INL/RPT-25-87430

Center for Securing Digital Energy Technology

MegaWatt Mayhem: Grid Operator Challenges with Large Loads

Existing Landscape and Grid Stability Implications

SEPTEMBER 2025

Luis Fernando Enriquez-Contreras,
S M Shafiul Alam,
Timothy R. McJunkin,
Megan Culler,
Jonathan Tacke, and
Emma Stewart

Idaho National Laboratory

Idaho National Laboratory

Battelle Energy Alliance manages INL for the U.S. Department of Energy's Office of Nuclear Energy

INL/RPT-25-87430

MegaWatt Mayhem: Grid Operator Challenges with Large Loads

Existing Landscape and Grid Stability Implications

Luis Fernando Enriquez-Contreras,
S M Shafiul Alam,
Timothy R. McJunkin,
Megan Culler,
Jonathan Tacke, and
Emma Stewart
Idaho National Laboratory

SEPTEMBER 2025

Idaho National Laboratory
Center for Securing Digital Energy Technology
Idaho Falls, Idaho 83415

http://www.inl.gov

Prepared for the
U.S. Department of Energy
Grid Deployment Office
Under DOE Idaho Operations Office
Contract DE-AC07-05ID14517

Page intentionally left blank

ABSTRACT

This report provides a summary of the challenges faced by United States electricity grid operators in accommodating and anticipating the rapid deployment of large loads, particularly data centers, based on academic literature and industry working groups. The report highlights the unique requirements and operational characteristics of data centers, which differ significantly from traditional industrial loads. Key issues addressed include utility planning considerations, with emphasis on the implications for grid operators, impacts to normal operations for grid operators, reliability considerations during periods of grid stress, and resilience considerations for the changing operational paradigms based on data centers. Real-world examples are used to highlight these challenges and the changes that grid operators must address. The findings underscore the necessity for coordinated efforts and innovative solutions from both grid operators and regulatory bodies to ensure the stable integration of large loads into the grid. This report is the first in a series that will explore the challenges of data center deployments based on several key power system perspectives.

Page intentionally left blank

CONTENTS

TABLE OF FIGURES

Page intentionally left blank

ACRONYMS

4CP	four coincident peak
AGC	Automatic Generation Control
AI	artificial intelligence
BESS	battery energy storage system
BTM	behind-the-meter
C2G	chip-to-grid
CAISO	California Independent System Operator
CIFP	Critical Issue Fast Path
CMLD	composite load modules
Co-Op	electric cooperatives
CPU	central processing unit
DOE	U.S. Department of Energy
EPRI	Electric Power Research Institute
ERCOT	Electric Reliability Council of Texas
FERC	Federal Energy Regulatory Commission
FIDVR	Fault-Induced Delayed Voltage Recovery
GPU	graphical processing unit
GW	gigawatt
IBR	inverter based resource
IEEE	Institute of Electrical and Electronics Engineers
IRP	integrated resource plan
INL	Idaho National Laboratory
IT	information technology
LLI	large load interconnection
MISO	Midcontinent Independent System Operator
ML	machine learning
Munis	municipal utilities
MVAR	mega volt-ampere reactive
MW	megawatt
NARUC	National Association of Regulatory Utility Commissions
NERC	North American Electric Reliability Corporation
NREL	National Renewable Energy Laboratory
PJM	Pennsylvania New Jersey Maryland Interconnection, LLC
PUE	power usage effectiveness
RTO	regional transmission organization

SCED	Security Constrained Economic Dispatch
SLA	Service level agreement
STATCOM	Static Synchronous Compensator
UFLS	under-frequency load shedding
UPS	uninterruptible power supply
U.S.	United States

Page intentionally left blank

MegaWatt Mayhem:

Existing Landscape and Grid Stability Implications

1. INTRODUCTION: GRID OPERATORS' CHALLENGES FROM LARGE LOADS

Over the past century, electric grid operations in the United States (U.S.) have been shaped by predictable patterns of load growth, dominated by residential, commercial, and industrial demand that evolved incrementally over time, well documented by load growth forecasting. Large loads, typically industrial facilities like manufacturing, oil & gas, and agricultural, were few in number, geographically dispersed, and often planned years in advance in close coordination with utilities. These loads, while significant, were generally integrated into the grid with well-understood operational practices and planning frameworks. However, this paradigm is shifting rapidly. Grid operators, the entities responsible for monitoring and controlling the electrical grid in a region to ensure a stable and reliable supply of electricity, are now facing a new class of large, high-density, and often clustered loads that are emerging at unprecedented speed and scale.

Data centers exemplify this transformation. Unlike traditional industrial loads, which might range from 50 to 500 megawatts (MW) (defined by the U.S. Department of Energy (DOE) as anything larger than 100MW [1]) and are often tied to specific manufacturing or resource extraction activities, data centers can exceed 1,000 MW at a single site and are frequently deployed in geographic clusters that collectively demand several gigawatts of capacity [2]. These clusters are appearing in regions that may not have been historically designed for such concentrated demand, creating new challenges for transmission planning, interconnection, and operations. The shift from a grid dominated by diffuse, predictable demand to one increasingly shaped by large, fast-moving, and digitally-driven loads represents a fundamental change in how grid operators must maintain reliability.

Assessing and addressing grid operator challenges with data centers is essential to supporting rapid grid growth and deployment of large loads. As of 2025, U.S. data centers consume approximately 4.4% of the nation's total electricity, a figure projected to rise to between **6.7% and 12% by 2028**—equating to a total power demand of **74–132 gigawatts (GW)**, depending on the scenario. Visualized in recent Lawrence Berkeley National Lab's 2024 Energy Usage Report [3], these statistics underscore the transformation of data centers from niche technical assets to a dominating force for large load integration, which grid operators must address with urgency.

Understanding these challenges is critical so that the United States remains competitive in the artificial intelligence race. Notably, resource adequacy outlook has evolved quickly based on expectations of load growth in data centers. A gross analysis of the number of times the term "data center" is used in the National Electric Reliability Council's (NERC's) 2023 Long Term Reliability Report [4], 6 times, versus the 2024 report, 26 times indicates the acceleration of the concern. Furthermore, demand growth in part due to data centers is mentioned in the risk summary for 7 of the 11 elevated risk regions. This report reviews the challenges faced by electric utilities and regulators as they work to connect more data centers to the grid and continue to ensure the reliability of the grid for all customers. It is essential to identify these problems so that solutions can be proposed that properly address the needs of the data centers, utilities, electrical regulators, and the general public.

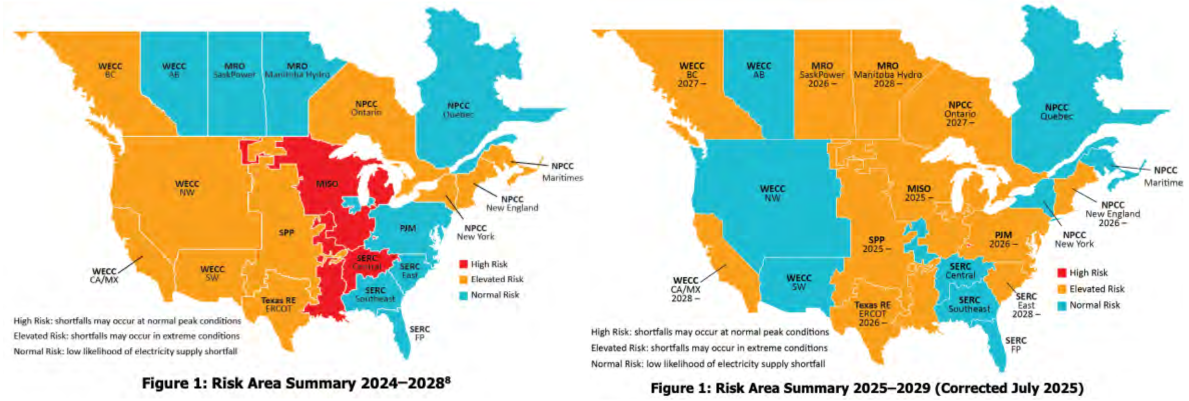


Figure 1: Risk Area Summary 2024–2028[8]

Figure 1: Risk Area Summary 2025–2029 (Corrected July 2025)

Figure 1: Risk summary from NERC's Long Term Reliability Report [4].

2. LARGE LOADS OVERVIEW

Data center loads are fundamentally different from traditional large load growth both in scale and impact. Note that data centers can come in many sizes; some may be purpose built for particular applications or dedicated for private organization use. In the context of this report, references to data centers refers to data centers which qualify as large loads, which may not include small, private facilities. This section explores four defining characteristics that distinguish these modern large loads: **size**, **location**, **demand shape**, and **uptime requirements**.

2.1. LARGE LOAD SIZE

The most visible difference is the sheer magnitude of power demand. Traditional large industrial facilities typically range from 50 to 500 MW, with power distributed across a sprawling physical footprint, such as a manufacturing campus or processing plant. In contrast, hyperscale data centers are increasingly scoped at 500 MW or more, with some sites exceeding 1,000 MW. For example, a recently announced project in Abilene, TX, will construct a single data center drawing 1.2 GW, which is expected to be completed by mid-2026 [5]. These gigawatt-scale facilities are no longer outliers but are becoming the new design norm.

A single data center campus can require dedicated substations, high-voltage transmission lines, and contingency planning that historically would have been reserved for major generation assets. The power density is also significantly higher, with dense racks of information technology (IT) equipment and cooling systems. This shift forces utilities to treat these loads not as incremental additions but as system-defining elements that can reshape reliability operations.

2.2. LARGE LOAD LOCATION

Unlike legacy industrial loads, which were typically sited near relevant resources (e.g. oil, minerals, crops), labor, or transportation hubs, modern large loads are often located based on digital infrastructure needs. Proximity to fiber optic networks, low-cost and reliable electricity, favorable tax incentives, and access to cooling resources (e.g. water) are now the primary siting drivers.

This decoupling of load siting from grid readiness has created a geographic mismatch. Many data centers are being proposed in rural or semi-rural areas with limited existing transmission capacity or weak substations. These regions were not designed to host gigawatt-scale demand, and the resulting expected strain on local infrastructure can delay interconnection timelines and increase costs. This mismatch between load siting and grid readiness creates significant challenges for utilities, who must now plan for high-impact demand in regions that were not designed to host it. The result is a growing tension between economic development goals and grid reliability constraints. For example, PJM has seen notable increases in capacity pricing, growth which is strongly correlated with data center deployments in the region [6] [7].

2.3. DEMAND SHAPE CHARACTERISTICS

The operational profile of large loads is also distinct. Some data centers, such as cloud servers, operate with a high and flat load factor, consuming power 24/7 with minimal variation. Other data center types, such as those used for AI, can vary their loading profiles suddenly and dramatically. Still others, like cloud hosts for online retail, may be typically high and flat, but experience capacity-driven peaks related to events not typically tracked by utilities, such as large-scale online sales. This contrasts with traditional industrial loads, which often follow production schedules and exhibit clear but predictable peaks and valleys in demand; load profiles and expectations are also communicated directly to utilities through defined interactions. While this flatness can simplify some aspects of forecasting, it reduces system flexibility and increases the need for fast-ramping generation or storage to maintain grid balance.

Moreover, some AI-driven workloads introduce short-term volatility. While the average load may be stable, certain compute-intensive tasks can cause rapid spikes, posing challenges for frequency regulation and voltage stability. These characteristics are not yet fully accounted for in traditional load modeling or interconnection studies, creating uncertainty in both short-term operations and long-term planning.

2.4. UPTIME REQUIREMENTS

Finally, uptime considerations for many types of data center are high priority. Unlike traditional industrial facilities that may operate on shifts or tolerate downtime, many data center loads require 24/7 availability with minimal tolerance for outages or voltage fluctuations. Many facilities target "five nines" reliability (99.999%), which translates to less than 5.3 minutes of downtime per year. Any interruption, whether due to grid faults, voltage sags, or planned outages, can result in data loss, service-level agreement (SLA) violations, or financial penalties.

