Large Loads 201: Tactical Insights on Forecasting, Ratemaking & Demand-Side Programs

Nina Peluso, Energy Futures Group
Stacy Sherwood, Energy Futures Group
Zack Tyler, Energy Futures Group
Carlos Peña, Energy Futures Group

ABSTRACT

As large electric load requests surge, they are reshaping utility planning, ratemaking, and demand-side program design in real time. This paper offers a tactical overview of how these requests influence state-level proceedings and utility decision-making. Drawing on targeted case studies, the authors first outline how planning processes are adapting to accommodate large load deployment, including the evolution of project-specific validation practices. Second, the authors present foundational ratemaking strategies to protect ratepayers while accommodating rapid load growth. Finally, the paper explores how demand-side programs may evolve in response to these shifts. This paper is intended as a "201-level" briefing for analytic practitioners, regulatory staff, and advocates seeking to understand how large load dynamics are actively shaping utility operations and regulatory frameworks.

Introduction

For decades, risk management has driven electric utility system operations and oversight. Regulatory structures were first developed to manage the risk of reserve shortages and subsequently to ensure the appropriate regulation of electricity as a public good. In many ways, today's headline topic, large load growth, is no different than any other the grid has faced, whether the widespread adoption of in-home appliances in the 1950s, the rise of efficient lighting in the early 2000s, or the policy-driven transition toward lower emissions. In each case, the goal was to balance cost, reliability, environmental, and other risks. The current period, characterized by massive prospective load growth concurrent with the impending retirement of aging thermal generation resources and a transmission system greatly in need of upgrade and expansion faces that same set of risks. However, the sheer scale and speed of the anticipated demand, paired with political pressure to reap the benefits of data center growth, has created a heightened sense of urgency across the industry.

Much like previous periods of transition, large load growth demands new strategies to assess and manage risk. In this paper, we draw a roadmap for consideration of large load growth across three key areas of electric utility management: planning and load forecasting, ratepayer protections and demand side management. Our recommendations, listed in Table 1 below, are key steps to assess and respond to risk. These strategies, if applied thoughtfully, should provide clarity around jurisdiction-specific costs and benefits associated with large load integration.

These strategies do not apply across all U.S. markets and utility territories equally, but we hope that they provide a useful starting point in addressing the emergence of large load growth.

Table 1. Strategies to Manage Emerging Large Load Growth

Planning & Load Forecasting	Ratepayer Protections	Demand Side Management (DSM)
Apply evolving best practices for large load forecasting	Implement targeted tariff and contract design	Increase understanding of data center typologies and efficiency gaps by type
Verify large load project realization using standard interconnection milestones	Set minimum demand charges	Prioritize mid-size and colocation data centers for DSM programs
Assess utility planning portfolios across multiple load forecasts	Track costs and apply rigorous cost allocation approaches	Target DSM interventions during retrofits at the end of server refresh cycles
Work within traditional planning processes	Assess opportunities for demand flexibility and clean energy integration	Institute Strategic Energy Management programs

The Evolution of Planning and Forecasting

Utility planning in the United States was formalized in the 1980s and 1990s via a patchwork deployment of state-level legislative and regulatory action. Each state and region abides by a different set of structures reflecting some combination of utility ownership model, market regulation, and political landscape. In this paper, we focus primarily on integrated resource planning (IRP) and similar long-term planning processes as the leading approach in most U.S. jurisdictions (Biewald et al. 2024). To the extent that a jurisdiction is not covered by such a process, the discussion below may still inform utility planning in other ways.

Traditionally, electric utility generation planning is procedurally divorced from transmission planning, interconnection processes, and ratemaking proceedings. While this approach has served reasonably well to ensure system reliability, cost stability, and compliance with policy mandates, some stakeholders have proposed novel solutions to break down the siloes between existing processes (Burdick et al. 2024). In the meantime, large load[1] integration is already on utility commission dockets across the country. While novel planning methods are introduced to and considered by utility commissions, regulators, utilities, and advocates face questions around large load integration now. Ultimately, generation planning in this context hinges on decisions about who absorbs risk: the prospective large load customer, the host utility, or the base of existing customers, both commercial and residential.

Below, we outline several ways these topics are already under consideration in state regulatory dockets. Though no response or strategy is yet considered a “best practice” given the nascency of these issues, we strive to identify key questions and strategies that can enable legacy regulatory processes to manage load requests with greater speed, scale, and system impacts.

[1] For the purposes of this report, we define “large load” broadly, to include customers including data centers, manufacturing, and industrial processes, characterized by both the speed and the scale of requests for customer interconnection.

Large Load Forecasting

One of the driving forces in most IRP processes is the planning reserve margin – utilities assess resource portfolios based primarily on their ability to meet that reserve margin at the lowest system cost, and in some cases, the highest system benefit. The peak electric load forecast, plus this planning reserve margin, is thus the central assumption which determines ‘necessary’ utility investment identified in an IRP. Where load forecasts are too low, a utility risks underbuilding resources, relying too heavily on market purchases, or even facing unserved energy events. If too high, a load forecast may spur excessive new builds whose costs are then spread across customers – raising rates without measurably improving service. As such, quality load forecasting is essential.

