# DODGING THE FIRM FIXATION FOR DATA CENTERS AND THE GRID

**Eric G. Gimon**

**Senior Fellow, Energy Innovation**
**November 2025**

# ACKNOWLEDGEMENTS:

I would like to thank my Energy Innovation colleagues for their patient review and feedback on various drafts of this report. I would also like to acknowledge the Energy Systems Integration Group for their conferences, taskforce meetings, and webinars which helped me get started on my research.


# TABLE OF CONTENTS

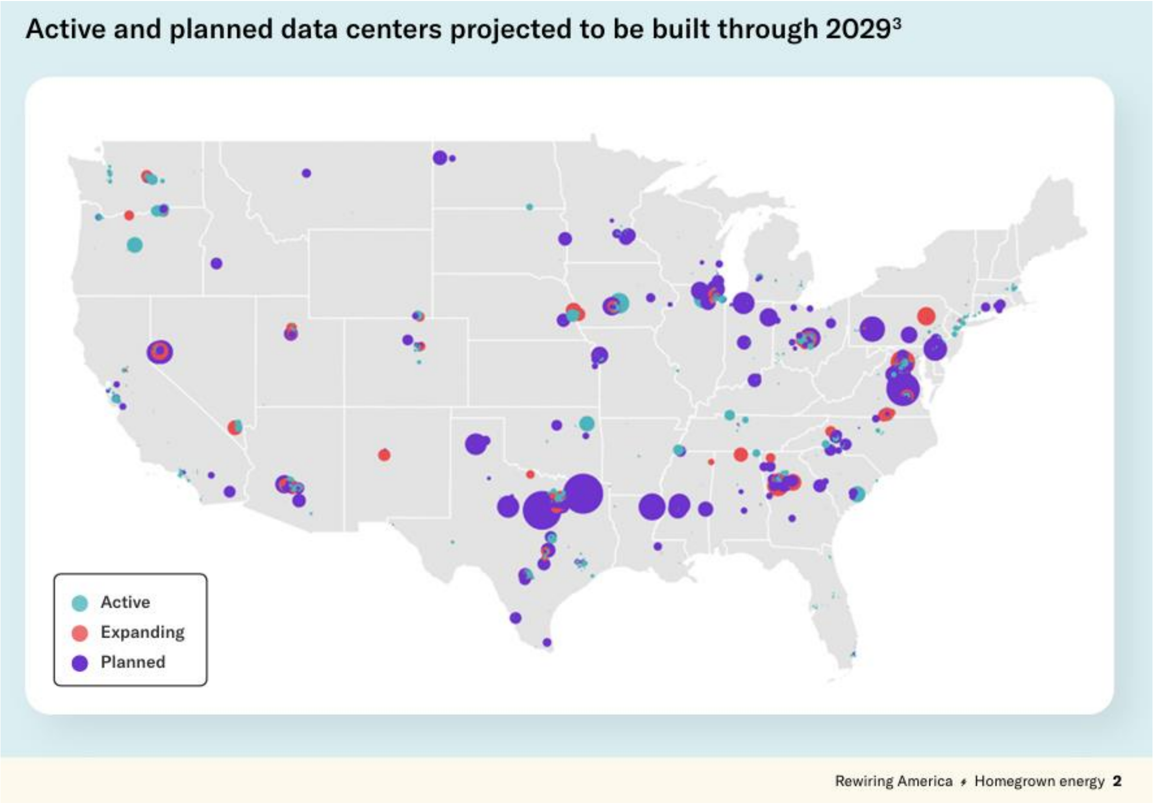


# EXECUTIVE SUMMARY

In multiple states[1,2] (see Figure 1) massive new data center campuses and a coterie of smaller ones have reversed years of flat or declining electricity demand, leaving utilities and policymakers scrambling for solutions.

*Figure 1 Existing and Projected data centers from Rewiring America's "Homegrown Energy" report[3].*




Faced with this onslaught of new demand, many utilities and developers are depending on old habits by adding new gas plants, refurbishing coal units, or turning to nuclear partnerships along with extensive grid upgrades near new load centers – the "firm fixation."

It reflects a belief that only firm resources and major transmission upgrades can handle data centers' needs. Yet this approach overlooks two essential truths: (1) power plants and data centers are both parts of a larger, interconnected system, and (2) data center loads, especially those driven by artificial intelligence (AI), are far more dynamic than the flat, baseload profiles they are often assumed to be. Firm fixation leads utilities and regulators to default to outdated firm-generation solutions instead of modern, modular approaches that consider the full complexities of today's power grid. At the scale of even the most compact new data centers, connecting to the grid is no small matter.

Regardless of approach, three features of recent growth are well known to the electricity industry and policy community, and to some extent the wider public. First, new data center load is being amplified by extreme investment interest in AI. Second, incremental load tends to be highly concentrated due to the nature of the growing individual server need for power and the geographic concentration of data centers. Third, the data center industry's appetite for new growth is so large, and other facility capital costs so high that new project owners are willing to pay more for power than average existing electricity consumers.

In a 2024 brief, Energy Innovation proposed instead that a portfolio of solutions – clean energy portfolios, advanced transmission technologies, demand-side flexibility, and efficiency – could work together to obviate the need to rush to meet demand with new fossil generation.[4] Reality so far has deviated significantly from this vision, setting up the power sector for failure: Either new demand will not be met or the negative cost and performance impacts of doing so on other grid users will challenge electricity markets and other long-standing arrangements in a dangerous manner.

The mad scramble to meet data center demand using traditional but crude resource investment methods can create potential missed opportunities to manage load growth that come from a deeper understanding of data centers. Of course, their electricity demand is problematic because it is concentrated, growing fast, and willing to outspend other users. However, it is also far more complex than the flat, 24/7 block it is often assumed to be. This primer identifies six defining features that provide a more nuanced version picture of data centers:

- **Agency and Split Incentives** – Multiple actors (developers, operators, and tenants) and ownership or usage types of data centers create a divided responsibility over grid interaction and access to energy-saving incentives that complicates energy decisions.
- **Clustering** – Facilities tend to concentrate geographically, amplifying local grid stress and transmission costs while creating systemic planning challenges.
- **Consumption Profiles** – Loads are not 24/7 blocks. Instead, they are choppy, with swings of hundreds of megawatts over short intervals, undermining assumptions of steady baseload behavior and potentially affecting the stability of the grid if safeguards are not put in place.
- **Flexibility** – While some AI-driven workloads can be scheduled for off-peak hours, this flexibility is uneven across facility types and even within users in the same data center campuses. While modest levels of curtailment or load-shifting based demand response during peak hours could ease interconnection bottlenecks and peak demand requirements, these may work best in combination with battery energy storage to overcome split incentives and other complexities.

- **Backup Requirements** – Current reliance on diesel for backup generation is unsustainable. Batteries and longer-duration storage are cleaner, more scalable options that provide knock-on benefits for the grid if allowed to participate as both backup and demand response.
- **Modularity** – Data centers grow in phases just as demand grows in phases rather than all at once, aligning poorly with "lumpy" firm large one-time investments in dispatchable power plants and infrastructure upgrades, while fitting well with modular renewables and battery deployments.

When examined as a whole, these features undermine the firm fixation logic. One-to-one matching of data centers with dedicated or "captive" firm power plants is particularly unwise for both the power generator and the new data centers, even given their willingness to pay for speed-to-power. Relying on captive plants for all supply such as pairing a nuclear plant with a large data center exposes them to outages, inflexibility, and stranded-asset risks, while hybrid co-location deals still rely heavily on the broader grid.

Most new demand will need to be served fully or in-part through the bulk power system, requiring upgrades in three key areas: **connection infrastructure, grid services (especially peak capacity), and bulk electricity supply.**

Once this is established, it's clear that data centers can tap the grid's advantages as a "system of systems" that pools variable demand and generation resources solutions together and ensures supply and demand match in real-time. As peak demand rises, this crucial service must be met, but not necessarily by firm generation. A deeper understanding of data center demand attributes yields a more complete solution set which includes data center flexibility, onsite storage, portfolios of clean energy, and others.

The challenges data centers pose include lengthy interconnection queues, peak stress, price impacts, and rising emissions – but these are not insurmountable. Three core lessons emerge for policymakers and stakeholders:

- The process of connecting any new **large load is a key leverage point.** It is the moment to ensure consumption tariffs reflect cost causation, encourage flexibility, and align incentives without imposing unworkable burdens later. Interconnection is the moment of maximum leverage: not to extract unreasonable concessions, but to ensure new entrants cover the full costs of the infrastructure they trigger, and to nudge data center developers towards solutions such as flexible demand or local storage that relieves local bottlenecks and supports the broader grid. Likewise, developers and customers should lean toward local fixes that speed access to the grid, improve power quality, and ease broader impacts—reducing the likelihood of being saddled with extraordinary requirements later.