To meet these expectations, data centers often request dual feeds from separate substations, deploy on-site diesel generators, and increasingly incorporate battery energy storage systems (BESS) for seamless backup and power quality management. Data centers need both to ensure their systems do not lose power suddenly (or at all) and receive the high power quality needed to protect sensitive power electronics in computing resources. In contrast, traditional industrial facilities may tolerate brief outages or schedule downtime around grid maintenance windows. This divergence in reliability expectations requires grid operators to rethink how they define and deliver resilience, not just at the system level, but at the individual node where these loads connect.

2.5. COMPARISON OF DATA CENTERS TO LARGE INDUSTRIAL LOADS

This section summarizes the key characteristics of large loads and what separates data centers from traditional large industrial loads.

Table 1: Summary of differences between hyperscale data centers and traditional large industrial loads.

Category	**Characteristic**	**Hyperscale Data Centers**	**Traditional Large Industrial Loads**
Load size	Overall Power Demand	**Very High.** Data centers are increasingly scoped to be 500 MW or larger, with gigawatt-scale centers no longer unusual.	**High.** Traditional industrial previously were the largest loads connected to the grid. However, they pale in comparison to max demand of hyperscale data

			centers. Industrial loads are typically in range of 50-500 MW per facility.
	Power Density	**Extremely High.** Compact footprint with dense racks of IT equipment and cooling systems.,	**High.** Power draw is spread out over a larger physical footprint, such as across a factory campus, and is associated with specific machinery and production lines
	Growth Trends	**Accelerating.** Driven by AI, cloud computing, and digital infrastructure demands. Load projections are increasing rapidly year over year.	**Incremental**. Growth is tied to economic cycles and industrial demand. Historically, build rates have been slow or negative, but may accelerate in the future.
Location	Geographic distribution	**Regionally Dense.** Clustered near fiber backbones, cheap power, and favorable climates	**Nationally Distributed**. Located near raw materials, labor, or transportation hubs.
	Siting Drivers	**Cost.** Proximity to fiber, low-cost and reliable power, tax incentives, and cooling efficiency.	**Labor and Logistics**. Proximity to supply chains, labor, raw materials, and logistics infrastructure.
Demand shape	Load Factor	**High and flat.** Data centers operate 24/7 with a steady base load as IT equipment is rarely fully "off" and needs constant cooling. The overall load factor can vary depending on the type of compute jobs, e.g. cloud computing vs. retail servers vs. AI training jobs.	**Variable and lower.** The load factor fluctuates with production schedules and shifts, creating distinct peaks and valleys in demand. Demand can drop significantly during nights, weekends, and holidays.
	Load Variability	**Potentially volatile.** While the overall load is stable, some AI-driven workloads can cause rapid and significant load fluctuations (as much as 90% in seconds), which can be disruptive to the power grid.	**Predictable and slower.** Load changes correspond to factory operations and are more predictable in their timing and duration.
	Power Demand Composition	**IT and cooling.** A data center's power is split between the IT equipment and high-intensity cooling systems, with the cooling load varying with external weather.	**Production-centric.** Power demand is tied directly to manufacturing, processing, or assembly machinery, as well as general facility needs like lighting and cooling.
	Flexibility Potential	**High.*** Large companies with multiple data centers are designed to shift compute jobs	**Moderate.** Some industrial processes can be paused or

		across regions or fail over to other facilities. Many data centers have significant backup capacity, which gives them potential to leverage on-site resources and act as flexible loads from the grid's perspective. *Note: High potential for flexibility does not mean that all data centers choose to operate flexibly. Additionally, some data centers may have dedicated use at a single facility that limits flexibility.	shifted, especially during demand response events.
Uptime Requirements	Reliability	**Extremely high.** Data centers require near-perfect uptime (often 99.999%), with any interruption being economically catastrophic. They often use advanced battery backup systems and on-site generators to ensure uninterruptible power.	**Moderate to high.** While crucial for factory operations, reliability requirements are typically less stringent than those for data centers. Brief, scheduled outages are often manageable.
	Backup Power	**Extensive.** On-site diesel generators, uninterruptable power supply (UPS) systems, and increasingly BESS	**Moderate.** May have backup generators, but not always to the same scale or redundancy.
	Grid Service Expectations	**High.** Often require dual feeds, voltage regulation, and priority restoration.	**High.** Many traditional large loads also have high grid service requirements that have been negotiated in plans with the utility.

3. CHALLENGES FOR GRID INTEGRATION AND OPERATORS

3.1. PLANNING AND COST

As a result of the new large loads, utilities are being forced to adapt their planning processes in real time, often without the benefit of long-term visibility or regulatory clarity. The challenges are multifaceted: interconnection queues are growing longer and more complex; infrastructure upgrades are becoming more capital-intensive and time-sensitive; and impact studies are increasingly strained by the scale and speed of proposed developments.

This section explores three critical dimensions of this evolving landscape: the interconnection process, the rising costs of infrastructure expansion, and the regulatory and procedural hurdles associated with permitting and impact assessments. Together, these issues highlight the urgent need for updated planning tools, cost allocation frameworks, and policy coordination to ensure that the grid can accommodate large loads without compromising reliability or affordability.

3.1.1. INTERCONNECTION

The interconnection process has become a critical pressure point for grid operators as they work to accommodate the rapid influx of hyperscale data centers. While interconnection agreements are typically negotiated by utility planning departments, they have direct and lasting implications for grid operations. These agreements define not only the physical connection to the grid but also the operational expectations, such as allowable ramp rates, voltage support, redundancy, and contingency behavior, that grid operators must manage in real time. This section explores how the interconnection process is evolving and how emerging policy responses, such as Texas Senate Bill 6 of Legislative Session 89(R) [8], are beginning to address the need for better coordination, transparency, and operational foresight.

The accelerated surge in demand has led to a backlog in interconnection requests, particularly in regions where transmission infrastructure is already constrained. The complexity of evaluating these requests, especially when multiple large-scale projects are proposed in the same area, can delay or even stall critical infrastructure development. Compounding the workload intensity of requests, it is not uncommon for prospective data centers to submit requests in multiple locations, making a final decision for siting based in expected costs and interconnection timeline [9]. For grid operators, this means preparing for a wide range of high-impact load behaviors, some of which may never come online, while still maintaining system stability and resource adequacy. In one response to this challenge, Texas now requires disclosure of duplicative interconnection requests as part of Senate Bill 6 [10], which will help organizations like the Electric Reliability Council of Texas (ERCOT) with more accurate forecasting and planning.

Small electric cooperatives (Co-Ops) and municipal utilities (munis), the rapid expansion of data centers are working to manage interconnection queues while operating with leaner engineering teams and fewer financial and technical resources compared to larger investor-owned utilities [11].. The sudden influx of such requests can quickly saturate their planning and operations bandwidth, leading to delays and uncertainty for both the utility and the data center developers.

Moreover, these smaller utilities often serve rural or semi-rural areas where transmission infrastructure is limited or outdated. Accommodating a large data center may require significant upgrades to substations, feeders, or even regional transmission lines, projects that are capital-intensive and time-consuming. Navigating the regulatory, environmental, and community engagement processes for these upgrades can further extend interconnection timelines. However, data centers bring big benefits to these smaller utilities too. Data centers can bring economic value to a region though jobs and industry growth, along with the direct income from data centers as large ratepayers. Without streamlined processes or external support, small co-ops and munis may find themselves caught between the urgency of economic development opportunities and the technical realities of grid readiness, making it difficult to meet the expectations of fast-moving data center developers.

3.1.1.1. CASE STUDY: DATA CENTER INTERCONNECTION QUEUES IN VIRGINIA

Virginia, particularly Northern Virginia, has become a global hub for data centers, but this rapid growth has exposed significant challenges in grid interconnection. Dominion Energy, the region’s primary utility, has faced mounting pressure as hyperscale data centers request large, concentrated loads that exceed the capacity of existing infrastructure. The interconnection queue has grown increasingly congested, with some substations reaching their limits and requiring major upgrades or new transmission lines—projects that can take years to complete [12] .

The situation has drawn the attention of federal regulators. In February 2025, the Federal Energy Regulatory Commission (FERC) launched a formal review of co-location issues in Pennsylvania New

Jersey Maryland Interconnection, LLC (PJM), the regional transmission organization that includes Virginia. FERC found that PJM’s tariff lacked sufficient clarity on how to handle large load interconnections—particularly those involving AI-enabled data centers co-located with generation assets—and directed PJM and its transmission owners to justify or revise their interconnection rules to ensure grid reliability and fair cost allocation to consumers [13]. This regulatory scrutiny underscores the broader challenge: the pace of digital infrastructure deployment is outstripping the ability of grid planning and policy to keep up. In response, the PJM Board of Managers launched the Critical Issue Fast Path (CIFP) process on August 8, 2025 [14]. While the specifics of this process are still in development, this initiative aims to rapidly develop reliability-focused solutions that enable timely interconnection of large loads without compromising resource adequacy.

3.1.2. INFRASTRUCTURE COSTS

For utility planners, the difficulty of addressing the costs of infrastructure upgrades needed to integrate data centers lies not only in securing the capital and coordinating construction, but also in determining how to allocate these costs fairly and transparently. When data centers request interconnection in areas with limited existing capacity, the utility must often build infrastructure that benefits a single customer but is paid for by a broader rate base, raising concerns about cost recovery and equity. A clear protocol for determining financial responsibility for electrical upgrades when a data center load is connected has yet to be established [15], however, strong models exist for doing this with clusters of generation resources, like solar, wind, or storage sites [16]. This challenge is especially acute for smaller utilities and cooperatives, which may lack the financial flexibility to absorb large upfront costs or the regulatory mechanisms to recover them efficiently. Even for larger utilities, the traditional cost allocation frameworks, designed for incremental load growth, are being tested by the sudden, concentrated demands of hyperscale data centers. Therefore, establishing grid repayment mechanisms in case of data center project cancellations is essential before any load is connected. Furthermore, there are concerns regarding the sufficiency of transmission and distribution capacity for data centers to connect them to generation sources.

Local residents have pushed back against the rapid expansion of data centers, citing concerns over rising utility costs, environmental impacts, and the burden of infrastructure upgrades. These groups argue that the costs of new substations, transmission lines, other infrastructure being unfairly passed on to ratepayers, despite the benefits accruing primarily to private developers and tech companies. The fear of increased cost is not unfounded. A 2024 study from the Virginia Joint Legislative Audit and Review Commission found that while data centers were currently paying their fair share of electricity rates, the expected growth in coming years would “likely increase system costs for all customers, including non-data center customers.” [17] This increase is expected to impact customers by $14-37 per month by 2040 without accounting for inflation in northern Virginia [17]. Across the nation, over a dozen bills have been proposed by state lawmakers that seek to ensure that data center growth does not result in higher rates for other customers [18] . There is a broad consensus on the need for data centers to provide proper compensation to the utility to ensure that costs are not passed down to other consumers in the utility network.

3.1.2.1. CASE STUDY: UPDATES TO INTERCONNECTION REQUIREMENTS IN TEXAS

Senate Bill 6 (SB6), signed into law in July 2025, makes several updates to interconnection request requirements for data centers and other large loads which attempt to address several of the interconnection challenges, including payment for upgrades, for large loads. ERCOT projects that statewide power demand could nearly double by 2030, driven largely by increased interconnection requests from large-load customers, including data centers and Permian Basin operators electrifying their operations [19], making the need for interconnection reform to address these large load customers an urgent issue to address. The bill requires the public utility commission (PUC) to ensure that large load customers contribute to the cost of interconnection, mandates that electric cooperatives and municipal utilities pass through those costs to the large-load customers, and calls for a reevaluation of the Four Coincident Peak (4CP) methodology used to set transmission rates [10]. It requires that applicants demonstrate financial commitment to the interconnection request through payment of an upfront fee ($100,000), control of the site (e.g. prove ownership or lease commitment) and provide proof that they can financially contribute to any required transmission upgrades. It also introduces a minimum transmission charge for behind-the-meter customers with on-site generation and requires large transmission-level customers to install equipment that enables load curtailment during firm load shed events to protect residential reliability,

3.1.3. IMPACT STUDIES AND PERMITTING

The permitting and billing process for adding new interconnections includes various obstacles such as land rights, environmental regulations, public opposition, lead times for equipment, and general permitting [15] [20] [21]. The permitting of grid interconnections for large data center loads has become an increasingly difficult problem to address.

