25.370) and Project 58481 (16 TAC §25.194); SB6 / PURA §37.0561. ercot.com; interchange.puc.texas.gov.
- DOE — §403 directive to FERC (Oct 2025); §202(c) emergency backup-generation order for PJM (Jun 30, 2026).
- FERC, six §206 show-cause orders to the RTOs/ISOs, June 18, 2026 — PJM (EL26-67), SPP (EL26-68), NYISO (EL26-69), MISO (EL26-70), with CAISO and ISO-NE issued the same day; reported at 195 FERC ¶ 61,211–61,216. FERC news release, “FERC Launches Aggressive Targeted Action to Speed Large Load Integration.” Analyses: Akin; Vinson & Elkins; Willkie Farr & Gallagher; Day Pitney; McGuireWoods; Troutman Pepper.
- FERC, Docket No. RM26-4-000 — Interconnection of Large Loads to the Interstate Transmission System (DOE §403 ANOPR, October 2025; >3,500 pages of comments).
- Akin, “FERC Issues Landmark Show Cause Orders on Large Load Interconnection.” akingump.com/en/insights/blogs/speaking-energy
- Troutman Pepper, Washington Energy Report, analysis of the June 18 orders. troutmanenergyreport.com
- McGuireWoods, client alert on the six §206 show-cause orders, June 2026. mcguirewoods.com
- ERCOT PGRR145 / NPRR1325 (Batch Zero), effective July 11, 2026; Market Notice M-B062326-01; Preliminary Long-Term Load Forecast 2026–2032.

- PUCT Project 58480 (16 TAC §25.370) and Project 58481 (proposed 16 TAC §25.194), implementing SB6 / PURA §37.0561.
- EUCI, "Utilities adopt large-load tariffs …" (Enverus estimate that gating cut AEP Ohio requests roughly in half). euci.com Primer — demand growth and load categories (Figures D1–D2)
- Grid Strategies, National Load Growth Reports (2023–2025), via Utility Dive ("Five-year US load growth forecast surges … to 128 GW," Dec 6, 2024): consistent FERC Form 714 five-year peak-growth series 23 → 39 → 67 GW (2022–24); the 128 GW headline adds ~61 GW of further planning-document updates; ~166 GW projected by 2030 in the 2025 report. gridstrategiesllc.com
- EIA, "After more than a decade of little change, U.S. electricity consumption is rising again" (Today in Energy) and "We expect rapid electricity demand growth in Texas and the mid-Atlantic" (July 31, 2025): ~0.1%/yr growth 2005–2020; 0.8%/yr 2020–2024; 2.2% forecast for 2025 and 2026. EIA, Annual Energy Outlook 2026: 2.1%/yr over the last five years after 15 years of nearly flat consumption. eia.gov. ITIF, "The United States Needs Data Centers …" (Nov 2025): ~760 GW July 2024 U.S. peak. itif.org
- Lawrence Berkeley National Laboratory, 2024 United States Data Center Energy Usage Report: 325–580 TWh/yr by 2030 (6.7– 12% of U.S. electricity); Goldman Sachs and S&P; Global data-center demand outlooks (2025–26); BofA Global Research US grid transformation note.
- World Resources Institute, "Powering the US Data Center Boom" and "How to Manage Growing Electricity Demand in the US"; Bipartisan Policy Center, "Electricity Demand Growth and Data Centers: A Guide for the Perplexed" (Apr 2026); Belfer Center, "AI, Data Centers, and the U.S. Electric Grid" (Feb 2026). wri.org
- Rabobank/RaboResearch, "The US scrambles to meet surging power demand by 2030" (Jun 2026), citing EIA, ICF, ACP, Grid Strategies. rabobank.com

Section 1 — Load-forecasting methodology and forecast trend (Figures 10, D3)