- **Demand side is a resource hiding in plain sight**. Household electrification and distributed resources can free up tens of gigawatts (GW) at costs comparable to new gas plants and on a faster timetable, offering a more pragmatic and equitable path to integration. Yet at the state and regional level, policy innovation still lags behind. However, several widespread mechanisms exist to channel data center owners and operators' willingness to pay into new solutions that help other existing customers accommodate rapid data center load growth in a fair, fast and equitable way. Because grid connection bottlenecks can be managed by multiple possible combinations of diverse resources, data centers don't need to do all the work of mitigating their grid impacts onsite or through a single counterparty. Once a data center has invested in flexibility and equipment to resolve local connection issues, additional constraints such as upstream transmission and grid services bottlenecks as well as large incremental amounts of annual electricity delivery can be addressed with demand-side solutions from other grid users. A recent report from Rewiring America proposes that many of the resources needed to meet data center load growth could come from sponsoring household upgrades instead of new generation.[5]
- **Storage and flexibility deliver a two-for-one win.** Batteries and managed demand not only ease all manner of data center impacts but can also accelerate renewable integration, providing cleaner, faster, and cheaper capacity than firm fossil solutions. Because batteries are increasingly essential for buffering, backup, and power quality, they also provide a built-in solution for integrating variable renewables—a two-for-one advantage. Furthermore, these renewable-plus-battery solutions can capitalize upon existing surplus interconnection to more quickly connect data centers to the grid in co-located arrangements.

This report challenges the electricity and data center industries to move beyond a firm fixation and adopt solutions that leverage the full capabilities of modern power systems.

The next section describes six defining features of data centers: agency, clustering, consumption profile, flexibility, backup needs, and modularity. We then pivot to explaining why traditional firm responses fall short within the broader context of how the modern grid supplies power to consumers, especially large, new consumers. We will look at how new, modular solutions can meet digital demand more effectively. These steps will depend on a more nuanced understanding of data centers, as opposed to how they are often imagined.

Our hope is that this information will empower policymakers to make wiser decisions when faced with AI growth and proposed public investments, avoiding a firm fixation on simplistic approaches and reaching for more realistic answers that embrace the full complexities of the challenge that rapid data center load growth presents today. By moving beyond simplistic assumptions, policymakers can avoid overcommitting to outdated firm resources and instead adopt strategies that embrace modularity, flexibility, and clean energy. We want to leave policymakers with three key takeaways to avoid falling into a firm power matching fallacy and to instead embrace the ability to mix and match resources to meet data center needs.

What began as a major strain on the grid can become the catalyst for building a smarter one, supporting both the digital economy's explosive growth and the clean energy transition.

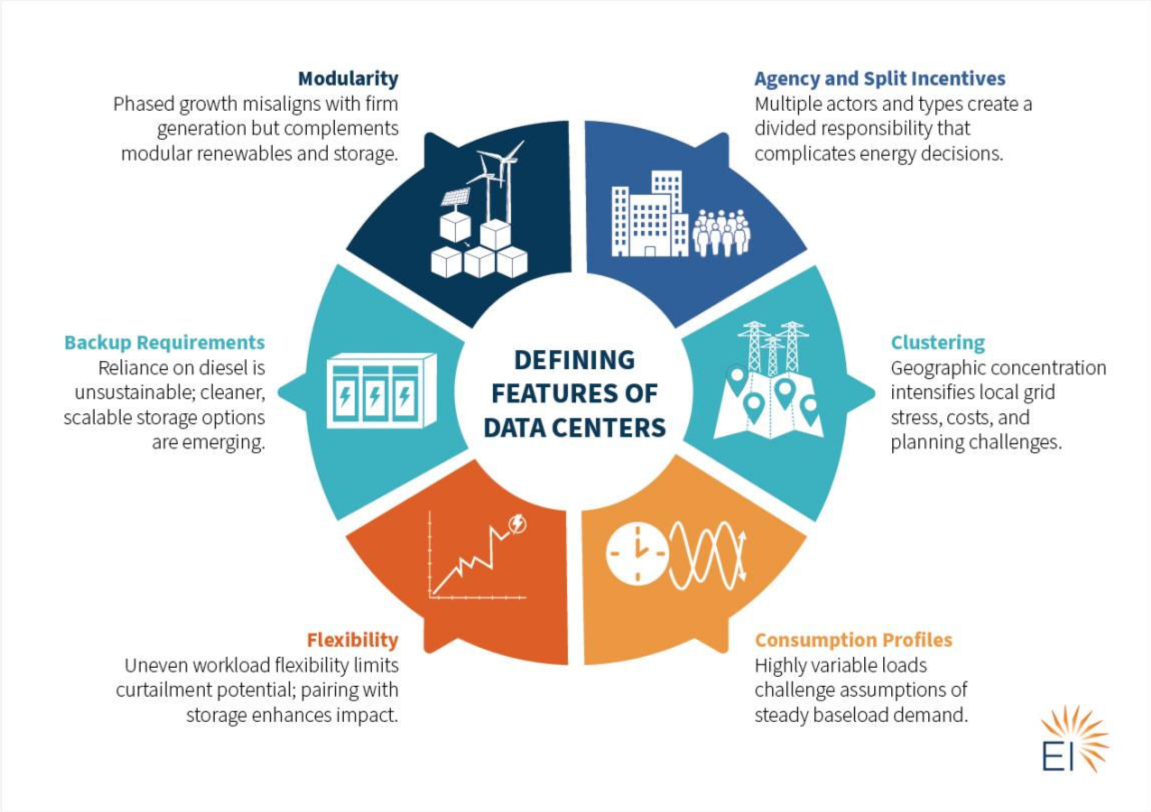
## COMMERCIAL AND INDUSTRIAL REALITIES THAT APPLY TO DATA CENTERS

**Actual data centers are not the simple "flat 24/7 block of demand" people imagine.**

Six different demand features of data centers explain the diversity of data center types (agency, clustering, and profile) and their internal workings (flexibility, backup, and modularity).

*Figure 2 Actual data centers are not the simple "flat 24/7 block of demand" people imagine.*

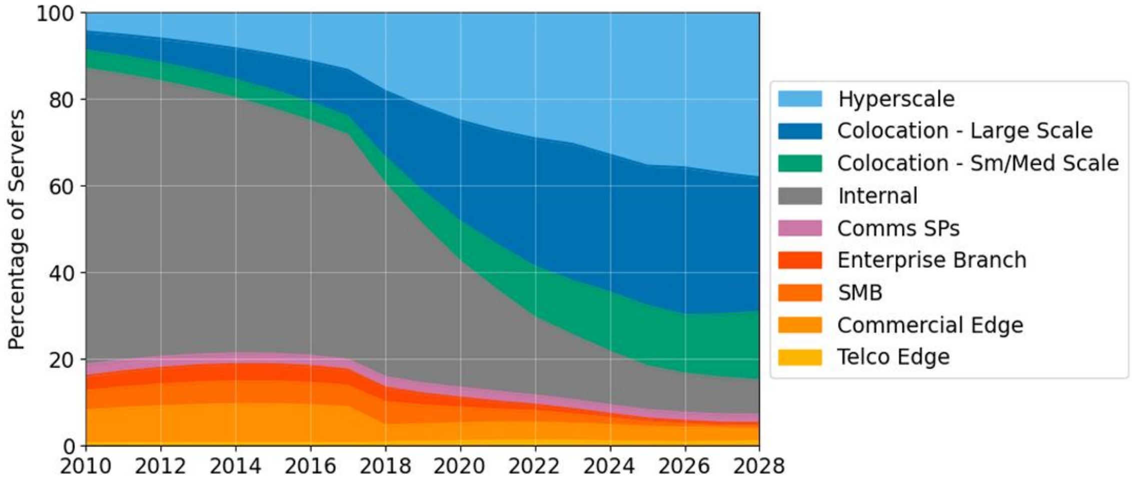



## Data Center Feature 1: Agency and the split-incentive problem

Planning *and* operating *a data center* involves many decision-makers. *Some data centers (often called co-locations or "colos") are facilities where customers can rent space to house their servers and equipment or just run their software on provided equipment. This means the facility is developed and owned by a different company* from those *that rent rack space, buy* computing *capacity, and ultimately consume electricity. Multiple actors force complicated decisions around electricity supply.*

If we want new data centers to adapt their development approach to better integrate with the grid and increase their "speed-to-power," policymakers must understand the planning, construction, and operation of modern data centers. Many different actors are involved, creating a classic split-incentive problem. Loosely speaking, apart from the users or clients, three groups of actors dictate the energy and resource impacts of data centers: developers, facility operators, and service providers. These tend to be separate entities. Overlap sometimes occurs, but usually not enough to prevent split-incentive issues. More than half (and an even larger fraction of the current pipeline)[6] of data centers are categorized as co-location facilities—large facilities that rent out space to multiple separate entities.

*Figure 3 Distribution of server types by data center type. 2024 United States Data Center Energy Usage Report[7]*




We illustrate the split incentives by cataloguing some of the key concerns for each of the three types of decision-makers in the life of a data center. In early stages, data center development is mostly a real estate bet: developers acquire land, water, and electric connection rights and then these rights pass on to the projects they sell. The natural incentive for developers is to keep the range of future owners they could sell to as wide as possible. Hence, they are unlikely to want to enter contracts or agreements (or support legislation) that might prematurely impair any of the land, water, and power consumption rights for their projects. For example, they may not want to agree to be a flexible consumer in return for faster interconnection (load interconnection currently takes three to 11 years) because that might scare off some prospective buyers.