Impact studies, a key piece of interconnection approvals, evaluate the reliability impacts of new generation and large loads to ensure the grid can handle the addition. High fidelity models are needed to ensure the accuracy of these studies. However, data center impact studies rely heavily on user-made models exist since there is a lack of commercially made composite load models (CMLD) on PSSE and PSLF [22]. The lack of validated models complicates interconnection studies and further delays permitting. Currently, existing models lack accuracy and many features are absent including voltage swell consideration, time delay, and frequency consideration in the load disconnect and reconnect logics of PSSE’s CMLD [22]. PSLF CMLD are also lacking accuracy and mismatching has been detected in the dynamic response [23].

The typical legal process takes anywhere from three to seven years to complete, depending on the electrical jurisdiction, such as ERCOT or FERC [24]. Data center operators, however, are pushing for deployment within six months to a year [24]. The industry has established a workaround by using "behind-the-meter" services [24]. In this process, the data center operator forgoes traditional transmission interconnections by establishing a direct electrical connection with a large generation source. This has sparked a significant debate among stakeholders, including utilities, operators, and policymakers, on how to approach this new connection method. Some argue that behind-the-meter services should be banned outright, contending that generation services are still grid-connected, and thus, the data center remains a grid-connected service. There is also a legal debate on whether behind-the-meter connections fall under FERC or regional transmission organization (RTO) regulations. Others take a less restrictive approach, suggesting that data centers should install large generation sources before interconnecting and/or pay for the transmission costs upfront. The misalignment between the speed of digital infrastructure deployment and the pace of regulatory approvals puts grid reliability and economic development at risk, underscoring the need for more streamlined and coordinated permitting frameworks.

3.1.3.1. CASE STUDY: PERMITTING CHALLENGES ACROSS THE COUNTRY

A report by Data Center Watch [25], a research firm funded by 10a Labs, which provides research and analysis for artificial intelligence companies and U.S. tech firms, identified multiple instances of permitting issues that contributed to cancellation or delays of data center projects. In Cascade Locks, Oregon, public backlash over a proposed $100 million facility by Roadhouse Digital led to the recall of two Port Authority board members in June 2023. The newly elected board subsequently terminated the project the following month. Similarly, in Warrenton, Virginia, voters ousted the entire town council in the November 2024 election due to their support for an Amazon data center proposal. The incoming council, composed entirely of opponents to the project, now holds the authority to halt its progress. The report also points to broader tensions in Virginia, where Delegate Josh Thomas introduced HB1601—a bill aimed at tightening oversight of data center siting by requiring applicants to evaluate potential impacts on natural and cultural resources. The legislation was driven by resident concerns over noise pollution and environmental degradation linked to data center expansion.

These examples underscore the broader challenge that even well-capitalized, experienced developers face when attempting to bring large-scale data centers online in regions with limited grid headroom. The permitting and interconnection study processes, often involving multiple layers of regulatory oversight and technical analysis, can stretch timelines by months or even years. For grid operators, permitting delays introduce uncertainty into long-term planning. Grid operators rely on predictable timelines to coordinate system impact studies, resource adequacy assessments, and capital investment cycles. When permitting is unpredictable or prolonged, it disrupts these processes and can lead to stranded assets, inefficient grid buildouts, or missed opportunities to align infrastructure with load growth. In a landscape where data centers are requesting interconnections at unprecedented scale and speed, permitting has become not just a regulatory hurdle—but a critical determinant of whether the grid can keep up.

3.1.4. LONG-TERM FORECASTING AND RESOURCE ADEQUACY

The rapid growth of hyperscale data centers is reshaping long-term load forecasting and challenging traditional approaches to resource adequacy planning. For grid operators, accurate long-term forecasts are essential to ensure that sufficient generation and transmission capacity is available to meet future demand reliably. However, the scale, speed, and unpredictability of large load development, particularly from data centers, are introducing new uncertainties that existing planning frameworks are not well-equipped to handle.

These facilities often request hundreds of megawatts of capacity with relatively short lead times, and their siting decisions are driven by factors largely outside the utility's visibility, such as access to fiber infrastructure, tax incentives, or land availability. This disconnect between developer decision-making and utility planning timelines makes it difficult for grid operators to anticipate where and when large loads will materialize. As a result, operators face a dual risk: underestimating future demand, which can lead to reliability shortfalls and emergency procurement, or overestimating it, which can result in overbuilt infrastructure and unnecessary costs for ratepayers.

The consequences of this uncertainty are already visible. For example, neither the Midcontinent Independent System Operator (MISO) nor the California Independent System Operator (CAISO) included substantial data center growth in their 2023 long-term forecasts [15]. This omission highlights a broader issue: many system operators are still catching up to the pace and scale of digital infrastructure development. Without more accurate and timely data on large load projects, resource adequacy studies may fail to reflect the true needs of the system.

To address this, grid operators are increasingly seeking earlier and more detailed engagement with large load customers. This includes requesting firm commitments, clearer timelines, and more transparent communication about siting decisions. By improving coordination between developers and operators, and by integrating large load scenarios into long-term planning models, utilities can better align infrastructure investments with actual system needs, ensuring reliability while avoiding unnecessary costs.

3.1.4.1. CASE STUDIES: PICK YOUR POISON – CHALLENGES WITH OVERESTIMATES AND UNDERESTIMATES

- A recent ERCOT forecast received pushback that it may have overestimated load growth. A May 2025 report by Ascend Analytics [26] found it unlikely that ERCOT's predicted load growth, driven in large part by data centers deployments, would be achieved. The primary barrier they identified was supply chain challenges around the infrastructure and workforce needed to actually build, deploy, and connect that much load to the grid on the proposed timeline.

- Georgia Power Company filed their integrated resource plan (IRP) in fall of 2023, which included a significant increase in near-term load growth predictions. Microsoft filed comments on the IRP pushing back against these predictions, challenging Georgia Power's modeling methods and claiming that the IRP both undervalued renewable energy and overestimated data center growth [27].
- According to NARUC [28], Arizona noted that forecasting data center load growth remains speculative, given uncertainties like fluctuating data center load estimates and emerging technologies such as Deep Seek. Arizona has three major utilities, and they collaborate closely on generation and transmission planning, as what happens to one utility often affects the other. The Commission suggested that utilities provide annual updates. Arizona utilities were criticized for overestimating growth in previous iterations of their IRPs but now there's a concern they are underestimating.
- Loudoun County, VA, and particularly "Data Center Alley" in eastern Loudoun, has the highest concentration of data centers in the world with approximately 200 data centers built and 117 in the development pipeline. In 2022, PJM informed Dominion Energy that it had underestimated power infrastructure needs. Dominion Energy subsequently informed "Data Center Alley" it will have less power available for at least five years. The power "constrained area" is created until new power lines can be built by 2027.

Forecasting issues exist around the country, with experts assessing that some forecasts overestimate large load growth, while other forecasts are underestimated. This growing divergence between actual and forecasted load growth underscores the urgent need for more adaptive and transparent forecasting methodologies. In a Black & Veatch survey, 45% of utility respondents stated that they have no confidence or are not very confident in their forecasting for data center loads [29]. As data center demand accelerates and becomes more volatile, grid operators must evolve their planning tools to avoid both underbuilding that risks reliability and overbuilding that burdens ratepayers

3.1.5. PLANNING SOLUTIONS FOR GRID OPERATORS

Solutions must go beyond administrative reforms and directly support the operational readiness of the grid. First, interconnection processes should include early and ongoing coordination between planning and operations teams. Interconnection agreements should explicitly define operational parameters, such as ramp rates, redundancy expectations, and curtailment protocols, and these should be reviewed and validated by grid operators before agreements are finalized. This ensures that operational impacts are not overlooked in the rush to approve new projects.

Second, cost allocation frameworks must be modernized to reflect the unique demands of large loads. Grid planners and operators need clear mechanisms to recover the costs of infrastructure upgrades, especially when those upgrades are driven by a single customer but impact system-wide reliability. This includes establishing repayment mechanisms in the event of project cancellations and ensuring that large loads contribute proportionally to the cost of transmission and distribution enhancements. For smaller utilities and co-ops, pooled funding models or state-level infrastructure grants may be necessary to avoid overburdening limited local resources.

Third, permitting and impact study processes must be streamlined and standardized. Grid operators rely on predictable timelines to coordinate system impact studies, resource adequacy assessments, and capital investment cycles. Delays in permitting introduce uncertainty that can lead to stranded assets or misaligned infrastructure. Solutions may include the development of fast-track permitting pathways for projects that meet predefined technical and siting criteria, as well as the creation of validated models for use in simulation tools. These models would allow operators to more accurately assess the dynamic behavior of data center loads and reduce delays in study completion.

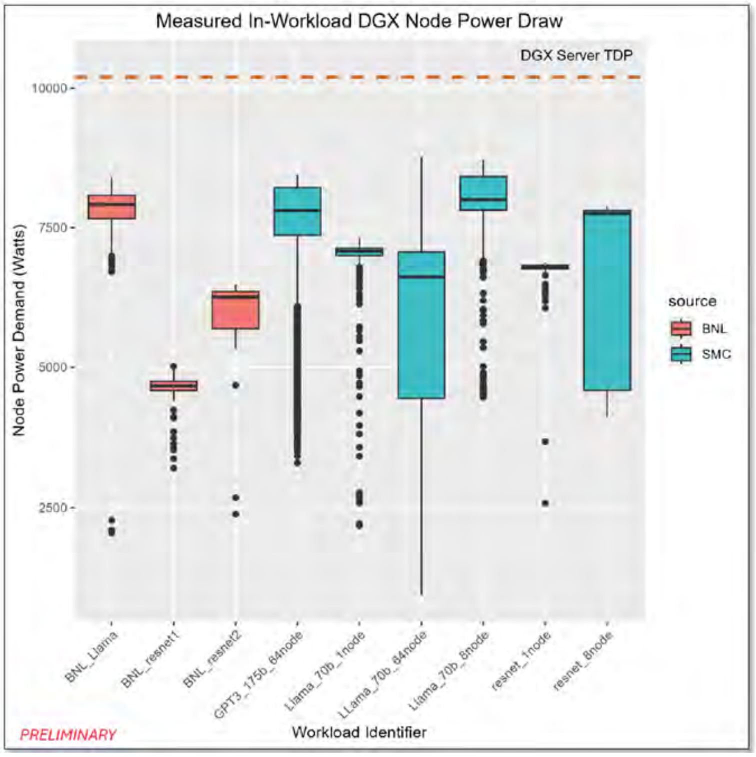
Ultimately, grid operators need more than just faster processes; they need better visibility, stronger coordination with planning teams, and tools that reflect the operational realities of integrating large, inflexible, and high-uptime loads. Without these reforms, the pace of digital infrastructure deployment will continue to outstrip the grid's ability to respond, putting both reliability and economic development at risk.

3.2. NORMAL OPERATIONS

While much attention has focused on planning and interconnection challenges from hyperscale data centers, their operational impact on grid management is equally significant. This section examines how these large, high-uptime loads are reshaping forecasting, dispatch, and real-time balancing. Load Forecasting Uncertainty

Accurate load forecasting is essential for grid operators to maintain system reliability, ensure resource adequacy, and manage real-time grid stability. The rapid growth and unique characteristics of hyperscale data centers introduce new layers of uncertainty across long-term planning, day-ahead scheduling, and real-time operations. This section focuses on the latter two, issues that grid operators deal with for near-term time horizons.