Unfortunately, utility load forecasts reflecting prospective new large load are to-date opaque, may be substantially overstated, and do not follow common procedures across jurisdictions. This lack of clear methodology is no surprise, given that the U.S. hasn’t seen load growth projections this high since the 1960s. Grid Strategies’ annual 2025 U.S. compiled load forecast shows that 2030 peak load has increased by a factor of six, to 166 GW. Further, Grid Strategies finds that the data center portion of these forecasts is likely overstated by about 25 GW (Wilson et al. 2025).

In an effort to identify clear standards and useable approaches to this emerging problem, the Energy Systems Integration Group (ESIG) Large Loads Task Force developed a report identifying ten recommendations for utility forecasting. Each reflects discussion with utility and regional load forecasters and consideration of diverse practices across the country. The ESIG report provides detail for each recommendation and outlines considerations when applying each in specific jurisdictions (ESIG 2025).

When engaging with large load forecasts, whether from the utility, commission, advocate, or academic perspective, the ten recommendations ESIG proposes constitute a roadmap for ensuring strong forecasting. As with all of the strategies proposed in this paper, they are best harmonized with other processes – for example, instituting clear financial requirements at critical project milestones, which can then be used to develop consistent load forecasting approaches.

Mapping Milestones & Collecting Data to Confirm Project Realization

One critical strategy for validating a utility forecast is to determine the likelihood of project completion by identifying the milestones that a prospective customer has hit, particularly in relation to interconnection. With each milestone that a customer hits, the certainty that they will indeed come online increases. Figure 1 below illustrates five of these key milestones and the general nature of the studies or agreements listed, as well as the financial or project action associated with it. Of course, each of these components is subject to jurisdiction-specific requirements. Some utilities, such as those in Northern Virginia, who have long experience managing data center loads, use forecasting approaches that apply a specific statistical likelihood to each project based on which milestones, like those below, it has hit. Others, who have little historical data to rely on and limited insight into other utility experience, may simply use generic demand blocks based on interest received (ESIG 2025).

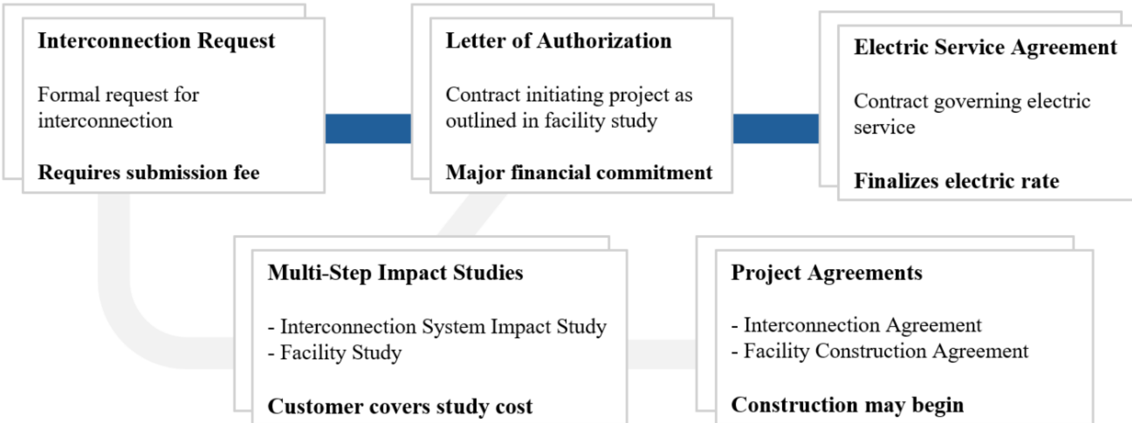

Figure 1. Standard Interconnection Milestones for Prospective Large Load Customers.

Approaches which fail to consider project milestones typically result in forecasts associated with much higher stranded cost risk given the lack of firm project commitments. While each utility must consider its own existing processes, timelines, and policy mandates, the strongest load forecasts typically weigh the completion of milestones above, such as the completion of a Letter of Authorization or Interconnection Agreement, far more heavily (ESIG 2025).

Beyond tracking major milestones, utilities can better assess project viability by requesting foundational data as soon as a prospective customer enters the pipeline. Those data should extend beyond basic project information to cover load characteristics, site and permitting detail, infrastructure readiness, and a verified project timeline. Each category contains critical-path items, without which a utility cannot reliably gauge project status. These include: (1) the developer's experience deploying large load projects, (2) the nature of the load (e.g., data center, steel mill), (3) the peak demand and ramp schedule to that peak, (4) the load models provided to the operating utility, (5) the zoning status of the proposed site, (6) whether a water connection is secured, (7) site control, and (8) the expected commercial operation date. Uncertainty related to any of these elements signals elevated project risk and should be reflected in the load forecast.

Assessing Scenarios Across Different Load Forecasts

In many IRP processes, utilities compare a set of resource scenarios across an array of system-level metrics – reliability, cost, emissions, etc. Typically, this takes the form of some kind of scorecard and either directly informs decision-making or serves as an illustration of key considerations in the final "preferred portfolio" selection. Some states have specific requirements for how portfolios are assessed and selected while others leave that discretion to utilities (Michigan 2022; Louisiana 2021). Of course, these assessments are central, particularly because selected portfolios may (1) inform a forthcoming RFP for resources, (2) inform future project approval proceedings, or (3) even in themselves represent a commission approval of included projects (Biewald et al. 2024). Given the importance of "preferred portfolio" selection, there are three key elements to oversee as large load forecasts grow.