Similarly, owner/operators that lease capacity to data centers customers do not necessarily have much insight into how flexible these customers are or how their customers' usage pattern might change over time. They are conservative about aspects such as whether the tenant-user would be interested in avoiding on-peak usage, participating in time-varying rates, accessing clean energy tariffs, or participating in a demand-response program. Obviously, renters must abide by some rules (via master service agreements or service-level agreements[i]) about behavior that impacts power quality (voltage, frequency, harmonics, transients, etc.) or broader

[i] A master service agreement is an umbrella standardized contractual framework between a utility and the "customer of record" (which could be a data center owner/operator, a tenant/end-customer, or a special purpose entity created to hold the contract) across multiple facilities in the utility's territory. A load serving agreement is more specific to power delivery at a given site.

electrical concerns (like grounding, interference, and surge protection), but that still leaves a lot of uncertainty for the data center owner/operator. Violations may also pass undetected until a severe problem occurs.

Because data centers are also large electricity consumers, utilities will want to know if contracts are backed by the ultimate users (e.g., hyperscalersii) or an intermediate company that could go bankrupt or disappear. Grid investments involve assets with multi-decadal lifetimes, while the service life of cutting-edge chips can be two to three years. Utilities and their regulators have a strong interest in recovering any incremental costs of investments needed to serve data centers and will look for contractual arrangements to make this happen.

### Data Center Feature 2: Clustering, data centers are attracted by similar conditions or to each other

*Data center locations tend to be concentrated in a few regions rather than evenly distributed. This clustering amplifies stress on already energy-dense grids. The main drivers are favorable conditions—reliable power, dense fiber, skilled workforce, tax regimes, and land—but anchor investments by hyperscalers or AI campuses could also accelerate the process. Policymakers should avoid treating projects as one-offs and consider the likelihood of a single facility snowballing into a larger cluster.*

"Clustering" describes how data centers in the U.S. tend to collect in a handful of regions rather than being evenly distributed. Clustering creates stress for the bulk power system because it takes already energy-dense loads and adds even more load nearby. The easiest explanation for clustering is that it derives from favorable existing conditions: reliable electricity, dense fiber connectivity, neighboring trained workforce, supportive tax regimes, and land availability.

Large anchor projects also draw in more data center development: Once a hyperscaler or AI training facility establishes itself, it signals viability, brings new infrastructure, and lowers costs for additional entrants. Policymakers wanting to provide support for a big project by promises of jobs and tax revenue, risk underestimating the impacts of this attractive force as welcoming one project may quickly lead to a cascade of follow-on facilities, with both outsized benefits and mounting strains.[8]

Recent history reveals a pattern whereby anchor investments amplify favorable local conditions into enduring centers of digital infrastructure. Northern Virginia's "Data Center Alley" grew from early fiber and internet exchange into the world's largest concentration of data centers. Amazon Web Services (AWS) was an early and steady

---

[ii] A hyperscaler is a cloud service provider or operator that builds and manages massive data center networks supporting millions of virtual servers and petabytes of data, operating globally and designed to scale seamlessly across regions. Examples might include Amazon Web Services (AWS), Microsoft Azure, Google Cloud Platform (GCP), Meta (Facebook), Apple, Alibaba Cloud, and Tencent Cloud.

investor in this cluster.[iii] Today, Data Center Alley reportedly handles roughly ~70 percent of the world's internet traffic, contains over 12 million square feet of commissioned data center space, and sustains hundreds of megawatts of power load.[9]. Reno's Tahoe-Reno Industrial Center became a global hub after Switch and Apple established major campuses, followed by Google and others[10]. Central Ohio offers a newer case: Google and AWS each invested in major builds, quickly attracting colocation providers.[11]. Atlanta and Phoenix look to be on similar paths[12].

In theory, diverse types of data centers should reinforce these patterns. Colocation facilities are drawn to network-dense hubs where they can maximize interconnection to other facilities. For example, enterprise servers might want to easily connect to multiple cloud providers—providers of cornerstone internet services stand to benefit from the reduced latency proximity affords, especially for content delivery like streaming video and games and so on. Hyperscalers could function as anchors, just like a department store in a shopping mall, investing billions into single campuses that create the vendor ecosystems others rely on. However, AI-focused facilities, with their unprecedented power needs, can also reshape the landscape by displacing other data centers competing for the same power network and generation resources.[13]

Electric power infrastructure both attracts and is stressed by clustering. Access to transmission lines and substations is a prerequisite, but as clusters grow, demand can overwhelm grids. Northern Virginia now faces multi-year waits for new hookups[14]. Reno's growth has raised water concerns and left Nevada utilities facing a potential doubling in necessary electrical infrastructure (also spurring them toward large renewable additions)[15]. Ohio illustrates the stakes most vividly: By March 2023, the utility AEP Ohio imposed a moratorium on new data center service agreements in Central Ohio, pending further study citing grid strain. Eventually regulators approved a new tariff[16] requiring data centers to pay for 85 percent of subscribed capacity whether it is used or not, with penalties for cancellation or under-performance and a four-year on-ramp[iv]. Clustering behavior can easily outrun planning and force regulators into reactive steps, introducing delays before more pro-active policies and tariffs can be put in place.

The policy lesson is not to avoid clusters—after all, they bring new jobs, tax revenue, and digital infrastructure—but to keep a skeptical eye on benefits claimed by developers and focus on smart planning. This should consider the multiple interests of stakeholders affected by a data center cluster and work in advance to align land use,

---

[iii] AWS is certainly not the only part of this story but has been called out as a major player. Dan Swinhoe, "The Amazon Factor in Virginia," Data Center Dynamics, November 6, 2024, https://www.datacenterdynamics.com/en/analysis/the-amazon-factor-in-virginia/. Amazon also touts its $51.9 billion investment in Virginia between 2011 and 2021 (capital + operations) in its data center infrastructure in Fairfax, Loudoun, and Prince William counties. Roger Wehner, "Learn About AWS's Long-Term Commitment to Virginia," Amazon, June 7, 2023, https://www.aboutamazon.com/news/aws/aws-commitment-to-virginia.

[iv] Under the decision, new data centers can access up to 50 percent capacity in the first year, 65 percent in the second, 80 percent in the third, and 90 percent in the fourth before getting full access to the grid.

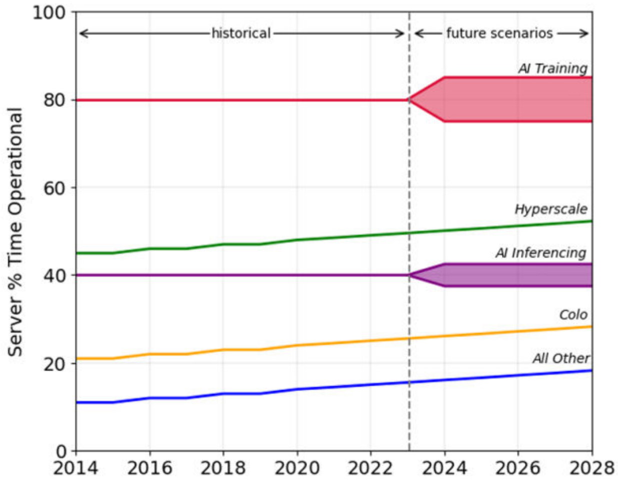
 

grid upgrades, generation, flexible loads, and permitting frameworks, ensuring that benefits can be captured without bottlenecks or backlash once clusters grow.

### Data Center Feature 3: Consumption profile

*Data center electricity usage is not steady or 24/7. Up close, it can be quite choppy and challenging. Batteries could act as a buffer—a keystone solution to managing power quality.*

*Figure 4 Server utilization by data center type. 2024 United States Data Center Energy Usage Report (LBNL).*

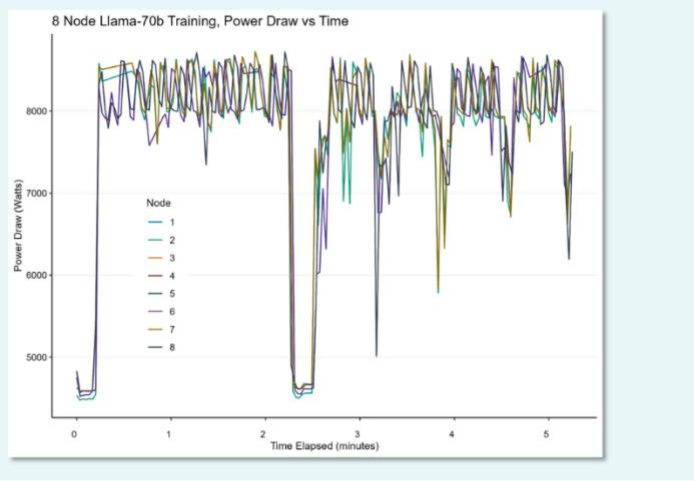



Data centers exhibit considerable variability, especially going between operational and idle states. In Lawrence Berkeley National Laboratory's (LBNL) 2024 United States Data Center Energy Usage Report, the authors explain that in 2014 colos reported 21 percent utilization rates, hyperscalers 45 percent – rising to an estimated 35 percent and 50 percent respectively in 2027. The same report models AI learning centers and AI inferencing at 80 percent and 40 percent utilization rates, respectively. These don't directly translate into electricity consumption load factors because some electricity is used for other purposes like cooling that don't follow a 1:1 relationship with computing load.

Even looking at the whole power consumption profile of a data center, it's important to differentiate between actual load factor (the percentage of possible 24/7 full power use that a data center in fact uses) and availability (the percentage of maximum power a data center expects to have if it wants it, i.e., the *option* to use power). Whether power comes from on-site generation or from the grid, it needs to be prepared to provide power when the data center wants it, and back off when the data center doesn't.