While data centers typically exhibit high and flat load profiles, certain workloads, especially AI training or batch processing, can introduce sudden, concentrated spikes in load that are difficult for grid operator's to anticipate and even harder to model. Although research has been conducted to model the load behavior of a single node in an AI model as illustrated in Figure 2, comprehensive modeling of entire data center operations remains complex and largely underexplored, making forecasting difficult [30].



Source: Newkirk, Alex C., Jared Fernandez, Jonathan Koomey, Emma Strubell, and Constantine Samaras. 2024. "AI Energy Use: Empirically Grounded Estimate of Training IT Power Demand and Pathways for Future Research."

Figure 2: Estimated power consumption at the node level of the different AI models. The analyses lack energy usage at the MW level, but the variation in average energy demand and range for different AI models highlights how at the MW level, greater uncertainty could occur [30].

The unpredictable and sudden variations may not always be visible to operators, particularly when data centers use behind-the-meter generation or storage that masks their net load. Without granular telemetry or coordination agreements, operators are left to manage system balance with incomplete information, increasing the risk of reserve shortfalls, inefficient dispatch, or frequency instability. This is an area currently being explored as part of DOE national laboratory research.

For grid operators, the core issue is not just uncertainty—it's unpredictable peak capacity use. Unlike traditional loads that follow well known economic or seasonal patterns, data center demand is driven by opaque digital workloads and business decisions that are often invisible to the grid. This makes forecasting not just harder, but fundamentally less reliable, requiring new tools, data-sharing protocols, and operational safeguards to maintain system stability.

It is worth noting that AI workloads still represent only a small percentage of overall data center use. Goldman Sachs estimates that 54% of data center workloads are traditional cloud computing applications, 32% are business functions like email and storage, and 14% are AI-driven [31]. As AI continues to rise and data centers process large AI workloads, the power variations will likely increase. However, power use is costly, and technology innovation focuses on increasing chip performance at lower cost and lower power consumption [32] [33]. Assessing the impact of AI data centers as an overall component to data center loads as well as the variability that these workloads introduce is something that will need to be continuously monitored by utilities.

3.2.1. GENERATION DISPATCH CONSTRAINTS

As discussed in detail already, data centers demand high, continuous power delivery with minimal tolerance for curtailment or interruption. Both mandatory curtailment and voluntary demand response programs are used for large loads, but analysis in some regions has suggested that the largest loads, hyperscale data centers, are less likely to participate in wholesale markets and demand response [34] [35] [36]. Curtailment and demand response are key levers that operators have historically used to manage grid stress. If large loads do not engage in voluntary programs or are not subject to mandatory shedding requirements under certain conditions, it places a greater burden on the supply side of the system.

Automatic Generation Control (AGC) plays a central role in this balancing act. AGC is a system used by grid operators to control the output of multiple generators at different power plants in response to changes in the load. It helps maintain the balance between supply and demand, ensuring the stability of the power grid. The timeframes for generator ramps can vary significantly depending on the type of generator. Traditional thermal power plants, for example, have slower ramp rates compared to more flexible resources like hydroelectric plants or battery storage systems. AGC systems need to account for these differences to effectively manage grid stability. Fast-ramping resources are crucial for responding to sudden changes in demand or supply. Control operations in the context of AGC and generator ramps include primary control, which involves the immediate response of generators to frequency deviations; secondary control, where AGC adjusts the output of generators to restore the frequency to its nominal value over a few minutes; and tertiary control, which involves manual adjustments to the generation schedule to ensure long-term balance and stability. Certain control operations may not be adequate for managing the rapid and unpredictable variations in demand seen with modern loads like data centers. For example, generators with slow ramp rates may not be able to respond quickly enough to sudden changes in demand, leading to frequency instability. Traditional control operations may also lack the flexibility needed to handle the dynamic nature of modern loads, and poor coordination between different control operations can lead to inefficiencies and increased risk of grid instability.

Furthermore, the operational constraints embedded in interconnection agreements can limit the operator's ability to respond dynamically. These agreements may guarantee firm service or restrict curtailment, even during emergencies, forcing operators to prioritize large loads over more smaller more responsive demand. In regions with high concentrations of data centers, this can distort dispatch decisions, requiring earlier unit commitments or longer generator runtimes, which in turn increase system costs and emissions.

3.2.1.1. CASE STUDY: NEW DEMAND RESPONSE PRODUCTS

In order to integrate data center load variability with existing generation dispatch methods, which include day-ahead and hourly dispatch decisions, some RTOs are considering new demand response participation programs or curtailment rules for large loads.

Texas SB6, mentioned earlier, will create a new reliability service in ERCOT to procure curtailment from large loads. Under this service, large load customers will be given at least 24 hours advanced notice. This allows ERCOT to proactively address anticipated generation shortfall. This new service will be applicable for loads 75 MW or more, and is off-limits to any large-load customer that “curtails in response to the wholesale price of electricity… or that otherwise participates in a different reliability or ancillary service.” [35].

In a similar move, PJM is exploring a new class of demand response for large loads. Their considerations would reflect the limited duration and reliability of operating backup generation for large loads, recognizing that these large loads would still need to have power, whether or not they were connected to the grid. It would also enhance and educate on existing opportunities to aggregate demand response across multiple data centers not geographically in the same location [37].

3.2.2. RAPID VARIATIONS IN DEMAND

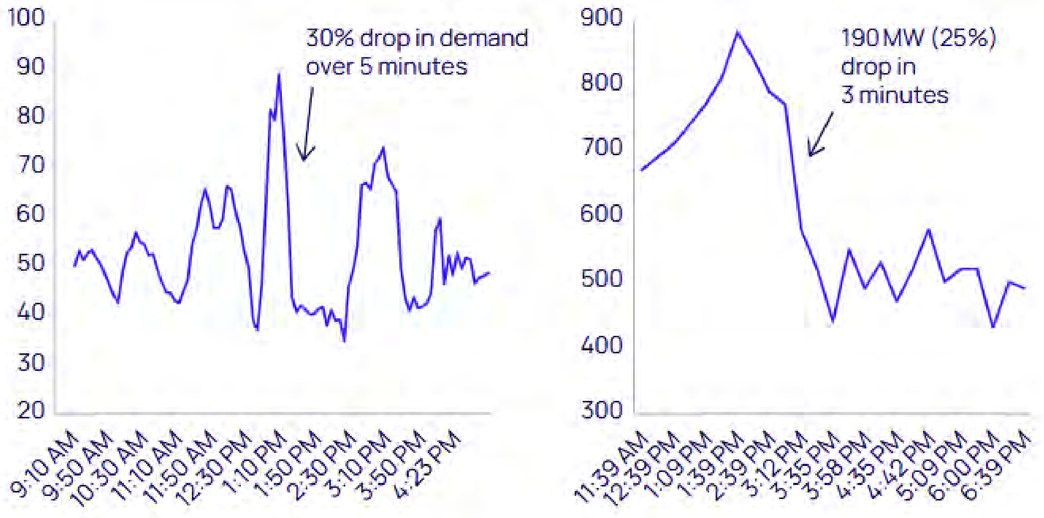
Data centers, especially those utilized for AI applications, can exhibit rapid and unpredictable variations in demand due to dynamic workload scheduling and the nature of AI training processes. This behavior poses significant challenges for grid operators, who must maintain voltage and frequency stability without the foresight of when these fluctuations will occur.

Once a data center is commissioned, its interactions with the grid present further challenges. Data centers can rapidly increase and decrease their demand from the grid due to fluctuating computing jobs, and a growing number of large loads can change their MW consumption rapidly enough to exhaust available regulation service. EdgeTunePower reported instances of steep ramps in power consumption for a 50 MW data center, with one instance showing a 24 MW ramp-up in approximately 0.3 seconds [38].

Another major consideration for data center operation is the risk of oscillations associated with data centers. This challenge is further exacerbated with the higher use of graphics processing units (GPUs) for artificial intelligence units compared to non-artificial intelligence data center that rely more on central processing units (CPUs). Data centers utilizing GPUs for AI applications have the difficulty that GPUs process tasks in groups, leading to a sudden surge of demand. Unlike traditional industrial loads, which tend to be relatively stable and predictable, AI-enabled data centers can exhibit sharp, short-term ramping behavior that varies by the hour or even minute. These fluctuations are often tied to dynamic workload scheduling, where computing tasks are shifted in real time based on performance needs or energy prices. This behavior is typically observed due to the operational nature of AI data centers. AI data centers generally engage in training to develop and refine models, and an interface state comprising two phases: the prompt generation phase, which analyzes the user's prompt, and the token generation phase, which produces the output for the user. [39] Currently, the training state results in a significantly higher surge in demand compared to the inference state. However, the training state can be scheduled, allowing for the possibility of conducting training sessions during non-critical periods in the future [39]. As highlighted in an Institute of Electrical and Electronics Engineers (IEEE) Spectrum report [40] on the DCFlex Initiative [41], companies like Google and Nvidia are now experimenting with “workload choreography” to align data center operations with grid conditions, but this remains an emerging practice rather than the norm. Without visibility into these rapid changes, grid operators struggle to maintain voltage and frequency stability, especially in regions with high concentrations of data centers.

The National Renewable Energy Laboratory’s (NREL) Chip-to-Grid (C2G) [42] initiative further underscores the complexity of managing these loads. The initiative notes that the lack of transparency around real-time data center operations, combined with the high power densities and cooling demands of AI infrastructure, makes it difficult to forecast and respond to load ramps effectively. Traditional resource adequacy studies, which rely on average or peak demand assumptions, are ill-equipped to capture the volatility introduced by AI-driven compute cycles. This mismatch can lead to under-provisioning of fast-ramping generation or grid services, increasing the risk of reliability events. To address this, grid operators are beginning to explore predictive models and demand response strategies tailored to the unique behavior of data centers, but these tools are still in early stages of development and deployment.

Despite attempts to mitigate these issues, the majority of large loads currently do not participate in their utility’s Security Constrained Economic Dispatch (SCED) which would penalize fast ramping. ERCOT implemented a SCED, but 255 five-minute SCED intervals were documented over a 1-year interval where the change in large load consumption exceeded the total procured regulation across ERCOT for that interval [34].



Source: Wood Mackenzie PowerRT Sensor Data

Figure 3: Real-time power consumption from two US hyperscale data centers (MW) [43]

This illustrates a need for reconsideration of ancillary service markets, who participates in them voluntarily or mandatorily, and the capacity of such markets or reserves.

3.2.2.1. CASE STUDY: SYNCHRONIZED MACHINE LEARNING WORKLOADS

Data from Google recently highlighted the high degree of variability created by machine learning (NL) and AI workloads. In a presentation at the 2025 Open Compute Project summit for Europe, the Middle East and Africa, a Google researcher highlighted the difference in computing power from non-AI workloads and AI workloads as seen in Figure 4. Google highlights that today’s ML workloads use synchronized computation across tens of thousands of accelerator chips [44]. AI workloads often occupy an entire data center cluster, or even multiple clusters. The peak utilization of ML workloads can approach the rated power of the underlying equipment, and if a cluster’s power usage is dominated by a few large ML workloads, the power fluctuation between idle and peak utilization rises and falls much more steeply. Depending on the workload's characteristics, these inter- and intra-job power fluctuations can occur very frequently.

Figure 4: Computation power comparison between non-AI and AI workloads [45]. These plots, showing power consumption over time on an equal scale of magnitude, suggest a 15x difference in a peak-to-trough transition over a period of milliseconds.

3.2.3. SOLUTIONS IN NORMAL OPERATIONS FOR LARGE LOADS

To manage the operational complexity introduced by large loads, grid operators need a new generation of tools, protocols, and planning frameworks that reflect the scale, speed, and inflexibility of hyperscale demand. Traditional assumptions about load diversity, predictability, and controllability no longer hold in regions with dense data center development. As a result, operators must shift from reactive balancing to proactive coordination—starting well before a load is energized.