- *Consider multiple futures.* Preferred portfolios should be selected based on performance across multiple load forecasts. Given the increasing uncertainty associated with these load forecasts as outlined above, it is inappropriate to select a preferred portfolio that is optimized to meet a single load forecast. Rather, the preferred portfolio should perform well across multiple potential futures. For example, in its 2025 IRP filing, CenterPoint utilizes a risk analysis assessment comparing outcomes across load forecasts (Centerpoint 2025). While CenterPoint falls short of explicitly utilizing this comparison in its portfolio evaluation, it demonstrates one possible evaluative approach.

- *Reduce reliance on simplified cost assessments.* Recent utility publications and commission statements, such as the AES Indiana 2025 IRP, demonstrate the overuse of simplified rate assessments purporting that new large customers will lower all rates (Massey 2026, California 2026, Supporting 2026). Typically, this assessment takes the form of dividing a system-wide cost estimate by system-wide energy. Given the magnitude of load growth, dividing cost by a larger denominator results in a lower $/MWh value. However, these $/MWh assessments are misleading at best and risk portfolios with massive, stranded cost risk being selected at worst. A $/MWh value is not a rate analysis, and should be reserved for an appropriate proceeding, or represented in full in the IRP proceeding. Further, the use of overstated large load forecasts referenced above exacerbates the issue with this type of analysis.
- *Use non-cost metrics.* Utilities now grapple with the difficult question of how much new load they can serve without sacrificing overall performance or compliance with state policy. While discussion about cost abounds, and will be covered in the next section, the use of multiple load forecasts to assess performance across many metrics, such as reliability and environmental impact, remains paramount. For example, a utility might model four load forecasts and find that the portfolios required to meet two of them fail to meet state zero-carbon goals. That result could force a reckoning on whether the utility can host new large load customers without relaxing those goals.

Reliance on Traditional Planning

This paper focuses primarily on the load forecasting elements of utility planning given they so strongly drive outcomes. However, in the face of the risks associated with new large load customers, it is paramount that regulators continue to rely on foundational planning assessments. In some jurisdictions, utilities have deemed load growth so urgent that they have initiated proceedings outside of standard planning processes. See, for example, the 2025 preapproval application of the Public Service Company of Oklahoma (PSO) for 1,299 MW of new generation (OCC 2025). These off-cycle applications carry major risk of stranded costs, system reliability oversight, and inappropriate cost allocation, discussed in more detail below. Indeed, in Oklahoma, PSO projected a $10/month residential customer rate increase to support a portfolio of projects primarily designed to meet new large customer needs (Garret 2025).

While utility planning processes are perceived as slow and onerous, attempts to sidestep appropriate consideration of load growth via separate filings should face considerable scrutiny. Where processes can be improved, utilities can and should continue to pursue them, but to bypass existing oversight mechanisms risks missing major system impacts from these new customers. As previously stated, large load growth threatens to shift risk onto existing ratepayers, and as long as that holds true, traditional planning oversight is essential.

Protecting Existing Ratepayers

The scale and speed of large load growth may negatively impact existing electric system customers if appropriate ratemaking strategies and protections are not established. Rapid load growth can impact existing customer bills, require procurement of new capacity, require major investments in transmission and/or distribution infrastructure, and limit or eliminate progress on renewable energy and emission reduction goals. As seen across proceedings since 2024, it is important to include safeguard provisions in tariffs and special contracts to protect existing ratepayers. These provisions include minimum contract length, minimum demand charges, rigorous cost tracking and allocation methodologies, and updated cost of service studies. As large loads come online, additional changes are likely necessary to provide reliable service at reasonable rates. Utilities may pursue a variety of revisions to rate design, standard contracts, and cost allocation methods, and consider clean tariffs and bring-your-own generation models.

Contract and Tariff Design

One of the primary ways to protect existing ratepayers is through a tariff and/or contract design that (1) ties the contract length closely to the cost recovery of assets needed to bring the large load online, (2) prevents stranded costs, (3) discourages over procurement, and (4) establishes collateral requirements.

Over the last two years, as utilities have established new large load tariffs, project contract terms have lengthened, reflecting an effort to closely tie the cost recovery of the new assets to the large load customer requiring their construction. The average minimum contract length prior to 2024 was five years; however, since 2024 average minimum contract length increased to 10.9 years (RMI 2025). The majority of the large load tariffs introduced in 2025 had a 15-year contract term, but some were as low as twelve (SEPA DELTa 2025). Given the length of time to recover the cost of assets associated with large load growth, a 15- to 20-year contract will limit the risk of existing ratepayers stranded with assets built to meet unrealized load.

While assets built to serve large loads may eventually serve overall system needs, if a large load customer exits early, they may not be obligated to cover those costs, leaving them to remaining customers. One strategy to prevent existing ratepayers from such stranded cost risk is the use of an exit fee clause whereby if a customer ceases service or defaults prior to the conclusion of the contract they must pay a substantial fee. Similarly, collateral requirements throughout the contract period that decline over time are another avenue through which utilities can ensure cost recovery from the customer responsible for an infrastructure investment.

Currently, there is a sense of urgency from data center developers to identify service territories with available capacity or willingness to add capacity quickly. Coupled with constraints on the supply chain for grid infrastructure, developers and data centers may in fact seek more capacity than actually needed to hedge against competition.[2] As a result, hyperscale data centers are seeking interconnection with massive energy needs ramping up over the next ten years. During that time, computing may see increased efficiencies or other operational changes that require less utility power. To protect against stranded costs, utilities should develop contracts and tariffs that require ramping requirements within a five-year period and that obligate the customer to pay for full contracted load if it cannot be reassigned.