*Figure 5 8 Node trace of power consumption in an AI learning cluster. 2024 United States Data Center Energy Usage Report (LBNL).*




Big swings in data center demand will clearly be a challenge, even for the most flexible on-site generation. Given the scale at which many data centers operate, these swings can still create problems for large regional grids[v]. The CEO of Hitachi Energy reportedly commented "there can be swings of 200, 300 MW within a ten-minute period as data centers move from learn vs stop learn mode, and that these types of swings would not be acceptable from other grid customers."

At smaller time scales, large numbers of similar chips in one place switch on and off and can create an aggregate resonance effect[vi]. Existing electrical standards are inadequate for screening out these behaviors, and utilities may not have sufficient sensors to properly trace back issues to a particular data center. In aggregate, the evidence points to data centers deteriorating power quality metrics in their environs.[17]

More research needs to be done that focuses on new large digital loads, including variable generation resources with inverters that center around things like low-voltage ride through or fault clearing. For more information, context, and solutions on some of the challenges with interconnecting these large loads, see GridLab's recent Practical Guidance and Considerations for Large Load Interconnections.

Data centers are not a "perfect baseload" fit to directly couple with large mechanical generators or even the grid, and they will need significant electrical equipment to buffer this connection and prevent extra wear and tear on co-located generation or nearby grid users. Even if some data centers can learn to be flexible, incorporating battery energy storage, especially as the hardware cost decreases, will likely become a key element in managing data center impacts on the grid. When good wind and solar resources are available nearby, batteries can play a dual role in managing both load and generation variability at multiple time scales. Consider the Lancium Clean Campus in under construction in Abilene, Texas: "In addition to the 1.2 GW grid interconnection,

---

[v] See the GridLab report *Practical Guidance and Considerations for Large Load Interconnections*, with special attention to July 2024 Northern Virginia Data Center Event called out in Figure 1.2.

[vi] Some of these resonance issues can potentially be solved by on-chip energy management and storage. Rouslan Dimitrov et al., "How New GB300 NVL72 Features Provide Steady Power for AI," Nvidia, July 28, 2025, https://developer.nvidia.com/blog/how-new-gb300-nvl72-features-provide-steady-power-for-ai/?utm.

Lancium's power plan for the site includes large-scale behind-the-meter battery storage and solar resources, which serve to ensure grid reliability, and economic and carbon optimization."[18]

### Data Center Feature 4: Flexibility, or the lack thereof

*Flexibility could be key to quickly connecting new data centers, especially those involved with AI learning. Managed demand is possible, but on-site batteries may be a better solution where split incentives or onsite needs make demand control too rigid or complex.*

Data centers can be flexible, but different functions involve different levels of flexibility. This is probably hardest to achieve for co-location data centers because the third-party owner which interfaces with the grid and with utilities is not the one deciding what servers inside its facility are doing. Additionally, data centers are tasked with fluctuating sets of applications, creating uncertainty about how reliable or persistent demand management can be as a means of providing flexibility.

Data centers fully owned by large hyperscalers provide a higher degree of control over the whole facility. But the diversity of services being provided, often with low latency (response times) needs, may create constraints on what the hyperscaler can do. Hyperscale data centers provide both regular services—like AWS' cloud computing—and AI workloads such as inference, which involves answering client queries using pre-processed AI models.

For AI learning data centers, which create these large learning models, the goal is to cram as many chips as possible into the same square mile with the fastest internal connectivity so that the collection can operate as one big parallel machine[19]. Much of the possible flexibility here comes from adjusting the timing of computing batches, yet matching these adjustments to power supply flexibility needs is not a given, especially when considering that data center operators will want to prioritize computation over flexibility. This is a consequence of the relatively larger size of the capital investment in computing hardware versus energy generation and distribution for most applications.

Flexibility is a particularly important quality for data centers because they are such a large component of load growth, and just a little flexibility would reduce the need for new peaking resources and speed up interconnection[20]. A 2025 analysis[21] by the Nicholas Institute for Energy, Environment & Sustainability at Duke University finds that just 0.5 percent to 1 percent flexibility opens significant space on the grid: 98 GW of new load could be integrated at an average annual load curtailment rate of 0.5 percent, and 126 GW at a rate of one percent. This level of flexibility is similar to what is provided by demand-response programs that exist today for other loads, but as far as speeding up interconnection, it may be the AI-driven hyperscalers and learning centers, acting more directly under their owners' control and schedules, that can achieve more.

AI loads are fundamentally more flexible than generic data center loads because they can be processed in batches, easily scheduled, and often internally orchestrated. For example, in a presentation[22] to the Texas grid operator Electric Reliability Council of Texas (ERCOT), the company Emerald AI demonstrated how it could implement flexibility at a data center. The company argued there is enormous potential to control AI data center load, and that "major hyperscalers are amenable to curtailing up to 25 percent for up to 200 hours in return for priority interconnection of 1 GW."

No one knows if any particular data center's operations will remain stable enough to guarantee a given level of flexibility or willingness to curtail over the lifetime of matching local grid upgrades. In some cases, the data center load can be flexible (willing to forgo some batches of work) but not exactly in the way that best serves the local grid. Some amount of local battery energy storage (providing multiple value streams like integrating local on-site variable energy, backup, and power quality services) could also help data centers be more flexible at their grid interface, especially those with less direct control over internal processes.

### Data Center Feature 5: Backup needed for disturbances and outages

*Most data centers require backup. Demand flexibility and short-duration batteries can either eliminate or lighten the load for traditional backup solutions.*

Many data center customers aspire to high availability—as much as 99.999 percent uptime—hence the need for backup power to take over in case of any grid failure. The Uptime Institute, a widely followed source for industry tier certification in data center design, build, and operations,[23] defines four reliability tiers (I through IV) with increasing expectations for performance under challenging conditions, with an eye towards worst-case scenario planning. Many data centers serving enterprise needs require at least a Tier III level of reliability, either because of a direct need, like maintaining accessibility to data under adverse conditions, or as a proxy for operational trustworthiness. For mission-critical operations—major banks, stock exchanges, the military, or hyperscalers serving global customers—a Tier IV level of availability may be required.

Because Tier III and Tier IV facilities require 72 and 96 hours of on-site power capacity, respectively, simple economics dictate that backup is usually in the form of diesel generators with fuel storage on-site. Batteries can also be used to help ride-through disturbances in power supply,[vii] providing faster response times and reducing fuel and maintenance expenses on diesel. However, with today's technology, battery energy storage systems (BESS) that can cover critical needs for three to four days are not

[vii] In current facilities, this ride-through comes via the uninterrupted power system (UPS) usually provided by old-school lead-acid batteries, but modern lithium-ion battery energy systems can provide these services along-side the bulk of backup power needs.

economically feasible, especially without some form of on-site generation to sustain their state of charge.[viii]

However, diesel does not scale well: As data centers get much larger, massive tank farms for the generators' on-site fuel require complex fire protection, spill containment, and environmental risk mitigation. Furthermore, many air districts (e.g., Virginia, California, or Oregon) place strict caps on generator run time and cumulative emissions in a site or region. Placing more than a hundred diesel generators on one site creates a cumulative permitting challenge and may well face serious local resistance along with the prospect of delays or outright rejection from regulators. Somewhat cleaner gas generators (turbines[ix] or reciprocating engines) are usually connected to a pipeline and require large propane or liquefied natural gas storage facilities to satisfy on-site capacity requirements.

Some large hyperscalers are opting to target better up-time based on statistical estimates rather than explicit proxies for reliability. For example, Microsoft has publicly committed to reducing the use of diesel generators by 2030. To that end, it contracted with Saft, a subsidiary of TotalEnergies, to install four battery energy storage systems, each in groups of four megawatt hours (MWh) and capable of 80 minutes of on-site power, to replace diesel backup.[24] In the U.S., Microsoft's newest Azure region in San Jose, California is also being built diesel-free, but is using natural gas turbines for backup (plus batteries for ride-through). In general, the U.S. grid is quite reliable, with the one-in-ten reliability standard[x] mostly achieved at the transmission service level.[xi] Most outages that do occur are less than one or two hours, so a battery can carry enough of the backup burden to get the facility to a high level of reliability while hardly, if ever, using on-site generation.

As longer-duration storage solutions like Form's 100-hour battery[25] or thermal batteries[26] connected to local renewables and steam turbines in local energy parks[27] emerge, data centers will be able to free themselves from fossil fuel backups while taking advantage of integrated design to combine multiple uses of batteries for flexibility, power quality, and backup.

---

[viii] To see how this is done in detail, see the NREL Vulcan platform demonstration in collaboration with Verrus. Deepthi Vaidhynathan et al., "Vulcan Test Platform: Demonstrating the Data Center as a Flexible Grid Asset" (National Renewable Energy Laboratory, June 2025), https://www.nrel.gov/docs/fy25osti/94844.pdf.

[ix] Although gas turbines face significant supply chain cost and delivery challenges currently. GridLab, *Gas Turbine Cost Report*, https://gridlab.org/gas-turbine-cost-report/.

[x] The one-in-ten reliability standard is a standard that applies for the bulk power system (i.e. transmission level) requiring transmission planners, system operators and reliability planners to aim for nor more than one "event" of involuntary load-shedding in ten years. If one "event" was 24 hours, that is already 99.97 percent up-time.