One critical solution is the integration of operational requirements into the interconnection and planning process. Operators should be involved early in the review of interconnection requests to ensure that ramp rates, telemetry standards, and curtailment capabilities are clearly defined and enforceable. This includes requiring real-time visibility into load behavior and behind-the-meter resources, which are often opaque under current agreements. Enhanced telemetry and forecasting data sharing between load operators and grid control centers can help mitigate the uncertainty introduced by AI-driven or batch-processing workloads.

Finally, operational flexibility must be built into the grid architecture itself. This includes investing in fast-ramping generation, grid-scale storage, and flexible demand resources that can buffer the system against sudden load changes. In some cases, it may also involve requiring large loads to install on-site equipment capable of participating in grid services—such as frequency response or emergency curtailment. These measures not only improve reliability but also ensure that the costs and responsibilities of grid stability are shared more equitably between utilities and large-load customers.

3.3. RELIABILITY RISK AND VOLTAGE RIDE THROUGH

The reliability of large loads and their ability to ride through voltage disturbances are critical concerns for modern power systems. As data centers, crypto mining operations, and other high-density computational facilities become more prevalent, understanding and managing their impact on the grid becomes increasingly important. This section delves into the complexities and challenges associated with large load reliability, voltage ride-through, and the phenomena of tripping and emergent behavior.

3.3.1. TRIPPING AND EMERGENT BEHAVIOR

Tripping refers to the automatic disconnection of a load or generator from the power grid in response to certain conditions, such as voltage disturbances or equipment failures. This is a protective mechanism designed to prevent damage to equipment and ensure safety. However, tripping can also lead to sudden shifts in power demand or supply, which may destabilize the grid.

Emergent behavior, in the context of power systems, describes the unexpected and complex interactions that arise when multiple elements of the grid, such as loads and generation sources, respond to disturbances. Emergent behavior can manifest as oscillations, voltage flickering, or other forms of instability that are not easily predicted by traditional analysis methods. Understanding and mitigating these behaviors are crucial for maintaining grid stability, especially as the penetration of non-linear loads and advanced power electronics increases.

Voltage ride-through has been identified as a high-priority issue that requires resolution [14]. ERCOT has observed that several new types of loads, including variable frequency drives and the power electronic drives needed for data centers and crypto mining, are particularly sensitive to voltage disturbances [28]. Electric Power Research Institute (EPRI) has raised concerns that the entire load may be disconnected due to terminal voltage exceeding acceptable limits [9]. Addressing this risk is challenging because historically, some load reduction or tripping during a fault/low voltage condition has been beneficial for the system. For instance, a sudden loss of approximately 1.6 GW of load led to an increase in frequency to 60.235 Hz as illustrated in Figure 3 [30]. An example from American Electric Power showcases the voltage ride-through capabilities and reclosing. Two different data centers, located in separate locations, were both able to ride through a fault without losing the load. This was accomplished using high-speed reclosing (HSR), which attempts to reconnect to a fault in approximately 80 milliseconds, with about 350 milliseconds between attempts. It took three attempts for the data centers to reconnect [31].

Despite successful examples of data center ride throughs, as the proportion of voltage-sensitive loads increases and system strength decreases, the risk of substantial load loss during a voltage disturbance also increases. In a July 2024 event identified via regulatory filings [32], a failed surge protector created power quality issues in Data Center Alley, a region outside of Washington DC that is reported serve approximately 70% of internet traffic. In response to the power quality issues, 60 data centers, out of more than 200 in this region, disconnected suddenly and switched to local backup generation, leading to a loss of 1500 MW of load. Grid operator and local utility Dominion Energy had to rapidly scale back generation to account for the sudden loss. This event as detailed in the sub section to follow, was not unique. A review of ERCOT filings revealed more than 30 near-miss incidents since 2020 triggered by large loads disconnecting. One of these included a loss of nearly 400 crypto miners, data centers, and oil and gas facilities disconnecting without warning, causing 112 MW of generation to shut down [32].

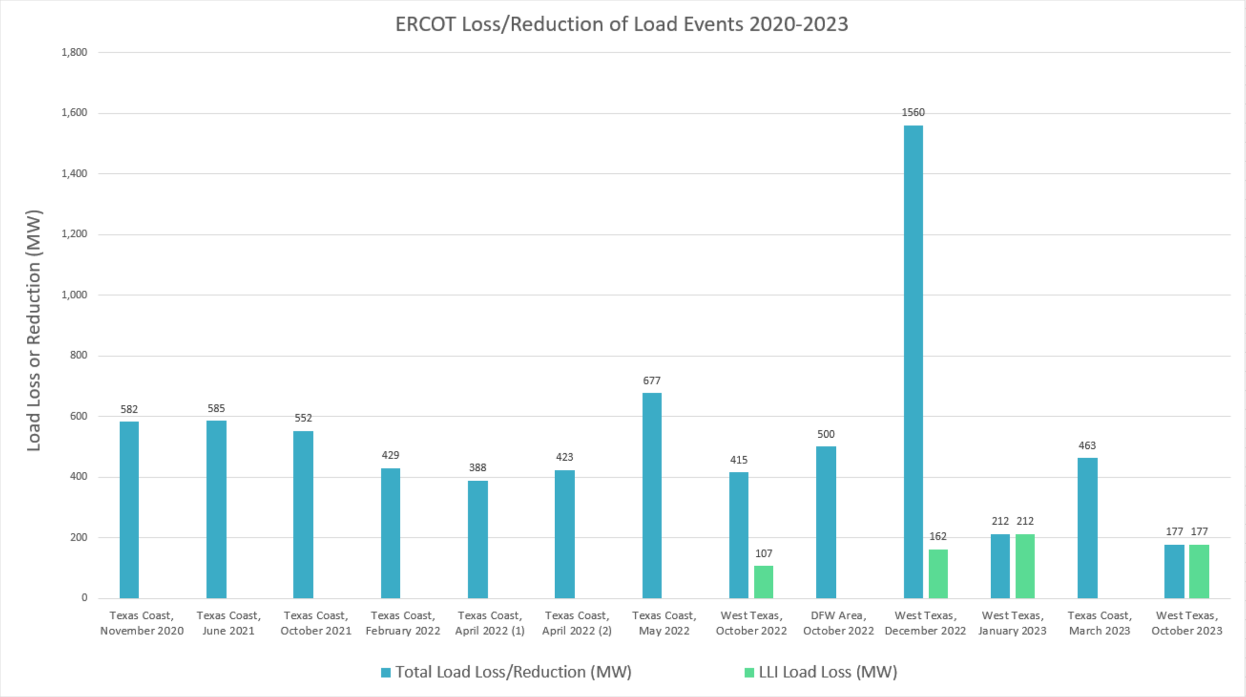


Figure 5: Large Load Loss and Large Load Interconnection (LLI) Events at ERCOT [46]

3.3.1.1. CASE STUDY: INSTABILITY IN DATA CENTER ALLEY

In July 2024, Dominion Energy experienced voltage and frequency overshoot on a 230 kV line due to the loss of a large load (approximately 1.5 GW) as illustrated in Figure 6 [47]. A lightning strike led to the failure of a lightning arrester during a thunderstorm, causing significant power quality issues on the grid. This event triggered the protective mechanisms in place at numerous data centers, resulting in their automatic disconnection to prevent damage from the compromised power quality. Approximately 1.5 GW of load was suddenly lost, leading to a frequency overshoot of 60.047 Hz as shown in Figure 7. The current practice involves manually restoring the connection of the UPS-backed large load to the grid after the grid-side disturbance is mitigated [15].

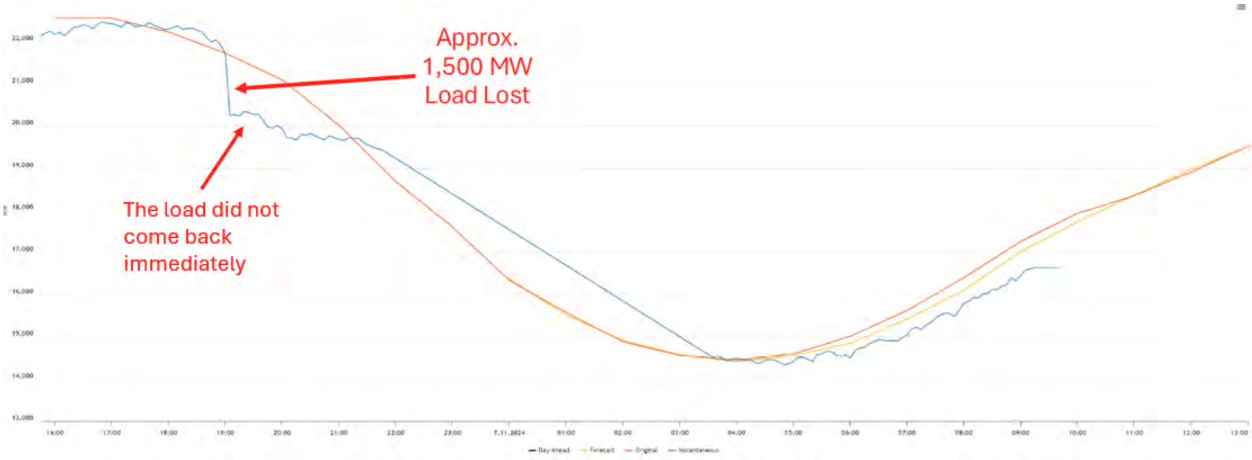


Figure 6: Example of large load loss in Dominion Energy territory. [47]

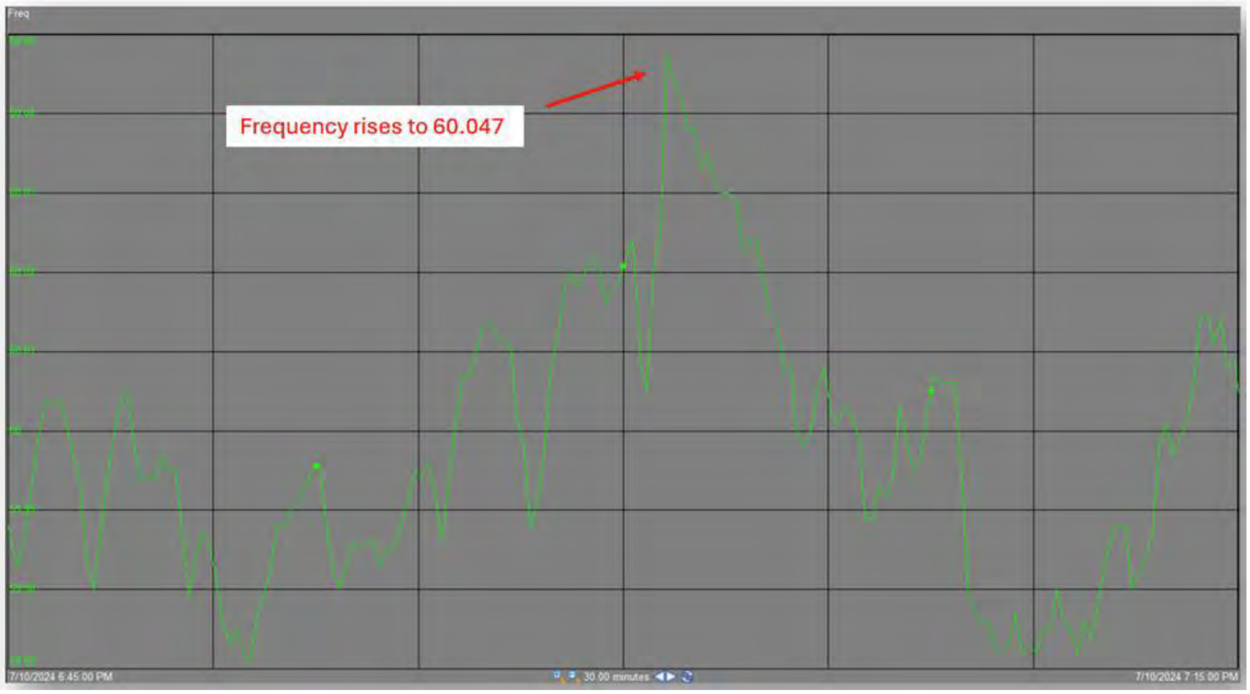


Figure 7: Frequency at time of large load loss in the same event [47].