[2] The industry is experiencing a supply chain shortage for transmission infrastructure, including transformers, which currently results in project timelines extensions by 4-5 years (Patel 2025).

Monthly Minimum Demand Requirements

Minimum billing demand is another avenue for ratepayer protection, especially given the growth of data centers with 24/7 usage patterns. Minimum billing demand requirements establish a minimum payment for the customer regardless of actual usage designed to ensure that the utility receives revenues for fixed infrastructure costs, grid reliability, and recovery of expenses for capacity reserved for a specific customer or customer class. The minimum billing demand should be designed with consideration for the need to accelerate cost collection within the contract term.

Recent large load tariffs have relied on an 80% minimum billing demand; however, there are some utility tariffs that have established an 85% or 90% minimum billing demand. Dominion Energy in Virginia imposed a split minimum demand charge for loads that are 25 MW or greater, by assigning minimum demand charges consisting of (1) a minimum distribution charge of 85% of contracted demand, (2) a minimum transmission charge of 85% of contracted demand, and (3) a minimum generation charge of 60% of contracted demand (SCC 2025). This design was proposed to protect existing ratepayers from rising electricity costs associated with data center growth in the service territory.

Tracking and Allocation Costs

Thoughtful tariff and contract development are the first steps to protect existing ratepayers. A second step is tracking utility investments tied to large load service, both those serving individual large load customers, and those that may benefit the entire rate base.

Utility costs incurred solely to serve one large customer should be tracked and associated with that customer using contributions in aid of construction (CIAC), thus avoiding inclusion in rate base through construction work in progress (CWIP) and limiting the cost exposure for existing customers.

Utility costs for infrastructure benefitting more than one large customer should be tracked using allowance for funds used during construction (AFUDC), delaying the inclusion into rate base until the assets are in service. This limits rate shock for existing ratepayers prior to the in-service date and ensures that appropriate costs are allocated to and paid by the new customer. When the AFUDC is included into rate base, the utility must prove prudency of the investment for existing ratepayers, as it must for all investments.

Critically, during the rate case where these investments are incorporated into rate base, a new cost of service study (COSS) should reflect these investments fully and ensure proper allocation to the large load class/customer. Given the nascency of hyperscale development and uncertainty around operation, commissions may consider required a new COSS the year following full customer operation reflecting actual operation to avoid cross-subsidization.

Demand Flexibility and Clean Energy Policies

Many data center developers request around-the-clock power delivery, though their specific operation patterns remain opaque. Some operators have raised concerns associated with shifting peak periods, unpredictable grid oscillation, and risks to system reliability during periods of high grid stress. Some developers have proposed flexibility programs similar to demand response and the use of interruptible rates paired with onsite energy or management systems as opportunities to mitigate critical peaks or avoid the creation of new system issues.

Utilities are considering flexibility opportunities through bring-your-own-generation (BYOG) that could be co-located or contracted to provide load flexibility services. Data center developers are particularly interested in the opportunity to develop projects faster with BYOG than they might otherwise via standard interconnection queues. Others have proposed BYOG structures with clean energy, though it is not yet clear what the emissions or grid impacts of such a structure would be.

Another opportunity, as mentioned above, may be interruptible rate design to incentivize data centers to shift computational tasks depending on grid needs. This might include delaying non-urgent tasks or shifting workloads to other locations. Operators could theoretically harness this flexibility to avoid critical peaks. This coupled with building and maintaining efficient data centers as discussed in the next section, regardless of size, could reduce overall energy demand. Of course, as more states pilot these programs and onboard new large load customers, state regulators must carefully assess the actual realized benefits of flexibility or BYOG policies.

Demand Side Management Opportunities for Large Loads

Data Center Typologies and Energy Use

Data centers vary widely in size, capacity, and operational purpose, ranging from hyperscale facilities to smaller installations supporting localized or application-specific computing needs. Common configurations include centralized server rooms, server cabinets, embedded data centers, distributed or edge facilities, colocation data centers, and hyperscale facilities.

At the largest scale, hyperscale data centers are distinguished by their massive computational capacity and strong emphasis on efficiency. These facilities typically deploy advanced cooling technologies, optimized power distribution architectures, and increasingly sophisticated workload management systems to improve performance and reduce energy consumption. In contrast, smaller facilities, including colocation, embedded, and edge data centers, generally consume less total energy but operate with fewer efficiency optimizations. They also may not employ an on-site energy manager.

Evaluating energy efficiency and demand management opportunities across different data center typologies is increasingly important. Improving efficiency in smaller facilities can reduce operating costs and energy waste while lowering environmental impacts. In addition, participation in demand response programs and the integration of distributed energy resources may enhance grid resilience and improve the flexibility of large digital loads (Shehabi et al. 2016). Despite these advances, significant efficiency differences remain across data center types.

Efficiency Gaps Across Data Center Types

According to the United States Data Center Energy Usage Report, small and medium-sized data centers typically exhibit higher Power Usage Effectiveness (PUE) values (ranging from approximately 1.68 to 1.91) indicating lower overall infrastructure efficiency compared with large-scale facilities (Shehabi et al. 2024). In contrast, hyperscale data centers and facilities optimized for artificial intelligence workloads often report substantially lower PUE values, approximately 1.14 and 1.22, respectively.