[xi] Actual recent figures for grid performance are quite good (see table 1.1). North American Electric Reliability Corporation, *2024 State of Reliability*, June 2024, https://www.nerc.com/pa/RAPA/PA/Performance%20Analysis%20DL/NERC_SOR_2024_Technical_Assessment.pdf.

### Data Center Feature 6: Modularity—data centers are built in phases

*Data centers expand in discrete phases, from racks to halls to entire campuses, with uncertain demand and rapidly rising power density. This modular growth pattern matches well with the modularity of renewables-plus-batteries deployment, which can be built in parallel to meet incremental load without the risks of lumpy firm power investments.*

For utilities and data center developers, timing capital investments can be challenging. Matching these investments with energy supply for their increasingly electric power sub-components compounds the challenge. Building a new data center means committing to constructing a large building and grid capacity without knowing if consumers will come, how quickly they will deploy, or how their consumption will evolve over time. Tenants in a co-location situation, hyperscalers, and AI data centers may not immediately have all the chips available (or face some other bottleneck) so may want to deploy in phases: slowly building up electrical demand over time until reaching full capacity, if all the anticipated demand materializes. With a chip service life of around two to three years, the balance between increased efficiency and increased computing power may mean newer chips could either increase or decrease electricity demand in each physical asset footprint.[xii]

The digital world's infrastructure is itself modular—built from discrete, substitutable units. Data centers are not just abstract systems of bytes and tokens; they are also collections of tangible components: chips, servers, and, above all, racks. The rack is the main unit of reference: a cabinet holding multiple slender servers or "rack units." Racks are grouped into "pods" of 20–30, and an enterprise client might deploy a couple of pods at a time in either a dedicated or co-location facility. Some tenants lease only a handful of racks in a shared space, while hyperscalers may build entire halls of 200–400 racks, with multiple halls forming a single phase of expansion on a large campus[28].

The modular nature of data centers lets developers manage financial risk by building in phases, with the option to add new capacity quickly but in a planned way. Each phase, however, carries high stakes not only in capital cost but also in power demand. A 2024 Uptime Institute report[29], states finds that four- to six-kilowatt (kW) racks remain common, with a trend towards higher consumption today. Meanwhile, AI applications and high- performance computing are pushing the development of liquid-cooled racks with incredible increases in power density. Vertiv, an Ohio-based company that designs, manufactures, and services critical infrastructure for data centers, reported in its 2024 Investor Event Presentation[30] that extreme rack densities already reach 250 kW

---

[xii] This is certainly a question in flux. Google has reported that over a recent 12-month period, the energy footprint of the Median Gemini Apps text prompt dropped by 33x! At a given facility this can be achieved by increased throughput or reduced energy use, or both. Amin Vahdat and Jeff Dean, "Measuring the Environmental Impact of AI Inference," Google Cloud, August 21, 2025, https://cloud.google.com/blog/products/infrastructure/measuring-the-environmental-impact-of-ai-inference.

per rack today and could exceed one MW within five years. That means a space the size of a bedroom closet could consume more power than a thousand average homes. As a result, a single phase of development for a data center might range from on the low end at 250 kW (two-dozen at 10–12 kW per rack) on the low end to 250 MW at the high end (a 1,000 liquid-cooled 250 kW racks) at the high end, with extra overhead for cooling.

The extreme end of data center development is exemplified by data center developer Vantage's recently announced plans [31] to build its $25 billion Frontier campus situated on 1,200 acres in Shackleford County, Texas, with an eventual total consumption of 1.4 GWs—close to average total consumptions of the states of either Rhode Island or Delaware. And this project is not alone: a September 2025 ERCOT staff report [32] to ERCOT's board details 130 GW of non-crypto data center load in the interconnection queue through 2030.[33] In the last few years, Texas has met new additional load with new, mostly clean generation. Of the 428 GWs of generation requests as of August 31, 2025, 204 GWs are for wind and solar and 180 GWs are for energy storage (together 90 percent of all requests).

Data center development may come in all levels of power consumption. However, because developers rarely build, install, and commission data centers in a single phase, projects of all sizes need a power supply that can grow and expand with them. When covering the incremental energy demand from a new data center, a large new single firm resource is an unwieldy indivisible capital investment. A modular approach with renewables plus batteries reduces risk and provides better economics: You're not committing to a single lump-sum investment in a 500 MW gas turbine; you can phase investments, optimize based on real usage, and spread spending—and risk—over time.

With computing loads that grow unevenly, modular investments let operators respond dynamically—deploy more solar, wind, or storage as AI racks come online. As a bonus, you can avoid supply chain bottlenecks because incremental installation bypasses the big lead times and equipment backlogs associated with large generator orders, enabling continuous expansion without project delays. Just as data centers grow in discrete steps, modular renewables and batteries let the grid grow in parallel.

These six features highlight why data center demand is complex, not just a flat, 24/7 block of constant load. We now turn to how supply options can, and cannot, match this demand.

# THE BEST WAY TO MEET DATA CENTER DEMAND IS DIVERSE RESOURCE PORTFOLIOS

When thinking about how to supply new demand from the rapidly growing data center industry, the key point to remember is **one-to-one matching with "firm" resources will not "solve" the load growth needs from data** centers.

In this section, we explain why single, stand-alone generation resource matching for any given industrial load has rarely been the historical course, and how and why that might change. We then describe the three resource buckets that new data center projects need to acquire to use the existing bulk power system. Finally, we discuss how the data center demand features described in the preceding section create further challenges and barriers in acquiring these resources.

## Debunking the one-to-one matching myth

If you imagine data centers as large capital assets running power through expensive electronics 24/7, it seems natural to imagine a dedicated "captive" 24/7 power plant built to match this demand, with historical precedent for this one-to-one matching. For example, in the post-war era aluminum producer Alcoa built smelters near cheap grid sources of hydropower in New York and the Pacific Northwest along with captive coal plants in Indiana and Texas to feed the company's aluminum smelters and mills. Today, industrial facilities use on-site combined-heat-and-power (CHP) plants to consume both the electricity and waste heat from fuel-driven power plants to operate industrial facilities with high end-use efficiency, and thus lower energy costs. According to the U.S. Energy Information Administration's (EIA) latest Manufacturing Energy Consumption Survey from 2022, U.S. manufacturers produce around 17 percent of their electricity needs on-site (Table 11.1) and that on-site generation is 97 percent co-generation (Table 11.3).[xiii]

### Single plants may not "play nice" with data centers

The demand characteristics of data centers described in the prior section raise immediate concerns regarding matching a captive plant with a data center. For example, while a data center may want 24/7 availability, its actual consumption will ramp up and down significantly with a profile that a large, single on-site generator might struggle to meet. Many fossil generators have a minimum dispatch level they cannot fall below, and "ramp rates" limits dictate how quickly they can adjust up and down. Furthermore, a modular, phased build-out does not lend itself to a single matching resource because in order to provide sufficient power for the full buildout,

[xiii] This survey defines co-generation as "the production of electrical energy and another form of useful energy, such as heat or steam, through the sequential use of energy. Cogeneration includes electricity generated from fossil fuels, such as natural gas, fuel oils, and coal; wood; and other biomass." In practice, the steam/heat is the main other energy output, so co-generation is often used as synonymous with CHP.

the single resource would have to operate at lower, inefficient, dispatch levels during earlier phases of data center construction and operation.

Beyond a mismatch with the demand characteristics of data centers described in the prior section, there are additional reasons to question using a captive plant as a 1-1 match for a data center.

### Captive power plants are not highly reliable alone

Table 4 from the North American Electric Reliability Corporation (NERC) 2024 State of Reliability Overview [34], shows the recent weighted forced outage rate (rate of unexpected failure) was 11.7 percent for coal, 7.7 percent for gas, 6.4 percent for hydro, and 2 percent for nuclear. Another relevant consideration is planned maintenance, like cleaning out coal boilers, maintaining and inspecting gas turbines, or refueling nuclear plants every 18-24 months.[xiv] This means a single supposedly "firm" plant will be unavailable for a double-digit percentage of time—not what data centers are looking for.

If an industry is set on self-supply, one strategy is to over-supply generation. Alcoa's Warrick, Indiana aluminum smelter and mill built three captive 144 MW coal plants alongside a 300 MW coal unit shared 50/50 with the local utility Vectren. With a total capacity of 732 MW but serving a local load of 550 MW,[35] the facility was clearly resilient to losing one unit and still running. But this effectively meant carrying 25 percent more capacity than necessary, without a guarantee of full reliability. Alcoa mitigated this extra cost by selling excess power to the grid and importing power from Unit 4 or the broader grid when necessary. This illustrates the general case that a grid connection remains both a sink for surplus and an important backup option; most on-site power is not fully independent and large loads will still want interconnection to the bulk power system. In fact, payment for grid backup (usually called "standby rates") is a common feature of CHP tariffs.[36].