3.3.2. HARMONICS AND OSCILLATIONS CAUSED BY DATA CENTERS

Modern data centers, despite using advanced power supplies and energy-efficient equipment, can still experience significant harmonic distortion due to fluctuating loads and non-linear devices. One key issue for grid operators due to the high penetration of power electronics and non-linear loads in data centers is oscillations in power systems caused by harmonics, distorted waveforms that arise when non-linear loads (like servers, variable frequency drives, and UPS systems) inject currents at frequencies other than the fundamental 60 Hz [48]. These oscillations can lead to a range of operational problems, including voltage distortion, overheating of transformers and conductors, and interference with sensitive equipment. In particular, even facilities equipped with power factor-corrected supplies and newer UPS technologies are not immune to these effects. Oscillations may become more pronounced during rapid load changes, such as when AI workloads ramp up or shift dynamically, causing harmonic currents to fluctuate in ways that traditional mitigation strategies may not fully address. Harmonic-mitigating measures, such as specially designed UPS systems and transformers, should be deployed to dampen these oscillations and maintain power quality in high-density computing environments

A Bloomberg report using data from Whisker Labs and DC Byte [49] found that the increase in data centers is putting strain on the grid and impacting the quality of power delivered to other customers. The report's analysis showed that more than three-quarters of highly distorted power readings in the U.S. are within 50 miles of large data center activity. New facilities adjacent to large US cities have compounded this, further stressing the power grid. More than half of the tracked households that showed the worst distortions of power quality are located within 20 miles of significant data center activity, which according to U.S. census figures, affects 3.7 million people that live in these impacted areas.

Quanta shared a complementary observation in a presentation to the NERC Large Load Task Force in November of 2024 [2], stating that “nonlinear loads bring a large number of harmonics to the grid, affecting power quality, causing energy losses and interferences to grid-connected apparatus, even resonances in the microgrid.”

3.3.2.1. CASE STUDY: VOLTAGE FLICKERING

In June 2022, Dominion Energy observed voltage flickering at a 115 kV substation serving the data center loads of Substations D1, D2, and D3, as illustrated in Figure 4 [50]. The voltage flickering commenced at 2:05 A.M. when the four hydro units at Station K began ramping down [50]. The issue was resolved at 3:58 A.M. when the STATCOM (Static Synchronous Compensator) gradually adjusted from 0 to -71 MVAR [50]. The voltage flickering was not caused by the initiating event, but rather by the data centers as they responded to the dynamic conditions on the grid. This instability, without mitigation as provided by the STATCOM, could exacerbate dynamic instability conditions on the grid, and even potentially lead to data center loads disconnecting due to the instability. While the event studied in this paper, as shown in Figure 5, was a 14.7 Hz oscillation event, flickering has also been observed during 10-11 Hz oscillation events [50].

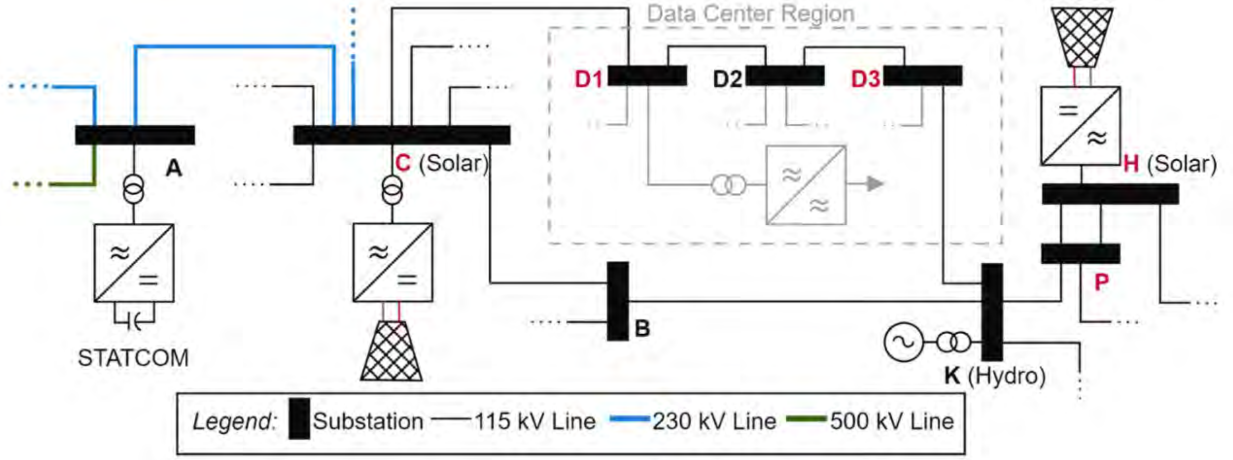


Figure 8: Region of Dominion Energy's power network highlighting STATCOM mitigation [50]

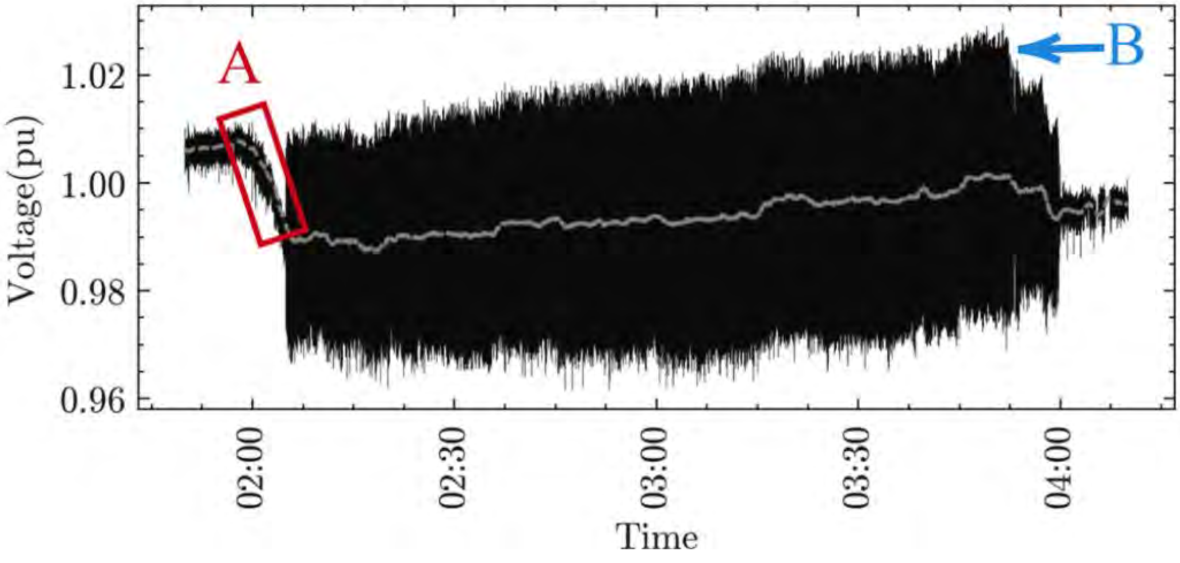


Figure 9: Voltage Magnitude at Substation D3. Label A indicates when four hydro units were ramped down preceding the oscillation. Label B indicates when the STATCOM setpoint was changed from 0 to 71 MVAR. [50]

3.3.3. SOLUTIONS TO ABSORB OR OPERATE THROUGH LARGE LOAD LOSS OR POWER QUALITY EVENTS

One method to partially mitigate the risk of large load loss due to dynamic grid instability events is to require large loads to ride through certain fault conditions, similar to inverter-based resources (IBRs). When IBR penetration was low, interoperability standards specified that IBRs should trip off during grid disturbances. However, as IBR penetration rose significantly, it became clear that the sudden loss of these resources during a voltage or frequency instability event could further exacerbate the issue. IEEE 1547-2018 updated the interoperability requirements to specify ride though curves to ensure IBRs would remain grid-connected during minor instability events [51]. A similar ride through requirement could mitigate risks of multiple data center trips in a single region. This approach comes with technical and operational challenges, as outlined in the IEEE standard 1668 Single-Phase and Phase-Phase Curve [15] [46]. Data centers are also resistant to this approach due to the sensitive nature of the power electronics and cooling equipment. Another approach is to establish new planning criteria to strengthen the grid, thereby reducing the size and area of voltage sag during a fault and, consequently, reducing load tripping. This approach necessitates the development of better dynamic models of large loads [15] [34].

Underfrequency Load Shedding (UFLS) is a critical measure designed to maintain the balance between generation and load when an imbalance causes the grid frequency to drop rapidly [39]. These programs establish thresholds that the frequency must not fall below, and if it does, some load will be automatically shed to prevent further frequency decline. In the United States, this threshold typically ranges from 59.5 to 59.3 Hertz, although it can vary by region. Standard regulations for UFLS have been established by the North American Electric Reliability Corporation (NERC) in the NERC Reliability Standard PRC-006-1 [52]. While data centers may be completely shut off in extreme cases, NERC PRC-006-1 aims to minimize the amount of load shedding required. Furthermore, data centers should consider adding backup capacity for their non-computational loads, such as cooling systems, to reduce the need for load shedding.

Additionally, non-computational loads can adversely affect power quality. Cooling motors, for instance, can stall, leading to a depressed voltage that is slow to recover after a fault, a phenomenon known as Fault-Induced Delayed Voltage Recovery (FIDVR) [53] [54]. One major cause of this issue is the excessive draw of reactive power by cooling motors, which can be up to five to six times their steady-state current. Given the significant impact that cooling and other non-computational loads have on the grid, data centers should provide utilities with detailed information on the percentage of computational loads. Data centers should strive to reduce their non-computational loads to achieve a higher Power Usage Effectiveness (PUE). One effective method is to use water-cooled server racks and employ evaporative cooling to dissipate heat instead of relying on chillers.

3.4. RESILIENCE AND LARGE LOAD OPERATIONS

Grid operators' planning and execution of resilience measures will increasingly need be aligned with large load operations. This is particularly important while serving high utilization and high uptime incentivized data centers. Large load operational factors that grid operators need to account into their resilience measure are described below.

3.4.1. RESTORATION AND PRIORITIZATION

Grid operators increasingly must account for large load characteristics, particularly those of data centers, into their restoration strategies following outages. would need to account for data centers' post-outage power ramping needs into their restoration planning and serving load prioritization. Unlike traditional loads, data centers often have tiered reliability requirements (e.g., Tier III or IV), which dictate stringent uptime expectations and complex restart behaviors. Restoration planning must account for the non-linear ramping of power demand post-outage, especially when multiple systems (cooling, IT, storage) come online in sequence or simultaneously.

Prioritization of which loads to reconnect to the grid first must consider not only the uptime and performance of the goals but also the increasing operational dependency that various critical infrastructure sectors, including the grid itself, have on AI-driven or cloud-tool systems and their supporting infrastructure. Traditional "critical load" are grounded in ensuring the continued continuity of service to critical infrastructure facilities, such as hospitals, defense, emergency response, or water systems, systems which directly impact life, safety, or public order. As these national critical functions (NCFs) increasingly rely on digitization, cloud services, and AI tools to drive modernization and efficiencies, the data centers hosting these tools become enabling infrastructure for the NCFs and must be prioritized appropriately. This means that grid operators must take into account not only what class of large load a customer is (e.g. industrial vs. agricultural vs. data center), but also must know and leverage information about that large load's function to prioritize service. Further complicating prioritization, data centers may serve multiple functions and may require a minimum level of service to host their most critical computing loads while others can be deprioritized or shifted to other facilities, options which must be communicated and coordinated with grid operators to enable maximum efficiency. Additionally, most data centers have on-site power backup, but to minimize any disruptions, grid operators must have knowledge of the duration of backup capacity available before grid service must be restored in order to factor that into their prioritization plans.