Since PUE approaches 1.0 as infrastructure efficiency improves, this disparity highlights a substantial performance gap between small and hyperscale facilities. The difference suggests considerable untapped opportunities for efficiency improvements in small and medium-sized data centers through enhanced cooling design, optimized power distribution, improved airflow management, and more advanced operational controls (EPA 2023). While large data centers may be efficient when constructed, they also undergo regular changes based on short equipment lifecycles and rapidly evolving digital infrastructure technology. Further, purely based on scale, efficiency measures at large data centers have an outsized impact.

Hardware Lifecyle and Operational Efficiency

Most data center operators refresh server hardware on a three- to five-year cycle. When upgrades are delayed beyond this window, operators can miss opportunities to deploy newer, more energy-efficient technologies with higher computational performance for lower power consumption (Masanet et al. 2020). As server efficiency improves through advancements in processor design, virtualization, and power management features, maintaining outdated or underutilized equipment can increase both operating costs and energy intensity.

Operators can address these challenges through periodic infrastructure assessments that evaluate system performance, energy consumption, utilization rates, and total cost of ownership. Regular technical reviews, combined with benchmarking against current technologies, can clarify whether modernization, consolidation, or virtualization strategies are economically and operationally justified. These refreshes create an opportunity for Demand Side Management (DSM) program intervention in the form of data center energy audits, discussed further at the end of this section. To the extent that efficiency code or standards were introduced at the state or federal level, operators could more easily measure performance against peer operations.

Efficiency Challenges in Small and Embedded Data Centers

Energy efficiency initiatives in data centers have historically focused on large-scale facilities. However, smaller installations, including server closets, server rooms, and localized data centers with less than 5,000 square feet of computer floor space, constitute a large share of computing infrastructure. Historically, these smaller sites house more than half of all servers in the United States (LBNL 2017). However, the rapid expansion of hyperscale and colocation capacity during 2024–2026 has likely change this share substantially (LBNL 2024), and smaller sites may no longer represent the majority of installed servers.

While they typically fall below the size threshold for utility demand-side management programs, issues at these facilities illustrate many of the structural barriers that limit efficiency improvements across the sector. Smaller and embedded data centers often lack the dedicated energy managers/staff, technical expertise, and financial resources needed to evaluate and implement efficiency measures. In addition, their energy use is frequently embedded within broader building operations, making costs more difficult to track and prioritize (EPA 2022). These same conditions also make small and embedded facilities difficult to reach through utility programs: operators may lack control over the space and infrastructure systems where savings reside, particularly in leased or shared environments, and energy typically represents a small share of operating expense, leaving efficiency a low priority. Partly for these reasons, programs aimed at small data centers have historically achieved limited participation and savings.

Even so, a small number of utilities have begun designing demand-side management offerings specifically for smaller and embedded facilities. Dominion Energy's Data Server Room Energy Efficiency Program, for example, targets server rooms located within offices, hospitals, universities, and other non-dedicated buildings, beginning with an optional audit and providing rebates for measures such as high-efficiency computer-room air conditioning (CRAC/CRAH) units, high-efficiency power supplies, temperature setpoint adjustments, and lighting occupancy sensors (Dominion Energy 2026). CenterPoint Energy's Data Center Energy Efficiency Program similarly offers no-cost technical assistance and performance-based incentives—paid on verified peak demand reduction and annual energy savings—to facilities including server rooms, network operations centers, and enterprise data centers, with eligibility extending to smaller, non-transmission-class customers (CenterPoint Energy 2026). Program participation remains modest, but these examples suggest that purpose-built program designs can begin to address the barriers described above.

Although this paper focuses on large load projects, similar efficiency opportunities apply across all data center types, including improved airflow management, optimized cooling systems, and better operational controls.

Infrastructure Energy Distribution and Cooling Efficiency

Table 3 shows typical end-use energy distribution across different data center configurations. The largest share of electricity consumption is associated with information technology (IT) equipment loads, followed by cooling systems and power conditioning equipment such as uninterruptible power supplies (UPS) and transformers (Shehabi et al. 2016).

Table 3. Typical End-Use Energy Distribution by Data Center Type

Space type	IT load	Transformer	UPS	Cooling	Lighting
Closet	50.0%	2.5%	–	46.5%	1.0%
Room	40.0%	2.0%	8.0%	49.2%	0.8%
Localized	50.0%	2.5%	10.0%	36.5%	1.0%
Mid-tier	52.6%	2.6%	10.5%	33.2%	1.1%
High-end	58.8%	1.8%	5.9%	32.4%	1.2%
Hyperscale	83.3%	1.7%	–	13.3%	1.7%

Figure 2 presents more recent data from the IEA showing electricity consumption by equipment type across enterprise, colocation/service provider, and hyperscale data centers. Although the categories are not identical to those used in Table 3, the relative pattern is consistent: IT equipment remains the dominant end use, while cooling continues to represent the largest non-IT load, particularly in enterprise and smaller facilities. The IEA reports that servers account for roughly 60% of electricity demand in modern data centers on average, while cooling ranges from about 7% in efficient hyperscale facilities to more than 30% in less-efficient enterprise facilities.