- PJM, 2026 Long-Term Load Forecast, issued January 14, 2026; Load Adjustment Request Implementation document (July 2025); PJM Inside Lines, "PJM's Updated 20-Year Forecast …" insidelines.pjm.com
- Utility Dive, "PJM trims near-term load forecast on stricter data center vetting" (Jan 15, 2026); Power Magazine coverage of the 2026 forecast. utilitydive.com
- PSE&G; 2026 Load Forecast documentation (100% / 50% confidence tiers by request stage); NOVEC bottom-up forecast documentation, via PJM. pjm.com
- ERCOT, "Long-Term Load Forecast Update (2025–2031) and Methodology Changes" (Apr 2025); NPRR1180 / PGRR107; PUC Subst. Rule 25.101. ercot.com
- NERC, Long-Term Reliability Assessment (January 2026; +224 GW / ~69% ten-year summer-peak revision vs. 2024 LTRA).
- Grid Strategies National Load Growth Reports (2023–2025) and Utility Dive coverage for the FERC-714 five-year forecast-trend series (23/39/67 GW) in Figure D3. gridstrategiesllc.com
- EIA, *Electric Power Annual* — Table 2.5 (retail sales to ultimate customers, 2014–2024) for Table D1 and Figure D1; Tables 4.2.A and 4.2.B (existing net summer capacity by energy source, 2014–2024) for the capacity series in Figure D1; Table 4.6 (capacity additions and retirements, 2024). EIA, *Short-Term Energy Outlook* for the 2.2%/yr forecast and the 2025 capacity factors used in Table 2A. eia.gov

Section 2 — Resource adequacy and the development-timeline gap

- MISO, Expedited Resource Addition Study (ERAS) and MTEP Expedited Project Review; Large Load Additions Workshop, January 30, 2026; IPWG materials, 2026.
- FERC Order No. 845, *Reform of Generator Interconnection Procedures and Agreements*, 163 FERC ¶ 61,043, Docket No. RM17-8-000 (April 19, 2018), and Order No. 845-A on rehearing — surplus interconnection service, defined as the unneeded portion of interconnection service established in an LGIA, such that total interconnection service at the point of interconnection would remain the same if the surplus service were used. Available only to the

extent it requires no new network upgrades; the original interconnection customer and its affiliates hold priority; the surplus agreement terminates with the original. ferc.gov

- PJM Board CIFP-LLA Decisional Letter, January 16, 2026; PJM stakeholder approval of the reliability backstop procurement, June 30, 2026.
- Capacity accounting (Table 2A): PJM, ELCC Class Ratings for the 2026/2027 Base Residual Auction and the 2025 PJM Effective Load Carrying Capability and Reserve Requirement Study — assumed resource portfolio, effective nameplate against installed capacity by class. pjm.com
- Transmission build-out and usable capacity: ERCOT, Permian Basin Reliability Plan (PUCT-approved 2024–25) and the 765 kV Strategic Transmission Expansion Plan (ERCOT board, December 2025); ERCOT, *2024 Regional Transmission Plan — 345-kV Plan and Texas 765-kV Strategic Transmission Expansion* (Jan 2025); Texas Public Policy Foundation, *An Economic Assessment of the 765-kV Strategic Transmission Expansion Plan* (2026) for the ~$33B capital total and the near-term zonal-flow finding; Burns & McDonnell *BenchMark* on MISO Tranche 2.1 ($21.8B, in service 2032–34); Utility Dive and Oncor/LCRA filings on route and mileage. FERC Order No. 881 (ambient-adjusted ratings, compliance July 12, 2025), Order No. 1920 (consideration of grid-enhancing technologies in regional planning), and the FERC explainer on dynamic line ratings; NYPA and PJM/AEP demonstrations for the reported 25% class of capacity gain. ercot.com; ferc.gov

Section 3 — Cost allocation and cost-shifting

- Monitoring Analytics (PJM Independent Market Monitor), 2025 State of the Market Report for PJM (Mar 2026) and capacity-auction cost analyses: data-center load = $6.5B (40%) of the $16.4B December 2025 auction, ~$6.2B of it unbuilt. Two distinct totals circulate for the three auctions (2025/26–2027/28, $47.2B cleared): $21.3B (45%) for forecast data-center load above existing data-center load, and $23.1B (~49%) including existing data-center load. This report uses the former throughout. monitoringanalytics.com
- Utility Dive, "Data centers were 40% of PJM capacity costs in last auction" (Jan 7, 2026) and "'Clear warning signs' as PJM wholesale power costs jump 54%" (Mar 14, 2026). utilitydive.com
- IEEFA, "Projected data center growth spurs PJM capacity prices by factor of 10" (capacity-price series $28.92 → $329.17/MW-day). ieefa.org
- NRDC (Tom Rutigliano), "Building Data Centers Without Breaking PJM" (Sept 2025) — origin of the $163B-through-2033 and ~$70/month figures, on the assumption that the capacity price collar lapses after 2027/28; Synapse Energy Economics, ~$100B through 2033 assuming mitigation continues. nrdc.org
- FERC, Docket ER26-1556-000, order accepting PJM's extension of the $325/$175 per MW-day price collar to the 2028/2029 and 2029/2030 Delivery Years (April 28, 2026); PJM, 2028/2029 Base Residual Auction results, July 14, 2026 ($16.4B cleared). Measured zonal bill impacts: IEEFA (~$18/mo western Maryland, ~$16/mo Ohio) and the D.C. Office of the People's Counsel (~$10 of a $21/mo Pepco increase).
- Bloomberg, "Data Centers Added $6.5 Billion to Big US Grid's Power Cost" (Jan 5, 2026); Canary Media, "PJM's capacity costs hit record as grid falls short on supply" (Dec 18, 2025).
- Sierra Club, "Data Center State Policies, 2026" (stranded-asset and cross-subsidization framing); Union of Concerned Scientists on ~$4.4B of data-center-related transmission cost. sierraclub.org