3.4.2. SOLUTIONS AND OPPORTUNITIES FOR RESILIENT LARGE LOAD PREPARATIONS

Enhancing the resilience of large loads requires a combination of technical upgrades, operational flexibility, and collaborative planning. site generation, whether diesel, natural gas, or battery storage, remains a cornerstone of resilience for data centers, but its effectiveness depends on fuel availability, maintenance practices, and integration with facility controls. Long duration energy storage also has the potential to serve data center during grid outage, especially when paired with intelligent energy management systems that can optimize discharge timing during grid outages. Long duration energy storage can be recharged using data centers' backup generation and/or when the grid is restored. Beyond hardware, flexible operation of data centers, such as deferring non-critical workloads or shifting compute tasks to other regions, can reduce stress on the grid and accelerate restoration timelines. Grid operators should collaborate with data centers for criticality-informed load segmentation that can not only enhance large load operational resilience but also support the grid operator in quicker and better-informed restoration of critical loads. Co-developed resilience plans that define critical load tiers, restoration priorities, and communication protocols will help with recovering quickly and supporting the broader grid ecosystem during times of stress.

3.5. EMERGING KEY ISSUES IN OPERATIONS

3.5.1. BLACKSTART

Blackstart planning traditionally focuses on restoring generation and transmission assets, but the growing presence of large, distributed loads like data centers introduces new complexities. In traditional blackstart, a bottom-up approach leverages blackstart generators to create small electrical islands and then connect those islands to reestablish the grid. The load restoration process in blackstart is critical to ensure that load and generation are always matched. The grid must be stable in voltage and frequency before reconnecting large loads, but load criticality is also taken into account.

Data centers are not straightforward assets to prioritize on a load restoration queue for a blackstart plan. As discussed previously, there is an argument that their criticality levels should vary depending on the computing loads that are supported at that facility. However, even critical loads with sufficient backup capacity may be move lower on the load restoration sequencing [53]. Their sensitivity to voltage and frequency stability means that a strong grid should be established before reconnecting to the grid to avoid the data center switching back to backup power due to power quality issues and disrupting overall blackstart processes by interrupting the delicate process of matching loads to generation.

The presence of backup capacity suggests a possibility of using BTM resources to aid in the blackstart restoration process by forming a local island. However, the capacity of those generation resources may not be sufficient to power much beyond the data center, even if the data center is not operating at full load. The substation may not be designed to feed generation back into the grid, and the on-site generation resources may not have the inertia to support blackstart operations, even if a grid-forming capability exists. An additional challenge raised by FERC is how co-located resources would be compensated for their contributions to blackstart if that is a service offered [55].

It is essential for grid operators to understand the nature of the data processed by data centers, whether it be cloud services, artificial intelligence computations, or crypto mining operations [53]. Conducting real-time drills can be beneficial for all parties involved, ensuring a clear understanding of which critical and non-critical loads need to be managed [53]. These drills are instrumental in both preventing blackouts and determining prioritization during a reconnection process.

Furthermore, data centers should strive to maintain open communication with grid operators, providing detailed information about their load profiles and the criticality of their operations [53]. A standard of communication between data center and grid operator should be established. This practice not only aids in the efficient management of power restoration efforts but also enhances overall grid resilience.

3.5.2. VISIBILITY AND TIMEFRAMES FOR SENSING AND MEASUREMENT

Data centers are customers, and like most utility customers, grid operators rely on utility-grade metering infrastructure that may only provide course-grained visibility. Traditional meters only collect data at 15-mniute or hourly intervals. However, the operational dynamics of large loads, especially those with high-performance computing or AI workloads, can shift significantly within seconds or minutes. This mismatch in temporal resolution can hinder real-time situational awareness and limit the operator’s ability to respond to fast-ramping events or unexpected load drops. Furthermore, the lack of sub-metering or telemetry from internal systems (e.g., cooling, storage, compute clusters) prevents operators from distinguishing between critical and deferrable loads. To address this, grid operators may need to explore advanced metering infrastructure, enhanced telemetry agreements, secure data-sharing protocols, and edge-based sensing architectures that allow for more granular, near-real-time monitoring. These improvements are essential not only for maintaining grid reliability but also for enabling advanced services such as dynamic load shaping, predictive dispatch, and event-driven demand response.

3.5.3. OWNERSHIP BY 3RD PARTIES

The operational complexity of large load integration is further compounded by the diverse ownership and development models of data center campuses. Unlike traditional industrial facilities, data centers are often developed through multi-entity arrangements involving real estate investment trusts, infrastructure developers, utility affiliates, and hyperscale cloud providers. In some cases, the land, electrical infrastructure, and IT equipment may be owned or financed by entirely separate entities, each with different operational priorities and risk tolerances. This fragmentation can create ambiguity in accountability during grid events, particularly when it comes to restoration coordination, load shedding participation, or telemetry sharing. For example, a utility may need to coordinate with a real estate developer for access to substation infrastructure, while also negotiating with a cloud provider for visibility into load behavior. These layered ownership structures can delay response times, complicate interconnection agreements, and hinder the implementation of resilience measures. To address these challenges, grid operators may need to establish multi-party coordination frameworks and standardized operational agreements that clearly define roles, responsibilities, and communication protocols across all entities involved in the data center ecosystem.

The operational complexity of data centers is amplified by their multi-tenant structure, where a single facility may host dozens or even hundreds of customers with varying SLAs, compute priorities, and energy usage patterns. From a grid operator's perspective, this fragmentation obscures the true nature of the load and complicates efforts to implement coordinated demand-side management or restoration prioritization. For example, one tenant may be running latency-sensitive financial transactions while another is training a non-critical AI model—yet both appear as a single aggregated load. Additionally, third-party ownership can introduce contractual and legal barriers to data sharing, making it difficult for utilities to obtain the operational transparency needed for grid planning and emergency response. To mitigate these challenges, regulators and utilities may need to develop standardized frameworks for load disclosure, criticality tagging, and grid-facing operational interfaces that allow colocation facilities to participate in grid coordination without compromising tenant confidentiality.

3.5.4. NATIONAL LOAD SHIFTING CAPABILITY

One of the most disruptive characteristics and potential opportunities of modern data center operations is their ability to shift computational workloads across geographic regions in response to cost, latency, or other signals. This represents a paradigm shift for grid operators, who rely on stable, regionally anchored load forecasts, especially if this shift is driven by internal orchestration that does not provide notification to grid operators. However, this flexibility also has the potential to support grid operator goals, such as shifting compute loads nationally to perform peak shaving, relieve congestion, or remove load from stressed regions. To manage this, there is a growing need for inter-regional coordination mechanisms, real-time load migration visibility, and policy frameworks that recognize the national scope of large load behavior. Grid operators may also benefit from predictive analytics partnerships with hyperscale operators to anticipate load shifts and proactively adjust dispatch and reserve strategies.

4. CONCLUSION

The rapid deployment of large loads, particularly data centers, is reshaping the landscape of electric grid operations in the United States. These modern loads, characterized by their immense size, high density, and unique operational requirements, present unique challenges for grid operators. Unlike traditional industrial loads, data centers demand continuous, high-quality power and exhibit load profiles that can fluctuate dramatically, complicating both short-term operations and long-term planning. The integration of large loads, particularly data centers, into the grid poses challenges across daily operations, reliability planning, resilience scenarios. Even traditional grid planning considerations, like interconnection, permitting, and resource adequacy, are changing grid operations with the consideration of data center characteristics. considerations. While many challenges have been identified by industry, there are many areas for which further investigation is needed. These areas are outlined in Table 2

Table 2: Areas requiring further investigation to find solutions for grid operators related to data center deployment.

Topic	Areas for Further Investigation
Planning and Cost	Policy, regulation, and impact studies must be streamlined to reduce interconnection wait times
	There is no consensus on an adequate method for fairly compensating utilities and other customers for additional costs incurred by connecting a data center to the grid.
	A lack of accurate composite load models prevents sufficient planning and experimentation with data, which slows down the interconnection approval process.
	Data centers are increasingly bypassing the typical permitting process by connecting directly to generation sources, a practice which faces technical and regulatory hurdles.

Topic	**Areas for Further Investigation**
Normal Operations	A data center can ramp up or ramp down its demand in milliseconds, leading to significant balancing challenges for utilities.
	Previous studies have analyzed the impact of GPUs on data centers at the kilowatt level, revealing voltage pulsations; however, it remains unclear if these voltage pulsations result in load ramps at the megawatt level
	The diversity of compute workloads (e.g., AI inference, batch processing) introduces unpredictable and highly variable demand profiles.
	Lack of real-time telemetry from behind-the-meter data centers limits grid operators' ability to forecast and respond to operational changes.
Reliability Risk	Data centers are very sensitive to voltage fluctuations and can easily disconnect from the grid. There is a stronger need for data centers to "ride through" voltage fluctuations.
	Voltage flickering events are a concern for data centers and collaborative solutions with utilities are needed to balance needs of both entities.
	Non-computational loads, such as cooling, can also create stability challenges for the grid, including excessive draw of reactive power.
Resilience Considerations	Restoration planning must account for the ramping behavior of large loads and the prioritization of segmented critical services within data centers.
	Blackstart plans that leverage data centers as blackstart resources and not only large loads are faced by technical limitations, but could provide an opportunity for novel ways to leverage backup power capabilities.
	Prioritization of data center loads in restoration or blackstart is not straightforward, and must account for the size, criticality, flexibility, and backup capacity of the asset.

To address these challenges, grid operators must adopt a proactive and coordinated approach, involving early and ongoing collaboration with data center developers, modernizing cost allocation frameworks, and streamlining permitting processes. Investing in fast-ramping generation, grid-scale storage, and flexible demand resources will be crucial to buffer the system against sudden load changes and ensure grid stability. The successful integration of data center large loads into the grid will require a concerted effort from all stakeholders, including utilities, regulators, and data center operators. By embracing innovative solutions and fostering collaboration, the industry can navigate the complexities of this new era and ensure a reliable and resilient electric grid for the future.

5. REFERENCES

[1] TheWhiteHouse, "Accelerating Federal Permitting of Data Center Infrastructure," The White House, 23 July 2025. [Online]. Available: https://www.whitehouse.gov/presidential-actions/2025/07/accelerating-federal-permitting-of-data-center-infrastructure/. [Accessed 01 September 2025].

[2] P. Shah, H. Sharma, K. Thomas, M. Henson and E. Mickelson, "LLTF Presentations," 14 November 2024. [Online]. Available: https://www.nerc.com/comm/RSTC/LLTF/LLTF_Presentaions_November_14_2024.pdf.

[3] A. Shehabi, S. J. Smith, A. Hubbard, A. Newkirk, N. Lei, M. A. B. Siddik, B. Holecek, J. Koomey, E. Masanet and D. Sartor, "United States Data Center Energy Usage Report," Lawrence Berkeley National Laboratory, Berkeley, California, 2024.

[4] National Electric Reliability Corporation (NERC), "2023-Long Term Reliability Assessment," NERC, 2023.

[5] Crusoe, "Crusoe expands AI data center campus in Abilene to 1.2 gigawatts," Crusoe, 18 March 2025. [Online]. Available: https://www.crusoe.ai/resources/newsroom/crusoe-expands-ai-data-center-campus-in-abilene-to-1-2-gigawatts. [Accessed 7 September 2025].