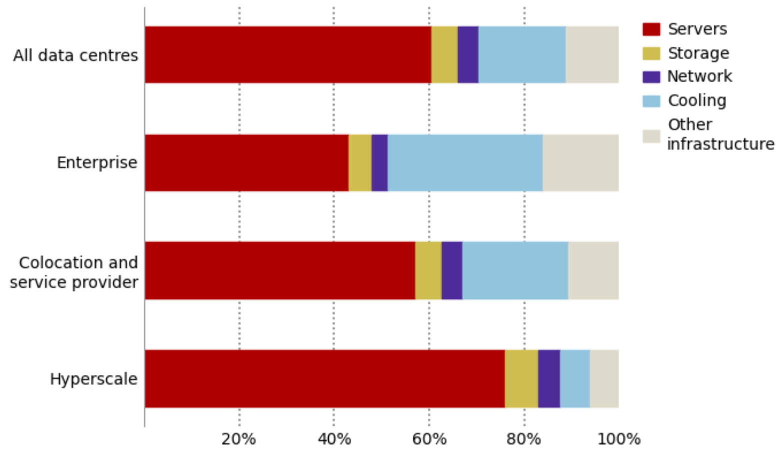

Figure 2. Share of electricity consumption by data center and equipment type. Adapted from (EIA 2025)

Given cooling represents a significant portion of data center energy use, it follows that efficiency improvements related to cooling may have substantial impact. The share of electricity consumed by cooling systems varies widely depending on facility type, cooling technology, and operating practices (Shehabi et al. 2016). Table 4 summarizes cooling approaches commonly employed across different data center configurations.

Table 4. Typical Cooling Approaches by Data Center Type

Space type	Cooling characteristics	Notes
Closet / Room	Minimal to basic cooling	Often lacks redundancy and may operate inefficiently
Mid-tier	Dedicated and redundant cooling systems	Efficiency varies depending on system design and operating practices
High-end	Advanced redundant CRAH or CRAC systems with economizers	Maintains stable environmental conditions to meet availability requirements
Hyperscale	Purpose-built high-efficiency cooling systems	Often incorporates economization and higher temperature setpoints; includes liquid cooling technologies for server racks.

Several strategies can improve cooling system efficiency, including improved airflow management, higher operating temperature setpoints, advanced cooling technologies, and coordinated system controls to minimize mechanical cooling demand (LBNL 2024).

- Airflow management. Implement hot- and cold-aisle containment, install blanking panels to seal unused rack spaces, and improve cable management to reduce air recirculation and bypass airflow.
- Higher temperature setpoints. Follow recommended environmental guidelines from ASHRAE TC 9.9. Increasing supply air temperatures to approximately 65–80°F can reduce cooling energy consumption while maintaining safe operating conditions for IT equipment (ASHRAE 2016).

- Economizers and liquid cooling. Deploy air-side or water-side economizers where climate conditions permit. For high-density computing environments, advanced solutions such as direct liquid cooling or immersion cooling can further improve thermal management efficiency and reduce reliance on mechanical refrigeration (IEA 2024).

Colocation and Wholesale Data Center Opportunities

Traditional colocation data centers continue to proliferate and can be strong candidates for demand-side management (DSM) interventions. While hyperscale facilities often attract national attention due to their size, colocation facilities may present a more practical target for DSM programs. These facilities are widespread, represent a growing proportion of system load, and their operators may be more receptive to technical assistance and financial incentives offered through traditional utility DSM programs. As discussed in earlier sections, these facilities often exhibit higher PUE values and therefore present greater value for energy efficiency improvements. This combination of scale, prevalence, and program responsiveness suggests that colocation facilities are a promising target for DSM interventions, particularly after initial construction when server refresh cycles and operational optimization activities begin.

DSM Program Design Opportunities

The discussion above drives toward several findings around DSM programs for data centers. First, we recommend that DSM programs prioritize mid-size and colocation data centers in their outreach and coordination efforts. Depending on jurisdiction, this customer segment represents a high proportion of system load where relatively small operational, or infrastructure improvements could produce meaningful energy savings. New construction opportunities may be somewhat limited, given many modern data centers are already designed with best-in-class technology, though cooling system selection and efficiency can always be re-evaluated, as cooling is the largest non-IT end use in most data centers.

Retrofit scenarios present a larger opportunity for DSM intervention. Opportunities may arise as early as three years after facility operation due to the server refresh cycle and the continuous improvement of IT infrastructure. Offering targeted audit programs at the retrofit stage could yield substantial energy savings because server refresh cycles regularly change equipment density, rack layout, and computing workloads, meaning optimal operating conditions shift over time. Activities such as rerouting cables, adjusting temperature setpoints, optimizing airflow management, balancing computing loads, and improving server utilization are some areas where DSM programs can support facility operators to achieve sustained energy savings.

These approaches are not merely theoretical; a small number of utilities have begun running data center–specific programs of this kind. CenterPoint Energy's Data Center Energy Efficiency Program, for instance, combines no-cost technical assistance with performance-based incentives tied to verified peak demand reduction and annual energy savings across IT, cooling, and power systems—effectively a custom, calculated-savings model scaled to large data center customers and designed to optimize energy use without compromising reliability (CenterPoint Energy 2026).

Strategic Energy Management (SEM) programs may also prove useful to improve data center energy performance. SEM emphasizes organizational practices, continuous improvement, and operational optimization. The evolving nature of data centers creates opportunities for SEM programs to support ongoing efficiency improvements. These programs are likely most effective in facilities with dedicated energy managers who actively track performance metrics and can implement organizational changes. SEM programs may be less effective in hyperscale facilities that already operate highly optimized internal energy management systems or in retail colocation environments where tenants control IT loads. Nonetheless, SEM initiatives can help facility operators identify operational improvements, monitor performance indicators such as PUE, and integrate energy management practices into routine facility operations.