Section 4 — Ride-through and dynamic performance

- NERC, Level 3 Essential Action Alert (May 4, 2026); Reliability Guideline: Risk Mitigation for Emerging Large Loads (May 2026); gaps assessment (March 2026); large-load registration criteria (comment through May 15, 2026); Project 2026-02 SAR; incident reviews of the July 10, 2024 and January 2026 events.
- ERCOT NOGRR282 / NPRR1308 (large-load ride-through and dynamic modeling); NERC Data Center Load Modeling Workshop, September 15–16, 2026.
- ERCOT, "Inverter-Based Resource and Large Load Ride-Through Events: Background and Mitigation," Reliability and Markets Committee, June 19, 2023. Primary source for two distinct events. The June 4, 2022 Odessa IBR disturbance cost 2,555 MW of generation and took frequency to 59.7 Hz. The December 7, 2022 west Texas

large-load event removed about 1,600 MW, took frequency to 60.235 Hz, and needed 12 min 30 sec to recover; the same deck records it as a three-phase fault caused by a breaker failure, affecting a mix of data centers, oil and gas, and other industrial load. The deck also records ERCOT's 2008 generator ride-through requirement and the five smaller west Texas events of October 2022. ercot.com

- NERC, *Incident Review: Voltage-Sensitive Crypto Load Reductions* (January 9, 2026) — 26 events of indirect load loss above 100 MW between January 2023 and September 2025, reductions of 17% to 95% of pre-disturbance consumption, sensitivity below roughly 0.7 per unit single-phase, and the four mechanisms behind the trips: neutral overcurrent protection, wye-wye rather than delta-wye transformation, constant-power overcurrent, and cooling-drive undervoltage settings as high as 0.96 per unit. NERC defines indirect load loss as loss not attributable to the outage of the connecting line. Also: ERCOT Large Flexible Load Task Force, March 4, 2025, for the delayed-clearing character of the December 2022 west Texas event and the substantial oil and gas share of it.
- Voltage-sensitivity mechanism: NERC, *Incident Review — Considering Voltage-Sensitive Crypto Load Reductions* and the July 2024 Eastern Interconnection incident review, for the switched-mode power supply front end, the ITIC (formerly CBEMA) tolerance curve and its 20-millisecond allowance at 0.7 per unit, the behaviour of static and diesel-rotary uninterruptible supplies, and the finding on neutral overcurrent protection and transformer winding configuration. IEC 62040-3 for the 0.7–0.8 per unit transfer threshold within one to three cycles. nerc.com
- Event record (Table 4A): ERCOT, *Large Load Loss/Reduction Events 2020–2024* (Performance, Disturbance and Compliance Working Group, November 19, 2024) — dated events, magnitudes, fault types and clearing times for October 12 and 31, 2022, December 7, 2022 and January 23, 2023. NERC, *Incident Review: Considering Simultaneous Voltage-Sensitive Load Reductions* (Jan 2025) for the July 10, 2024 event, and the two January 2026 incident reviews of crypto-mining load reductions. ERCOT Large Load Working Group, September 19, 2025 — the ~2,600 MW instantaneous load-loss threshold from the 2030/31 study case. ESIG, "Engaging with Large Loads" (Nov 2025) — ERCOT's reduction of System Operating Limits and its characterisation of event impacts to date as minimal. nerc.com; ercot.com; esig.energy