[6] C. Kunkel, "Projected data center growth spurs PJM capacity prices by factor of 10," Institute for Energy Economics and Financial Analysis, 30 July 2025. [Online]. Available: https://ieefa.org/resources/projected-data-center-growth-spurs-pjm-capacity-prices-factor-10. [Accessed 7 September 2025].

[7] C. Waldoch and T. Ennis, "Capacity Constraints in PJM have Grown Beyond Dominion," GridStatus, 25 July 2025. [Online]. Available: https://blog.gridstatus.io/pjm-ppl-capacity-constraints-datacenters/. [Accessed 7 September 2025].

[8] S. King, "Texas Senate Bill 6, Legislative Session 89(R)," Texas Legislature Online, 20 June 2025. [Online]. Available: https://capitol.texas.gov/BillLookup/History.aspx?LegSess=89R&Bill=SB6. [Accessed 03 September 2025].

[9] NERC, "Large Loads Frequently Asked Questions," NERC, 2025.

[10] *Texas Senate Bill 6, 89th Legislature Regular Session,* 2025.

[11] M. Christian, "Co-ops Discuss Risks, Opportunities of Demand Surge From AI, Data Centers," NRECA, 11 June 2024. [Online]. Available: https://www.electric.coop/co-ops-discuss-risks-opportunities-of-demand-surge-from-ai-data-centers. [Accessed 7 September 2025].

[12] H. Agrawal, "Data Centers Want Power, Regulators Say Wait," Yale Campus Press, 4 August 2025. [Online]. Available: https://campuspress.yale.edu/newgen/data-centers-want-power-regulators-say-wait/.

[13] Federal Energy Regulatory Commission (FERC), "FERC Orders Action on Co-location Issues Related to Data Centers Running AI," FERC, 20 February 2025. [Online]. Available: https://www.ferc.gov/news-events/news/ferc-orders-action-co-location-issues-related-data-centers-running-ai.

[14] D. E. Mills, "PJM Board Letter Re: Implementation of Critical Issue Fast Path Process for Large Load Additions," 8 August 2025. [Online]. Available: https://www.pjm.com//-/media/DotCom/about-pjm/who-we-are/public-disclosures/2025/20250808-pjm-board-letter-re-implementation-of-critical-issue-fast-path-process-for-large-load-additions.pdf.

[15] North American Electric Reliability Corporation, "Large Load Task Force Kickoff Presentations," North American Electric Reliability Corporation (NERC), 2024.

[16] Federal Energy Regulatory Commission (FERC), "Explainer on the Interconnection Final Rule," FERC, 23 January 2025. [Online]. Available: https://www.ferc.gov/explainer-interconnection-final-rule. [Accessed 7 September 2025].

[17] Joint Legislative Audit and Review Commission, "Data Centers in Virginia: JLARC Report 598," 9 December 2024. [Online]. Available: https://jlarc.virginia.gov/pdfs/reports/Rpt598.pdf.

[18] A. Brown, "Lawmakers Fear AI Data Centers Will Drive Up Residents' Power Bills," National Conference of State Legislatures, 22 April 2025. [Online]. Available: https://www.ncsl.org/events/details/lawmakers-fear-ai-data-centers-will-drive-up-residents-power-bills.

[19] E. Foxhill and K. Guo, "Texas electricity demand could nearly double in six years, grid operator predicts," The Texas Tribune, 20 June 2024. [Online]. Available: https://www.texastribune.org/2024/06/20/texas-electricity-demand-forecast-ercot/.

[20] North American Electric Reliability Corporation, "NERC Large Loads Task Force NERC LLTF Meeting and Workshop," 10 April 2025. [Online]. Available: https://www.nerc.com/comm/RSTC/LLTF/LLTF_April_Meeting_&_Technical_Workshop_Presentations_.pdf.

[21] Milrose Consultants, "Zoning, Permits, and Compliance: Breaking Down the Data Center Approval Process," Milrose Consultants, 3 March 2025. [Online]. Available: https://www.milrose.com/insights/data-center-approvals.

[22] K. Thomas and N. Giffin, "Data Center Modeling," Elevate Energy Consulting, 8 July 2025. [Online]. Available: https://www.scribd.com/document/897334117/Data-Center-Modeling.

[23] S. Jenkins, "Datacenter Load Modeling," January 2024. [Online]. Available: https://www.wecc.org/sites/default/files/documents/meeting/2024/12%20-%20JenkinsS%20-%20Data%20Center%20Load%20Modeling_January%202024.pdf.

[24] B. A. Moore, "Large Load Interconnection: Some Legal and Policy Issues," Butler Snow, L. L. P., 13 February 2025. [Online]. Available: https://www.butlersnow.com/news-and-events/large-load-interconnection-some-legal-and-policy-issues1. [Accessed 16 July 2025].

[25] Data Center Watch, "$64 billion of data center projects have been blocked or delayed amid local opposition," Data Center Watch, 2025. [Online]. Available: https://www.datacenterwatch.org/report.

[26] B. Giacobone, "Is the Texas Grid Overestimating Load Growth?," Ascend Analytics, May 2025 2025. [Online]. Available: https://www.ascendanalytics.com/blog/is-the-texas-grid-overestimating-load-growth. [Accessed 28].

[27] Microsoft Corporation, "Microsoft Comments on Georgia Power's 2023 Integrated Resource Plan Update: Georgia Public Service Commission Document 218199," 1 April 2024. [Online]. Available: https://psc.ga.gov/search/facts-document/?documentId=218199.

[28] National Association of Regulatory Utility Commssioners (NARUC), "NARUC Staff Surge Call – Forecasting Load Growth: Assumptions and Risks, February 10, 2025," 10 February 2025. [Online]. Available: https://pubs.naruc.org/pub/EBDE4EE2-CE58-1BDD-47C4-7990D22A7A97.

[29] Black&Veatch, "Data Centers Add Twist to Complex Challenge of Load Forecasting for U.S. Power Utilities," Black & Veatch Corporation, 2024. [Online]. Available: https://new.bv.com/en-US/perspectives/data-centers-add-twist-to-complex-challenge-of-load-forecasting-for-u-s-power-utilities. [Accessed 01 September 2025].

[30] National Academies, "Report on Electric Grid Resilience," 2023. [Online]. Available: https://www.nationalacademies.org/documents/embed/link/LF2255DA3DD1C41C0A42D3BEF0

989ACAECE3053A6A9B/file/DE3FB82D37402D695C97156F50D4AF5476F5DF930A6D?noSaveAs=1.

[31] Goldman Sachs, "AI to drive 165% increase in data center power demand by 2030," Goldman Sachs, 4 February 2025. [Online]. Available: https://www.goldmansachs.com/insights/articles/ai-to-drive-165-increase-in-data-center-power-demand-by-2030. [Accessed 5 September 2025].

[32] CloudExpat, "Cloud AI Platforms Comparison: AWS Trainium vs Google TPU v5e vs Azure ND H100," CloudExpat, 27 March 2025. [Online]. Available: https://www.cloudexpat.com/blog/comparison-aws-trainium-google-tpu-v5e-azure-nd-h100-nvidia/. [Accessed 5 September 2025].

[33] J. O'Donnell and C. Crownheart, "We Did the Math on AI's Energy Footprint: Here's the Story You Haven't Heard," MIT Technology Review, 20 May 2025. [Online]. Available: https://www.technologyreview.com/2025/05/20/1116327/ai-energy-usage-climate-footprint-big-tech/. [Accessed 5 September 2025].

[34] P. Gravois, "ERCOT Large Load Loss/Reduction Events 2020-2024," 19 November 2024. [Online]. Available: https://www.ercot.com/files/docs/2024/11/18/ERCOT%20Large%20Load%20Events_PDCWG_19Nov2024.pdf.

[35] B. Martucci, "Texas law gives grid operator power to disconnect data centers during crisis," Utility Dive, 25 June 2025. [Online]. Available: https://www.utilitydive.com/news/texas-law-gives-grid-operator-power-to-disconnect-data-centers-during-crisi/751587/. [Accessed 01 September 2025].

[36] T. Horger, "Large Load Additions Workshop," 9 May 2025. [Online]. Available: https://www.pjm.com/-/media/DotCom/committees-groups/workshops/llaw/2025/20250509/20250509-item-02---large-load-additions-workshop---presentation.pdf. [Accessed 7 September 2025].

[37] S. Bresler and T. Horger, *Large Load Additions, PJM Conceptual Proposal and Request for Member Feedback,* Valley Forge: PJM, 2025.

[38] Elevate Energy Consulting, "Practical Guidance and Considerations for Large Load Interconnections," 2025.

[39] P. Shah and H. Sharma, *Interconnection and Integration of Large Loads,* 2024.

[40] J. Tilton, "Big Tech Tests Data Center Flexibility Google, Nvidia, Oracle, and Data4 adapt three data center hubs to local power grids," *IEEE Spectrum,* 12 June 2025.

[41] Electric Power Research Institute, *DC Flex,* 2025.

[42] "Chip to Grid Data Center Initiative," 2025. [Online]. Available: https://www.nrel.gov/computational-science/research/chip-to-grid-data-center-initiative. [Accessed 7 September 2025].

[43] C. Seiple and B. Hertz-Shargel, "US power struggle," Wood MacKenzie, June 2025. [Online]. Available: https://www.woodmac.com/horizons/us-data-centre-power-demand-challenges-electricity-market-model/. [Accessed August 2025].

[44] H. Gan and P. Ranganathan, "Balance of power: A full-stack approach to power and thermal fluctuations in ML infrastructure," Google Cloud, 11 February 2025. [Online]. Available: https://cloud.google.com/blog/topics/systems/mitigating-power-and-thermal-fluctuations-in-ml-infrastructure. [Accessed 01 September 2025].

[45] M. Lyengar, "Data Center Evolution - Powering the AI Revolution Sustainably," in *2025 OCP EMEA Summit*, Dublin, 2025.

[46] Electric Reliability Council of Texas, "ERCOT Large Load Events," 2024.

[47] North American Electric Reliability Corporation, "Incident Review: Considering Simultaneous Voltage-Sensitive Load Reductions," 8 January 2025. [Online]. Available: https://www.nerc.com/pa/rrm/ea/Documents/Incident_Review_Large_Load_Loss.pdf.

[48] P. Matwijec and E. Spears, "Mitigating DC Harmonics," Eaton, 2014. [Online]. Available: https://www.eaton.com/content/dam/eaton/markets/data-center/Mitigating-DC-harmonics.pdf.

[49] L. Nicoletti, N. Malik and A. Tartar, "AI Needs So Much Power, It's Making Yours Worse," Bloomberg Technology, 27 December 2024. [Online]. Available: https://www.bloomberg.com/graphics/2024-ai-power-home-appliances/.

[50] C. Mishra, L. Vanfretti, J. Delaree Jr, T. J. Purcell and K. D. Jones, "Understanding the Inception of 14.7 Hz Oscillations Emerging from a Data Center," *Sustainable Energy, Grids and Networks,* p. 101735, 2025.

[51] *IEEE Standard 1547-2018 - IEEE Standard for Interconnection and Interoperability of Distributed Energy Resources with Associated Electric Power Systems Interfaces*, 2018.

[52] *NERC PRC-006-1 - Reliability Standard for Automatic Underfrequency Load Shedding,* 2011.

[53] a. Chalamala, V. Rabl, S. Abdulsalam, O. (. Adeosun, S. Baker, L. B. Tjernberg, V. Carter-Ridley, T. Horger, K. Garg, A. Gledhill, B. Hederman, R. Kirby, C. Martin and Molly Mooney, "Data Center Growth and Grid Readiness," IEEE, 2025.

[54] R. Bianchini, *Panel Discussion - Data Center Infrastructure,* 2025.

[55] S. Rwejuna, J. Triggs and A. Bryant, "FERC Orders Review of Co-Located Generation for Data Centers in PJM," White & Case, 28 February 2025. [Online]. Available: https://www.whitecase.com/insight-alert/ferc-orders-review-co-located-generation-data-centers-pjm. [Accessed 5 September 2025].