Large operators are also beginning to pair efficiency with demand flexibility, expanding the DSM toolkit available to utilities serving hyperscale loads. In 2025, Google entered agreements with Indiana Michigan Power and the Tennessee Valley Authority to deliver demand response by curtailing machine-learning workloads during grid events—its first programs targeting ML loads—and by early 2026 reported roughly 1 GW of demand response capacity contracted across utility partners, including Entergy Arkansas, Minnesota Power, and DTE Energy (Google 2026). In June 2026, Google extended this strategy into the wholesale market by funding a three-year, 100 MW virtual power plant in PJM through the aggregator Voltus under a "bring-your-own-capacity" arrangement—reportedly the first time a hyperscaler has directly financed a virtual power plant to help meet data center capacity needs (Voltus 2026). Although such efforts remain early-stage, they demonstrate that demand-side engagement with large loads is moving from concept toward contracted practice.

Taken together, these findings suggest that DSM programs focused on operational optimization, targeted audits, and mid-life infrastructure improvements offer high potential for improving energy performance in data centers.

Conclusion

Utility planners, regulators, advocates, developers, and prospective new customers all must adopt proactive strategies to manage risk associated with large load growth. The recommendations outlined in Table 1 and detailed across this paper provide a roadmap to align large load integration with long-term system goals.

References

ASHRAE (American Society of Heating, Refrigerating and Air-Conditioning Engineers). 2016. Data Center Power Equipment Thermal Guidelines and Best Practices. Atlanta, GA: ASHRAE. ashrae.org/file%20library/technical%20resources/bookstore/ashrae_tc0909_power_white_paper_22_june_2016_revised.pdf.

Biewald, B., D. Glick, S. Kwok, K. Takahashi, J.P. Carvallo, and L. Schwartz. 2024. *Best Practices in Integrated Resource Planning.* Berkeley, CA: Lawrence Berkeley National Lab. eta-publications.lbl.gov/sites/default/files/2024-12/irp_best_practices_2024_synapse_lbnl_24-061_0.pdf.

Burdick, A. J. Hooker, L. Alagappan, M. Levine, and A. Olson. 2024. *Integrated System Planning: Holistic Planning for the Energy Transition.* Energy and Environmental Economics. ethree.com/wp-content/uploads/2024/10/E3-ISP-Whitepaper.pdf

CenterPoint Energy, 2025. *2025 Integrated Resource Plan, Volume 1.* CenterPoint Energy. centerpointenergy.com/en-us/Documents/Midwest/PUBLIC-CEIS-2025-IRP-Volume-1-of-2.pdf, page 147.

CenterPoint Energy. 2026. Data Center Energy Efficiency Program. Houston, TX: CenterPoint Energy. centerpointenergy.com/en-us/SaveEnergyandMoney/Pages/Data-Center-Energy-Efficiency-Program.aspx.

California State Assembly, 2026. Committee on Utilities and Energy, and Committee on Privacy and Consumer Protection. *Joint Oversight Hearing on AI's Energy Impacts: Background Paper*. autl.assembly.ca.gov/system/files/2026-01/01.28.2026_joint-privacy_ai-oversight-hearing-background.pdf

DCEO (Illinois Department of Commerce and Economic Opportunity). 2024. *Data Center Investment Tax Exemptions and Credits.* Springfield, IL: Illinois Department of Commerce and Economic Opportunity. dceo.illinois.gov/expandrelocate/incentives/datacenters.html.

DCEO (Illinois Department of Commerce and Economic Opportunity). 2024. *Data Center Investment Program Annual Report.* Springfield, IL: Illinois Department of Commerce and Economic Opportunity. dceo.illinois.gov/content/dam/soi/en/web/dceo/aboutdceo/reportsrequiredbystatute/2024-data-centers-annual-report-submission.pdf.

Dominion Energy. 2026. Data Server Room Energy Efficiency Program. Richmond, VA: Dominion Energy. domsavings.com/business-program/data-server-rooms-program.

EIA (International Energy Agency). 2025. Energy and AI, Paris https://www.iea.org/reports/energy-and-ai, Licence: CC BY 4.0

EPA (U.S. Environmental Protection Agency). 2023. 5 Simple Ways to Avoid Energy Waste in Your Data Center. Washington, DC: U.S. Environmental Protection Agency. energystar.gov/products/data_center_equipment/5-simple-ways-avoid-energy-waste-your-data-center.

EPA (U.S. Environmental Protection Agency). 2022. Embedded Data Centers. Washington, DC: U.S. Environmental Protection Agency. energystar.gov/products/data_center_equipment/embedded-data-centers.

ESIG (Energy Systems Integration Group). 2025. *Forecasting for Large Loads: Current Practices and Recommendations.* A report by the Large Loads Task Force. esig.energy/wp-content/uploads/2026/02/ESIG-Large-Loads-Forecasting-report-2025.pdf

Garret, M., 2025. Responsive Testimony of Mark E. Garret. OCC Case No. PUD2025-00064. public.occ.ok.gov/WebLink/DocView.aspx?id=20432961&dbid=0&repo=OCC&searchid=fdd7a8bb-f446-4c4b-a72f-eca575844af6

Google. 2026. "Google Signed 1 GW of Data Center Demand Response." Mountain View, CA: Google. blog.google/innovation-and-ai/infrastructure-and-cloud/global-network/demand-response-data-center-milestone/.