Section 5 — Co-location, BTM, and electrically proximate load

- FERC, §206 order on PJM co-located load, *PJM Interconnection, L.L.C.*, 193 FERC ¶ 61,217, Docket Nos. EL25-49-000 *et al.* (December 18, 2025); April 16, 2026 compliance order; order on rehearing (June 18, 2026). Show Cause Order, 190 FERC ¶ 61,115 (February 20, 2025), which consolidated the AD24-11-000 technical conference of November 1, 2024 and Constellation's complaint in EL25-20-000. Paper-hearing schedule: PJM initial brief February 16, 2026; responses March 18; replies April 17.
- ERCOT NPRR1325 (WLPUN / PCLR; Form X / Form W); SPP HILL process, approved by FERC January 2026.
- American Bar Association, Infrastructure & Regulated Industries Section, "The Jurisdictional Collision over Large Loads and Data Center Interconnection" (Spring 2026). americanbar.org
- FERC fact sheet and order, *PJM Interconnection, L.L.C.*, 193 FERC ¶ 61,217, Docket Nos. EL25-49-000 *et al.* (December 18, 2025), for two points relied on here. The order requires the Eligible Customer serving co-located load to choose among four transmission service options: NITS; a new interim non-firm service open only to customers seeking NITS, running until the network upgrades behind that request are complete; a new Firm Contract Demand service; and a new Non-Firm Contract Demand service. And the order expressly declines to address jurisdictional matters regarding the interconnection of retail loads served through a co-location arrangement (Section 7). A separate reliability informational report from PJM fell due January 19, 2026. ferc.gov
- FERC, *PJM Interconnection, L.L.C.*, 189 FERC ¶ 61,078, Docket Nos. ER24-2172-000 and ER24-2172-001 (November 1, 2024) — order rejecting the Susquehanna amended ISA, with the Phillips dissent and Christie concurrence attached; rehearing denied, ER24-2172-002 (April 2025). The order supplies the 300 MW / 480 MW / 960 MW figures, the 2023 ISA history, the November 2023 fallback event, and the Exelon/AEP cost-shift estimate of up to $140 million a year — which Susquehanna answered as forgone retail revenue under the PPL LP-5 tariff rather than a shift of incurred cost. ferc.gov

- FERC, 195 FERC ¶ 61,162 (June 1, 2026) — one-time limited waiver transferring 760 MW of Capacity Interconnection Rights from Eddystone Units 3 and 4 to the Crane Clean Energy Center, granted over the protest of the PJM Independent Market Monitor on all four waiver criteria. Constellation's waiver request of March 31, 2026 states the December 2030 upgrade date and the second-half-2027 restart target. Related: DOE section 202(c) orders keeping Eddystone online as an energy-only resource from May 2025.
- Talen Energy Corporation, Form 8-K and press release, "Talen Energy Expands Nuclear Energy Relationship with Amazon" (June 11, 2025) — 1,920 MW at full contract quantity through 2042, full volume no later than 2032, transition to front-of-the-meter service at the spring 2026 refuelling outage, with PPL Electric Utilities responsible for transmission and delivery. Constellation Energy, "Constellation to Launch Crane Clean Energy Center" (September 20, 2024) for the twenty-year Microsoft agreement and the absence of any co-located data center. sec.gov
- Behind-the-meter generation (Table 5A): unit size ranges are the standard class ratings for reciprocating gas engines, aeroderivative and frame turbines, fuel cells, storage and standby diesel as deployed at data-center scale, and are given as indicative engineering ranges rather than as a survey of installed projects. Proximity: FERC's June 18, 2026 orders introducing "electrically proximate" large load and the two-substation working example.