### What happens when the power plant is no longer needed?

An interesting postscript to the Alcoa Warrick plant story is that Alcoa announced it would shut down its aluminum smelter in 2016 (although it had partial restarts post 2018) because of poor market conditions [37] and transferred major rolling mill and finishing operations to Kaiser Aluminum in 2021.[38] It is now left with an unattractive coal generation asset, whose generation capacity now exceeds Alcoa's local demand and

---

[xiv] This is on average a 32-day process. Aaron Larson, "Planning Is Key to Successful Nuclear Refueling Outages," POWER Magazine, September 1, 2023, https://www.powermag.com/planning-is-key-to-successful-nuclear-refueling-outages/.

will likely struggle to sell surplus capacity in the broader power market along with most coal assets,[39] which comes with significant environmental remediation liabilities.[xv]

This is always the risk with a captive power plant: One day the load will vanish because of changing economics. Investors will want to know if Plan B exists and that the captive plant is in and of itself an attractive asset with a bright future.

### Co-location of prime power generation assets with data centers today

Grid bottlenecks create considerable talk about co-locating "prime power" generation[xvi] with new data centers. This might include leveraging existing nearby assets (for example, the Talen-Susquehanna deal which co-locates a data center next to a nuclear plant[40]), restarting mothballed generators, or building new on-site resources (such as gas plants). But these arrangements are not true one-to-one matches of generation with load, since they still depend heavily on a grid connection for full functionality, sometimes at the expense of other consumers.

For example, in the Talen-Susquehanna deal, the data center is physically adjacent to a pair of nuclear units. It is unlikely the units' output will ramp precisely in step with data center consumption. Therefore, matching local supply with demand creates net output—nuclear output minus on-site data center consumption—variability which must be managed by the grid operator. During the refueling of one nuclear unit, the other must pick up the data center load, thereby reducing exports to the grid. In effect, the grid acts as backup.

Hybrid arrangements using on-site power together with the grid can address bottlenecks. They combine physical and financial hedges. The Talen-Susquehanna deal, for instance, was eventually reshaped into a power purchase agreement after regulatory push-back.[41]. These hybrid deals share many of the properties—and many of the drawbacks—of other on-site generation deals discussed above.

If land is available, the best way to provide on-site prime power is not with a single firm resource but by using an energy park[42] with renewables and batteries, with backup from longer- duration storage and/or gas generators.[43] Then most of the generation is

---

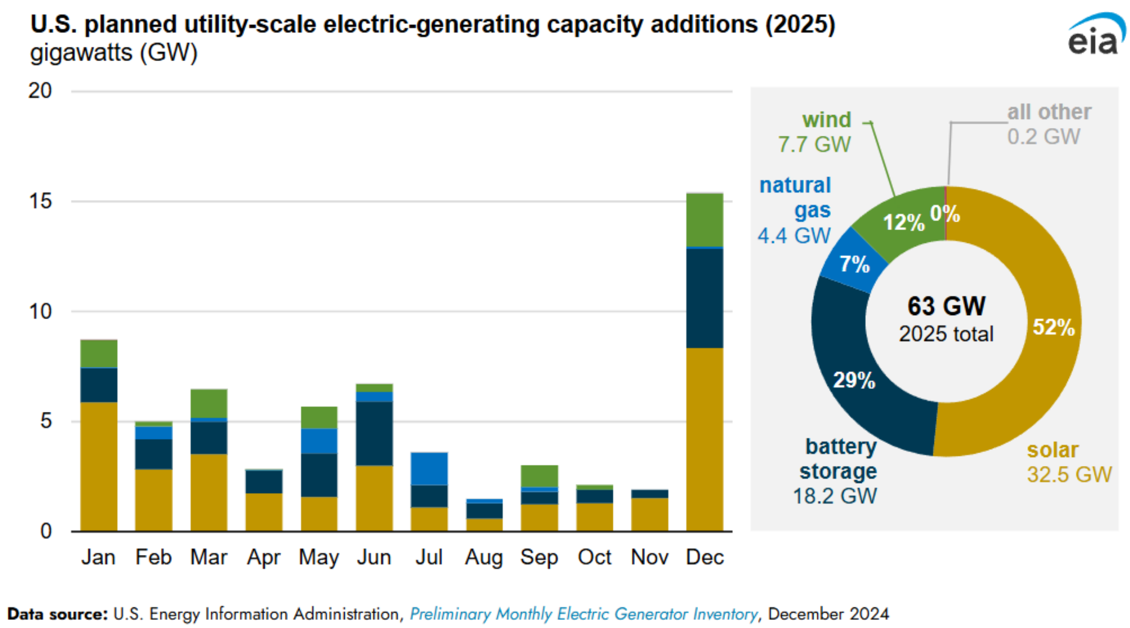
In 2024, Sierra Club and Environmental Integrity Project intervened in a case against Alcoa Warrick for "100+ permit violations" in 2022 and 2023, including releases of mercury, aluminum, chlorine, copper, fluoride, nickel, and zinc into the Ohio River. Environmental Integrity Project, "Groups Intervene in State Action to Stop Aluminum Smelting Plant's Illegal Dumping of Heavy Metals in Ohio River," February 21, 2024, https://environmentalintegrity.org/news/groups-intervene-in-state-action-to-stop-aluminum-smelting-plants-illegal-dumping-of-heavy-metals-in-ohio-river. There are also lingering questions regarding compliance for Warrick's "ash ponds" under the Environmental Protection Agency's Coal Combustion Residuals rules. Hoosier Environmental Council, "Today's EPA Action Means More Coal Ash Cleanup for Indiana" (press release), April 25, 2024, https://www.hecweb.org/wp-content/uploads/2024/04/PRESS-RELEASE-Todays-EPA-Action-Means-More-Coal-Ash-Cleanup-for-Indiana.pdf.

[xvi] Prime power is the continuous everyday power that powers the data center, as opposed to backup power There is also "bridge power" which is local generation which acts as prime power until a grid connection is put in place, and then becomes either backup or just a part of the supply portfolio.

clean, faster, and cheaper to deploy than other generation and helpful for hyperscaler emissions commitments. In addition, the combined resource is more reliable, with fewer large points of failure than a handful of fossil units – a good Plan B if load never fully materializes or decreases. Included energy park resources would reflect a microcosm of trends in the wider U.S. market where the large majority of generation coming online[44] and waiting in interconnection queues are renewables and batteries.[45]

*Figure 6 Solar, battery storage to lead new U.S. generating capacity additions in 2025. US EIA.*




## Providing new electricity supply for data centers from the bulk power grid

Given that most data centers will need to get some, if not all, of their power from the bulk power system, it is helpful to review how large commercial or industrial loads do this. The power grid is a system of systems including physical transmission and distribution poles and wires, the generation and loads they connect, operations and dispatch, power markets, and power purchase agreements.

As soon as a large new load decides to connect to the bulk power system, its needs can be disaggregated and met in many ways.

The grid resources a new data center project must collect to successfully draw from the bulk power system fall into three broad buckets: connection, grid services, and bulk electricity.

## New large data centers will require connection and network upgrades

How data centers connect to the grid depends on their size: scale matters. Smaller enterprise and co-location data centers (tens of MW or less) will often connect to a distribution system's high-end network (i.e., somewhere between 13.8 and 69 kilovolts) and may tie into an existing distribution substation with a new feeder. The utility typically owns and operates the primary substation equipment, while the data center customer owns the step-down transformer to its facility. The local utility conducts the impact studies and plans local upgrades to ensure compliance with NERC standards. Too many connections in the same area may trigger transmission upgrades and inclusion in transmission planning studies. In some geographies, like Virginia, this may involve an independent system operator (ISO) such as PJM [xvii] in planning and approving upgrades.

Larger data center campuses will have their own complex internal grid that connects directly to a bulk power system transmission substation. The data center must file a large load interconnection request with the local transmission owner or ISO. Tariffs and agreements will include matters like covering study costs, equipment ownership, and who pays for upgrades. The state may also require approvals for siting, environmental review, and cost recovery. The connection process can become long and painstaking once local capacity on the grid becomes tight. In Virginia's Dominion utility territory, data centers larger than 100 MW face up to a seven- year wait for power hookups.[46]

One important feature of new connection costs is that they are usually covered by the new load because cost causality is clear. Unfortunately, this may not hold true for more upstream transmission impacts where transmission upgrade costs are traditionally socialized more widely. A recent Natural Resources Defense Council (NRDC) report[47] tells the story in PJM: "Tight supply conditions led PJM to approve a $5 billion transmission expansion project to meet new data center demand in Virginia, where data centers already account for around a quarter of the state's electricity demand. The costs for this project were distributed by the Federal Energy Regulatory Commission (FERC), PJM, and utilities using varying cost allocation methods. Maryland residential customers were left with a bill of approximately $330 million, and Virginia residents had to foot $1.25 billion for transmission designed largely for a handful of data center customers in only a small region of the state."

## Data centers create new stresses on a bulk power system planned around peak demand; they also consume other grid services

The main grid service data centers require regardless of size is peak capacity: the ability to serve up to their maximum interconnection rating during periods of system peak.

---

[xvii] Also referred to as a regional transmission operator, PJM covers 13 states in the mid-Atlantic and is one of the largest power markets in the world.

Going back to PJM (often a source of current examples because it already serves so many data centers), the ISO's board chair communicated about future reliability concerns because: "PJM's 2025 long-term load forecast shows a peak load growth of 32 GW from 2024 to 2030. Of this, approximately 30 GW is projected to be from data centers."[48]

PJM's conundrum is how to keep the grid reliable as data center demand grows faster than new generation. Its "non-capacity-backed load" proposal would classify very large new loads (less than 50 MW) as customers outside the capacity market.[49]. The idea is to avoid shifting costs to others, but critics say that the 50 MW cutoff is arbitrary, curtailment rules could distort market signals, and contract and siting decisions may be disrupted.[50]. PJM is still debating whether the non-capacity-backed load should be voluntary or mandatory in shortage zones before filing at FERC for the 2028/29 delivery year.[51]

One challenge with resources like peak capacity is that once a project has been approved for interconnection, been built, and paid its share of costs, it becomes a load like any other. At that point, it is very difficult for the market to discriminate against it without creating efficiency concerns or legal risks. Data centers do more than strain peak supply; like all large loads with some variation, they also draw on ancillary services and other grid management resources.