IEA (International Energy Agency). 2024. Energy and AI. Paris: International Energy Agency. iea.blob.core.windows.net/assets/de9dea13-b07d-42c5-a398-d1b3ae17d866/EnergyandAI.pdf.

LBNL (Lawrence Berkeley National Laboratory). 2024. Data Center Energy Efficiency Actions Master List. Berkeley, CA: Lawrence Berkeley National Laboratory. datacenters.lbl.gov/sites/default/files/2024-04/DCEE%20Actions%20Master%20List_090920_final.pdf.

LBNL (Lawrence Berkeley National Laboratory). 2017. Small Data Centers. Berkeley, CA: Lawrence Berkeley National Laboratory. datacenters.lbl.gov/small-data-centers.

Louisiana Public Service Commission Staff. 2021 *Integrated Resource Planning Rules for Electric Utilities in Louisiana.* Baton Rouge, Louisiana: Louisiana Public Service Commission. lpscpubvalence.lpsc.louisiana.gov/portal/PSC/ViewFile?fileId=ZHWue07L6hc%3D.

Masanet, E., A. Shehabi, N. Lei, S. Smith, and J. Koomey. 2020. "Recalibrating Global Data Center Energy-Use Estimates." Science 367 (6481): 984–86. science.org/doi/10.1126/science.aba3758.

Massey, Joel, 2026. "Public Service Commission enacts new lightning initiative to fast track large-load electricity projects." *MyArkLaMiss.com*. myarklamiss.com/news/public-service-commission-enacts-new-lightning-initiative-to-fast-track-large-load-electricity-projects/

Michigan Public Service Commission Staff. 2022. *Revised Integrated Resource Plan Filing Requirements.* Lansing, MI: Michigan Public Service Commission. mi-psc.my.site.com/sfc/servlet.shepherd/version/download/0688y000003OYZ2AAO

Oklahoma Corporation Commission, 2025. *Application of Public Service Company of Oklahoma for an Order of the Commission Preapproving the Acquisition and Cost Recovery of Generation Facilities.* Cause No. PUD2025-000064. public.occ.ok.gov/WebLink/DocView.aspx?id=19510270&dbid=0&repo=OCC&searchid=02b4ef3f-dded-430b-96d0-eb60cc19cb9b

Patel, Sonal C. 2025. "Out of Sync: The Infrastructure Misalignment Undermining the U.S. Grid. Power Magazine. powermag.com/out-of-sync-the-infrastructure-misalignment-undermining-the-u-s-grid/.

RMI (Rocky Mountain Institute) 2025. Alyssa Perez, Sarah Wang, Lauren Shwisberg. "Large Energy Users Want Power. Here's How to Protect Other Ratepayers from the Costs," Rocky Mountain Institute, November 7, 2025. rmi.org/large-energy-users-want-power-heres-how-to-protect-other-ratepayers-from-the-costs/.

SCC (Commonwealth of Virginia State Corporation Commission), 2025, Final Order. Case No. PUR-2025-00058, Application of Virginia Electric and Power Company for a 2025 biennial review of the rates, terms and conditions for the provision of generation, distribution, and transmission services pursuant to §56-585.1A of the Code of Virginia. scc.virginia.gov/docketsearch/DOCS/89g601!.PDF, page 25.

SEPA (Smart Electric Power Alliance) DELTa (Database of Emerging Large-Load Tariffs) 2025. sepapower.org/large-load-tariffs-database/.

Shehabi, A., S. J. Smith, A. Hubbard, D. Sartor, R. Brown, M. Herrlin, J. Koomey, E. Masanet, N. Horner, I. Azevedo, and W. Lintner. 2024. *United States Data Center Energy Usage Report.* Berkeley, CA: Lawrence Berkeley National Laboratory. escholarship.org/uc/item/32d6m0d1.

Shehabi, A., S. J. Smith, D. Sartor, R. Brown, M. Herrlin, J. Koomey, E. Masanet, N. Horner, I. Azevedo, and W. Lintner. 2016. *United States Data Center Energy Usage Report.* Berkeley, CA: Lawrence Berkeley National Laboratory. eta-publications.lbl.gov/sites/default/files/lbnl-1005775_v2.pdf.

"Supporting Growth in Central Indiana with Transparency and Commitment." AES Indiana, September 19, 2025. AES Indiana. aesindiana.com/press-release/supporting-growth-central-indiana-transparency-and-commitment-0.

Voltus. 2026. "Voltus and Google Announce Bring Your Own Capacity Agreement." San Francisco, CA: Voltus. voltus.co/press/voltus-google-bring-your-own-capacity-pjm.

Wilson, J. S. Meyer, Z. Zimmerman, and R. Gramlich. 2025. *Power Demand Forecasts Revised Up for Third Year Running, Led by Data Centers.* Grid Strategies. gridstrategiesllc.com/wp-content/uploads/Grid-Strategies-National-Load-Growth-Report-2025.pdf

AI-assisted tools used: ChatGPT
Used for: Grammar and spell check; Citation formatting
Extent: Light editing
Human review: The authors reviewed and verified all content, calculations, and citations.