Section 6 — Flexible and curtailable load

- Nicholas Institute, Duke University, "Rethinking Load Growth" (Norris et al., Feb 2025) — curtailment-enabled headroom across the 22 largest balancing authorities, which serve 95% of US load: 76 GW at 0.25% of maximum uptime, ~100 GW at 0.5%, 126 GW at 1%, i.e. roughly 22, 44 and 88 hours a year, with an average curtailment event near two hours.
- Berkeley Lab data-center demand range; Goldman Sachs demand outlook (May 2026); ICF load-flexibility analysis (June 2026); GridLab/Telos NV Energy case study.
- Norris et al., *Rethinking Load Growth*, and the Nicholas Institute webinar deck of February 19, 2025, for the basis of the figures used here. The study defines the curtailment rate as annual curtailed MWh divided by the new load's maximum potential annual consumption. Headroom on that basis: 76 GW at 0.25%, 98 GW at 0.5%, 126 GW at 1%, 215 GW at 5%. Hours in which any curtailment occurs: 85, 177 and 366. Average event duration: 1.7, 2.1 and 2.5 hours. Retention: at least half the new load in 88% of curtailment hours, three-quarters in 60%, nine-tenths in 29%. The study adds a constant new load in all hours against each balancing authority's maximum observed seasonal peak, and its stated limitations list network constraints first among the factors that would reduce the estimate, with unused reserve capacity among those that would raise it. The authors describe the result as a first-order estimate and set out a research plan to recompute it with the transmission network represented. nicholasinstitute.duke.edu
- EPRI, DCFlex initiative — pilot finding that 10–40% load modulation is achievable without breaching service-level agreements, reported in ITIF, "The United States Needs Data Centers, and Data Centers Need Energy" (Nov 24, 2025); EPRI DCFlex *Regulatory, Policy and Technology Trends Digest* (Sept 8, 2025) for the filed structure of the I&M;–Google demand-response product (unlimited curtailments, 12-hour summer and 15-hour winter duration caps). epri.com; itif.org
- Indiana Michigan Power and Google, special contract, IURC Cause No. 46276 (filed July 30, 2025; approved 2026) — Clean Capacity Arrangement plus custom demand response, with credit amounts and contract capacity redacted; Citizens Action Coalition Exhibit 1 (Oct 23, 2025) objecting to the scope of confidential treatment and the absence of a public ratepayer-impact analysis. Google, "Signed 1 GW of data center demand response" (Mar 2026) and "How we're making data centers more flexible" for the I&M;, TVA, Entergy Arkansas, Minnesota Power and DTE agreements and the earlier Omaha Public Power District demonstration. iurc.portal.in.gov; blog.google
- Demand response in the capacity market: PJM, "PJM Auction Procures 134,479 MW of Generation Resources" (Dec 17, 2025) — demand response at about 5% of the cleared supply mix, and cleared DR rising from 5,531 MW to 7,299 MW between the 2026/27 and 2027/28 auctions following the ELCC accreditation reform approved by FERC in May 2025, with offered megawatts essentially unchanged. pjm.com

Section 7 — Jurisdiction and the non-RTO grid

- FPA §201(b); FERC order rejecting Tri-State Generation & Transmission's High Impact Load tariff, 3–0 (Docket ER25-3316, October 2025).
- Utility Dive, "FERC rejects Tri-State's large load tariff over retail jurisdiction issues"; and coverage of NARUC / NCSL comments in RM26-4-000. utilitydive.com
- Mintz, "FERC Decision Highlights Jurisdictional Limits on Large Load Tariffs." mintz.com
- Balch & Bingham, "FERC's Regional Path for Large-Load Integration." balch.com/insights
- Womble Bond Dickinson, "Federal–State Jurisdictional Issues Raised by Data Centers and Large Load Interconnections." womblebonddickinson.com
- Snell & Wilmer, "FERC Sets June Action on DOE's Large Load Interconnection Plan." swlaw.com
- RTO Insider, "FERC Rejects Tri-State's High Impact Load Tariff." rtoinsider.com
- PUCO approval of the AEP Ohio data-center tariff (≥25 MW; 85% minimum take; 12-year term), 2025.
- Virginia SCC, Case PUR-2025-00058 (Dominion GS-5 rate class); Georgia Power PLL-18; Pennsylvania PUC model tariff (2026).
- Latitude Media, "The unsettled landscape of large-load tariffs." latitudemedia.com
- Columbia Climate Law Blog, "What Local Governments Should Know about Large-Load Tariffs" (EEI count: 23 states approved, 7 pending as of May 2026). blogs.law.columbia.edu/climatechange
- Smart Electric Power Alliance tariff survey (65 tariffs across 34 states, 2025), via Latitude Media.
- Buckeye Institute, "Undermining Ohio's Competitive Edge"; Ohio Statehouse News Bureau coverage of the AEP Ohio tariff. buckeyeinstitute.org
- electroneconomics, "The Tariff Lottery" (Georgia Power 11 GW under contract; developer-side ramp-risk analysis). electroneconomics.substack.com