### If incremental demand is not met with increased supply, prices and emissions will rise

As the recent Nicholas Institute report[52] points out, some amount of flexibility from data centers could significantly reduce costs and delays associated with connection and peak demand constraints from new data centers. The report estimates peak load bottlenecks could be avoided for around 100 GW of so-called "curtailment-enabled headroom" on the U.S. grid. However, even if data centers avoid consumption during the most problematic hours, they still need power the rest of the time. Absent new supply on those same grids, the extra generation available off-peak will be from more expensive, and typically dirtier, marginal generation units.

Data centers' need to draw most of their power from existing units is thus a problem for other electricity customers because absent new matching supply, it will drive up their wholesale electricity costs. It is also problematic for the data centers themselves, which frequently are tied to corporations that have carbon reduction goals which are incompatible with increased emissions from existing fossil power plants. Conversely, new supply (especially cheap and clean supply) arriving quickly enough to offset data center consumption without requiring a large amount of new grid infrastructure creates potential for "beneficial electrification[53]" where more power over the same wires reduces other consumers' costs[54].

## Further consequences: Challenges and barriers specific to data centers

Connecting large new data center loads through the lens of three resource buckets faces three broad challenges required by all such loads. But these resource buckets also interact with the six more specific data center demand characteristics outlined in the preceding section.

### Connection challenges specific to data centers

Because the source of many connection issues—or at least more expensive upgrades—come down to a limited set of hours and circumstances, flexibility is often cited as a master key for easing or speeding up connection. But flexibility is not always as simple to implement as first imagined, and other connection challenges specific to data centers are not necessarily circumvented with a touch of flexibility.

- **Agency:** Especially for co-location data centers, the operator is stuck between wanting to be more flexible to satisfy grid constraints and the imperative to be as generic as possible in contracts with tenants to accommodate as broad a class of customers as possible. Typical quality of service and service-level agreements also act as a barrier for to tapping flexibility.[55] Intervenors in public utility cases also question whether policies for ensuring new data centers cover all their incremental costs are effective[56].
- **Clustering:** Clustering leads to many data centers on the same part of the grid, necessitating more upstream transmission upgrades, as in the Virginia case mentioned cited by NRDC, mentioned above.
- **Consumption profile:** Big swings in power demand and power quality impacts on other consumers make data centers trickier for utilities and transmission providers to study and interconnect than simple 24/7 constant loads. Standard protection schemes and the collective behavior of 60 data centers recently caused a large reliability problem in Virginia in July 2024 when these data centers all dropped off the grid at once and caused a sudden surge in excess electricity that strained grid resources.[57]
- **Flexibility:** Some data centers are not flexible at all; others could be flexible but not in a manner consistent or predictable enough to satisfy the engineers running interconnection studies. These engineers are only likely to be satisfied after adding sophisticated energy management systems and large batteries, along with the promise of judicious backup power.
- **Backup:** As mentioned, backup power could be leveraged to facilitate connection or provide so-called “bridge power”[58] for data centers that cannot wait for interconnection. Unfortunately, backup power (used either for

flexibility or bridge power) tends to be dirty, leading to siting and local community environmental concerns.[59][60]

- **Modularity:** With a modular or phased build-out, a data center may ask up front for a large enough connection to accommodate all future phases, leading to stranded asset risk if all phases do not materialize.

### Grid services challenges specific to data centers

Just as for solving connection issues, flexibility can help tamper the impact of new data centers on system-wide needs like peak capacity issues. A large overlap exists between local grid and larger grid issues with peak planning. However, as described in the prior section, flexibility is not always easy to implement or deploy in a manner which solves all challenges. Furthermore, the specific features of data centers tend to create additional challenges beyond help from simple load flexibility measures.

- **Agency:** Utilities often see new fossil resources like gas peakers as the easiest way to resolve new peak demand issues from data centers.[61] Because gas turbines are increasingly expensive, this may not be a good deal for other utility customers and may also entrench future emissions, working against many data center providers' and host states' clean power goals. Because eventual data center owner/operators tend to build where developers have prepared the ground, the fact that these developers may perceive emissions goals as secondary to "speed-to-power," and that utilities choose their own procurement path creates an agency mismatch.
- **Clustering:** The clustering of data centers tends to amplify their effects on the regional grid, with sharper surges in demand for grid services that cannot be accommodated fast enough through new resources builds.
- **Consumption profile:** While data centers don't run all the time, they plan their infrastructure for peak computing demand. This creates a knock-on effect for the bulk power system, which plans for peak power demand.
- **Flexibility:** Flexibility is not always a simple feature to implement or deploy.
- **Backup:** On-site backup power is a poor substitute for system resources because of expense as well as siting and local environmental concerns.
- **Modularity:** While the broader grid is in a good position to adjust to a phased build out of data center demand, this requires either coordination with the local utility or strong forward signals in the market to avoid disruptive demand shocks for grid services.

### Bulk electricity challenges specific to data centers

Certain subcategories here (consumption profile, flexibility, modularity) concern time domains that do not apply when considering total annual consumption.

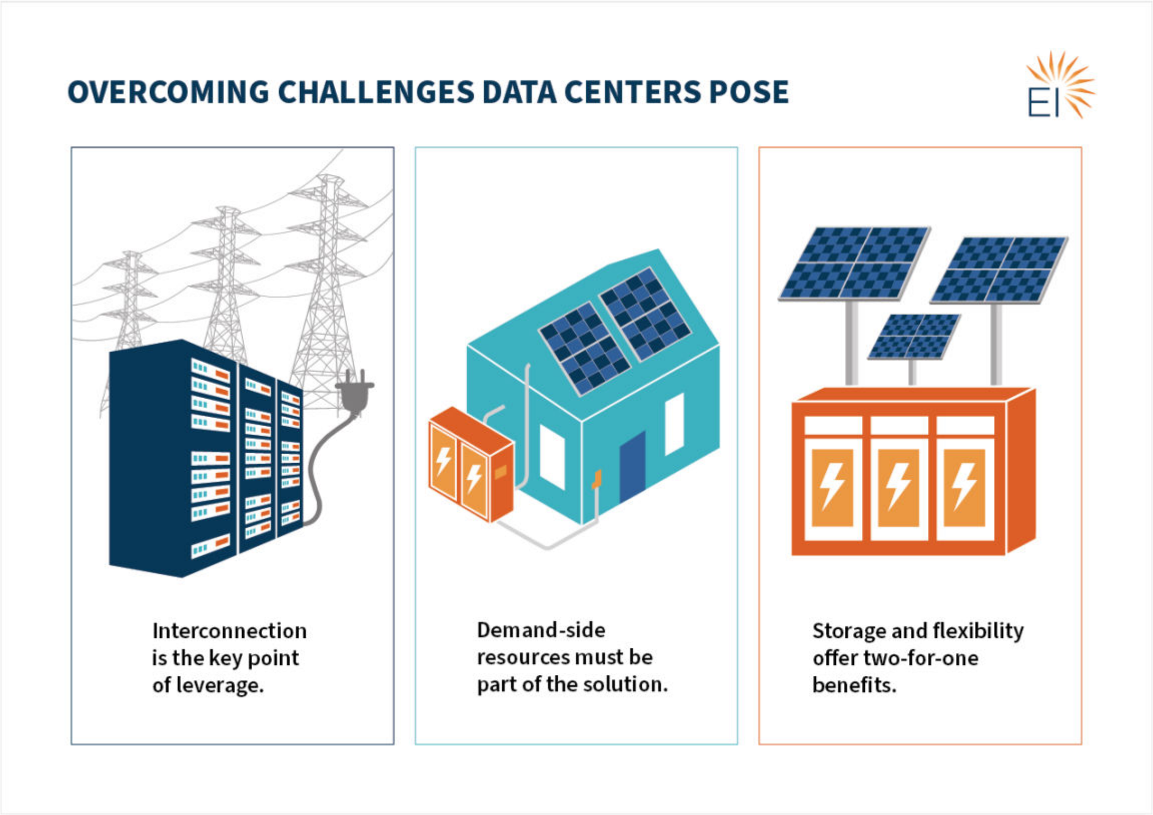
 

- **Agency:** As with grid connection and services, intermediate entities between electricity provisioners and data center owner/operators with emissions goals may not consider the environmental impacts of reliance on using spare capacity from existing marginal resources.
- **Clustering:** Clustering means more annual electricity drawn from the same grid. This creates a greater need for new supply, amplifying the problems of price and emissions increases.
- **Consumption profile:** A variable but not necessarily predictable profile for large data center loads could create new challenges for grid operators, even at off-peak times.
- **Flexibility:** Data center demand flexibility could potentially alleviate emission concerns by targeting off-peak consumption more towards time periods with lower marginal emissions. This requires an extra layer of control on top of whatever that data center might have already committed to ease grid connection and mitigate impacts on grid operational resources.
- **Backup:** Supplying too much of the actual annual electricity use from backup power creates problematic environmental impacts.