Section 8 — Off-grid and islandable load

- NERC, 2026 State of Reliability report (released July 2026); documented 2025 customer-initiated load-loss events (~1,800 MW February, ~1,300 MW June; nine ERCOT crypto events >100 MW); Jim Robb "five-alarm fire" remark.
- Wall Street Journal, Jennifer Hiller, coverage of the NERC State of Reliability report (July 3, 2026), as reported.
- Data Center Knowledge, "NERC Flags AI Data Center Grid Risks in Report" (registration category; PERC1 model). datacenterknowledge.com
- Ascend Analytics, "Can US Interconnection Queues Survive Data Center-Driven Load Growth?" (BTM gas as stopgap; LCOE vs. ERCOT forward price), May 5, 2026. ascendanalytics.com
- EUCI, on AEP subsidiaries' ~750 MW of stranded generating capacity acquired for data centers that did not materialize. euci.com
- NERC, Level 3 Essential Action Alert (May 4, 2026) and Reliability Guideline (May 2026) on behind-the-meter resources and coordinated islanding; Carbon Direct and Utility Dive analyses. carbon-direct.com

Section 9 — 2018–2035 timeline

- FERC, Order No. 841 (Feb 2018), Order No. 845 (Apr 2018), Order No. 2222 (Sept 2020), Order No. 2023 (July 28, 2023) and Order No. 2023-A (Mar 21, 2024), Order No. 1920 and Order No. 1977 (May 13, 2024); FERC explainers and fact sheets. ferc.gov
- ERCOT, Large Flexible Load Task Force (established 2022); 2024 Regional Transmission Plan (Dec 2024) with 345 kV and 765 kV STEP; Permian Basin Reliability Plan (PUCT-approved Oct 2024); Baker Botts, "Bitcoin and Large Loads in ERCOT" (2022 LFLTF origin). ercot.com

- NERC, Project 2026-02 Computational Loads (SAR posted Apr 1, 2026; “bridge” standard targeted year-end 2026; broader standards from 2027); Large Loads FAQ and Rules of Procedure computational-load registry proposal. nerc.com
- NERC, 2025 Long-Term Reliability Assessment (published January 2026): PJM high-risk from ~2029 and ~210 GW summer peak by 2035; MISO elevated/high-risk from 2027–2028 and ~18 GW data-center load by 2035. nerc.com
- DOE, Section 202(c) emergency order directing backup-generation availability in PJM (June 30, 2026). pjm.com
- White & Case, Crowell & Moring, K&L; Gates — order-by-order dating of FERC Orders 2023 and 1920. Section 9 — the four clocks: supply chain and political economy (Figure 15)
- GE Vernova, Q1 2026 results (Apr 22–23, 2026): combined gas backlog and slot-reservation agreements grew from 83 to 100 GW (slot reservations 43→56 GW), targeting ≥110 GW by year-end 2026; data-center electrification orders of $2.4B in Q1. SEC Form 8-K; Power Engineering; Industrial Info Resources. sec.gov
- Utility Dive, “GE Vernova expects to end 2025 with an 80-GW gas turbine backlog that stretches into 2029” (Dec 11, 2025); electroneconomics, “Slotting In the Future: Gas Turbine Reservations and the AI Power Land Grab” (Dec 15, 2025) — slots sold out through 2027, delivery conversations at 2029–30; Siemens lead times exceeding 40 months. utilitydive.com
- MultiState, “State Data Center Legislation in 2026” and “Federal AI Data Center Policy Meets Resistance” (Apr 2026): 300+ datacenter bills across 30+ states in 2026; ≥18 states on separate rate classes; 11 states on moratoriums. datacenterbans.com for the New York Responsible Data Center Development Act (passed Jun 4, 2026), New Jersey Data Center Fair Share Act (signed Jul 7, 2026), Virginia $0.011/kWh tax (eff. Jul 1, 2026). multistate.us
- Brookings, “The pledge to protect ratepayers from AI data center costs needs enforcement” (Jul 2026): ≥48 projects (~$156B) blocked or stalled by local opposition in 2025; Ratepayer Protection Pledge (Mar 4, 2026), seven hyperscaler signatories; White House expansion reported by Reuters (Jul 13, 2026). brookings.edu

International experience

- Ireland — CRU Large Energy User Connection Policy Decision Paper (Dec 2025) ending the 2021 de facto moratorium; 100% onsite generation, 80% renewable matching, siting rules for >10 MVA; EirGrid Large Demand Facility Fault-Ride-Through Grid Code recommendation (Apr 1, 2026). Data centers ~22% of 2024 metered demand, ~31% projected by 2034; Dublin ~1,150 MW. William Fry, Arthur Cox, DC Byte, KPMG Ireland; EirGrid Winter Outlook 2025/26. cru.ie
- Netherlands — TenneT congestion declarations (zero large-connection headroom in much of Noord-Holland to ~2036; 5+ GW queues); shift toward flexible/non-firm connections and mandatory congestion management; 9 GW off-peak capacity vs. 70+ GW of applications; battery “congestion mitigator” contracts. RAP, “Gridlock in the Netherlands” (2024); Taylor Wessing; EnergyStorage.news (Apr 2026); Mordor Intelligence. tennet.eu
- Singapore — 2019 moratorium; 2022 pilot allocation; IMDA Green Data Centre Roadmap (May 2024) releasing ≥300 MW gated on PUE < 1.3 and ≥50% green power. DatacenterDynamics; IMDA; Mordor Intelligence. imda.gov.sg

Technical Annex (distribution and planning paradigms, Figure T6)

- Distribution-system impacts (T.9): EPRI hosting-capacity analysis frameworks; NERC and IEEE guidance on large-load distribution interconnection, service-transformer loading, voltage regulation, harmonics, and reverse power flow; utility distribution interconnection screens.
- Planning paradigms (T.10): deterministic vs. probabilistic planning and production cost modeling (e.g., chronological unit commitment/economic-dispatch tools); NERC ProbA methodology; the deterministic TPL-001 contingency framework alongside probabilistic LOLE/EUE adequacy studies (cross-referenced to T.3, T.6, T.7).

Technical Annex (Figures T1–T5)

- Schneider Electric, “The new power dynamic: AI data centers begin at the UPS” (Jul 2026); OPAL-RT, “AI workload variability and its impact on data center power stability” (Mar 2026); Oracle Cloud, “GPU Power Smoothing for Large-Scale AI Training” (Mar 2026). blog.se.com

- Peer-reviewed / preprint: “Wide-Area Power System Oscillations from Large-Scale AI Workloads”; “Modal Analysis of Spatial Load Correlation in AI Data Center-Dominated Power Systems”; “EasyRider: Mitigating Power Transients in Datacenter-Scale Training Workloads” (Stanford); “Power for AI Data Centers” (Energies, MDPI, Jan 2026). GPU swing 30–150% in ms; >1,000 MW/s aggregated; 14.7 Hz forced-oscillation observation. arxiv.org
- NERC — Reliability Guideline: Developing Load Model Composition Data (WECC CMPLDW); Level 3 Essential Action Alert: Computational Load Modeling (May 2026, requiring PERC1 or equivalent, EMT at weak-grid POIs; TPL-001 non-consequential load-loss expectations); Data Center Load Modeling Technical Reference and PERC2 specification (in development). TPL-001-5.1 planning standard. nerc.com
- Keentel Engineering, NERC Large Loads Action Plan and SPP HILL process guides (study sequence: load-flow, short-circuit per TPL-001-5.1, RMS dynamic, EMT/SCRCCT screen); ERCOT “Officer Letter Loads” attestation practice. keentelengineering.com
- Resource adequacy (T.6): NERC 2025 Long-Term Reliability Assessment (Jan 2026) and 2026 Summer/2025–26 Winter Reliability Assessments — LOLE 0.1/yr standard, LOLH, EUE, PRM/ARM, ProbA methodology; ERCOT ELCC switch and DR rising to ~53 GW by 2030; MISO PY2026–27 LOLE Study (PRM 7.9% summer / 18.9% winter). NERC “Assessment of Gaps” white paper (Mar 2026). nerc.com
- Capacity accreditation of co-located load-plus-resource systems: “Risk-Based Capacity Accreditation of Resource-Colocated Large Loads in Capacity Markets” (2025); PJM Manual 20A, Resource Adequacy Analysis. arxiv.org
- Scenario analysis & reliability assessment (T.7–T.8): NERC May 2026 Reliability Guideline (probabilistic scenarios across weather/load/outage on a network-aware footprint); Carbon Direct, “Inside NERC's Level 3 Alert on data center loads” (May 2026); Grid Strategies and EPRI growth-scenario framing; TPL-001-5.1 contingency categories P0–P7. carbon-direct.com