# TAKEAWAYS FOR POLICYMAKERS AND OTHER STAKEHOLDERS

*Figure 7 Three key takeaways*




Many organizations (Clean Air Task Force/Brattle,[62] GridLab,[63] NRDC,[64] the Bipartisan Policy Center,[65] and the Regulatory Assistance Project[66]) have provided detailed and useful guides for coping with the challenges of meeting new data center loads.

This section distills the key lessons from the features and challenges discussed above.

### Interconnection is the key point of leverage to influence when and how data centers join the grid

Accommodating large, dense new loads affects every grid participant, and the challenges show up at multiple scales. Geographically, they range from the substation where the data center connects to the entire interconnected system. In time, they span from sub-second transients to hours of local and bulk stress to the accumulation of annual demand.

When issues are tied directly to a data center's load or its immediate connection, cost-causation principles are easier to apply. But at larger scales, like meeting new annual demand or rising peaks across a region, the problem is less about the nature of data centers than the pace and size of their growth. At that point, they can reasonably argue for being treated like any other customer buying power "at the pump," without special obligations.

This tension is what policymakers need to keep in mind. Interconnection is the moment of maximum leverage: not to extract unreasonable concessions, but to ensure new entrants cover the infrastructure costs they trigger, and to nudge them toward implementing solutions like flexible demand or local storage that relieve local bottlenecks and support the broader grid. Likewise, developers and customers should lean toward local fixes that speed access to the grid, improve power quality, and ease broader impacts—reducing the likelihood of being saddled with extraordinary requirements later.

## Using other demand-side resources

In one of our earlier reports on meeting the load growth challenge,[67] we pointed out the importance of using demand-side resources to meet this challenge most efficiently. Often the discussion of demand-side solutions focuses on direct measures at a data center, especially in the wake of the efficiencies revealed in the DeepSeek announcement.[68] However, data centers can meet their resource needs with other demand-side resources elsewhere on the grid. Recently, Voltus, an aggregator of distributed energy resources (DERs), announced a deal with Cloverleaf Infrastructure—a data center developer—to meet new capacity needs from data centers with market-accredited capacity from DERs.[69] This kind of transaction compensates other existing customers and thus helps accommodate the rapid rise of data center loads fairly, speedily, and equitably. Since connecting to the grid can involve mixing and matching resources to relieve bottlenecks, once a data center has invested in the flexibility and extra equipment needed to resolve local connection issues, there is no reason why more upstream connection issues, grid services bottlenecks, and the need for a large amount of annual electricity delivery cannot be resolved with demand-side solutions from other grid users.

A recent Rewiring America report proposes many of the resources to meet data center load growth could come from sponsoring household upgrades.[70] The report finds that if hyperscalers paid 50 percent of the up-front cost of installing heat pumps in the tens of millions of U.S. households that currently use inefficient electric heating, cooling, and water heating, they could free up a total 30 GW of capacity on the grid. In addition, if hyperscalers paid 30 percent of the up-front cost of rooftop solar and storage in every single-family household in the U.S., they could add 109 GW of capacity on the grid. The cost of these upgrades would be comparable to the report's estimate of $315/kW-year to build and operate a new gas power plant.

### Storage and flexibility relieve data center challenges; they can also ease interconnection of new variable renewables

On-site prime generation solutions built around renewables and flexibility (modulating demand and using batteries) may provide cheaper, cleaner, and faster means for meeting new and existing data center demand. Because batteries are increasingly essential for buffering, backup, and power quality, they also provide a built-in solution for integrating variable renewables—offering a two-for-one advantage.

Furthermore, these renewable-plus-battery solutions can take advantage of existing surplus interconnection[71] to more quickly connect data centers to the grid in "power couples."[72]

By exploring the nuanced solutions, policymakers can avoid overcommitting to outdated firm resources and instead adopt strategies that embrace modularity, flexibility, and clean energy. Doing so will support both the digital economy's explosive growth and the clean energy transition.

# END NOTES

[1] Ana Boyd and Todd Olinsky-Paul, *Load Growth: What States Are Doing to Accommodate Increasing Electric Demand* (Clean Energy States Alliance, July 2, 2025), https://www.cesa.org/resource-library/resource/load-growth-what-states-are-doing/.
[2] Cushman & Wakefield, *2025 Global Data Center Market Comparison*, May 7, 2025, 31, https://www.cushmanwakefield.com/en/insights/global-data-center-market-comparison.
[3] Rewiring America. "Homegrown Energy: How Household Upgrades Can Meet 100 Percent of Data Center Demand Growth." September, 2025, https://www.rewiringamerica.org/research/homegrown-energy-report-ai-data-center-demand.
[4] Eric Gimon, Mike O'Boyle, and Michelle Solomon, *Meeting Growing Electricity Demand Without Gas* (Energy Innovation, March 28, 2024), https://energyinnovation.org/report/meeting-electricity-demand-without-gas/.
[5] Rewiring America. "*Homegrown Energy: How Household Upgrades Can Meet 100 Percent of Data Center Demand Growth*." September, 2025, https://www.rewiringamerica.org/research/homegrown-energy-report-ai-data-center-demand.
[6] Cushman & Wakefield, *2025 Global Data Center Market Comparison*, 3.
[7] Shehabi, A., Smith, S.J., Hubbard, A., Newkirk, A., Lei, N., Siddik, M.A.B., Holecek, B., Koomey, J., Masanet, E., Sartor, D. 2024. 2024 United States Data Center Energy Usage Report. Lawrence Berkeley National Laboratory, Berkeley, California. LBNL-2001637. Retrieved from https://eta-publications.lbl.gov/sites/default/files/2024-12/lbnl-2024-united-states-data-center-energy-usage-report.pdf
[8] Ginger Woolridge, "How Ashburn, VA Became the Colocation Mecca Known as Data Center Alley," *Lightyear* (blog), November 12, 2021, https://lightyear.ai/blogs/ashburn-colocation-data-center-alley.
[9] Woolridge, "How Ashburn, VA Became the Colocation Mecca Known as Data Center Alley."
[10] Data Center Frontier. *Reno Rising: Region Attracts Cluster of Cloud Campuses*. 2023. https://www.datacenterfrontier.com/cloud/article/11430825/reno-rising-region-attracts-cluster-of-cloud-campuses
[11] "Data Centers Are Booming in Ohio's Digital Heartland." Data Center Frontier, January 31, 2024. https://www.datacenterfrontier.com/site-selection/article/33035576/data-centers-are-booming-in-ohios-digital-heartland.
[12] CBRE Research. *Global Data Center Trends 2025: North American Inventory Growth Includes Phoenix*. June 2025. https://www.cbre.com/insights/reports/global-data-center-trends-2025

[13] "Hyperscalers Reshape the Colocation Pricing Landscape." Datacenters, July 17, 2024. https://www.datacenters.com/news/hyperscalers-reshape-the-colocation-pricing-landscape.
[14] "Data Centers Face Seven-Year Wait for Dominion Power Hookups." *Bloomberg*, August 29, 2024. https://www.bloomberg.com/news/articles/2024-08-29/data-centers-face-seven-year-wait-for-power-hookups-in-virginia.
[15]Alan Halaly, "Nevada's Data Center Boom Is a Power, Water Conundrum," *Las Vegas Review-Journal,* July 27, 2025, https://www.reviewjournal.com/local/local-nevada/nevadas-data-center-boom-is-a-power-water-conundrum-3403770/
[16] Sonal Patel, "Regulator Approves AEP Ohio's Landmark Data Center Tariff," POWER Magazine, July 10, 2025, https://www.powermag.com/regulator-approves-aep-ohios-landmark-data-center-tariff/.
[17] Leonardo Nicoletti, Naureen S. Malik, and Andre Tartar, "AI Needs So Much Power, It's Making Yours Worse," Bloomberg, December 27, 2024, https://www.bloomberg.com/graphics/2024-ai-power-home-appliances/
[18] "Crusoe Expands AI Data Center Campus in Abilene to 1.2 Gigawatts," Lancium press release, March 18, 2025, https://lancium.com/press-release/crusoe-expands-ai-data-center-campus-in-abilene-to-1-2-gigawatts/
[19] AFL Hyperscale, *AI Data Centers: Scaling Up and Scaling Out,* white paper (December 2024), https://www.aflhyperscale.com/wp-content/uploads/2024/12/AI-Data-Centers-Scaling-Up-and-Scaling-Out-White-Paper.pdf
[20] GridLab, *Data Center Flexibility: NV Energy Case Study Report*, accessed October 16, 2025, https://gridlab.org/portfolio-item/data-center-flexibility-nv-energy-case-study-report/
[21] Tyler H. Norris, Tim Profeta, Dalia Patino-Echeverri, and Adam Cowie-Haskell, *Rethinking Load Growth: Assessing the Potential for Integration of Large Flexible Loads in US Power Systems* (Durham, NC: Nicholas Institute for Energy, Environment & Sustainability, Duke University, February 11, 2025), https://nicholasinstitute.duke.edu/publications/rethinking-load-growth.
[22] Emerald AI, *Emerald AI: ERCOT Large Flexible Load Task Force Discussion* (March 28, 2025), https://www.ercot.com/files/docs/2025/03/27/Emerald-AI-LFLTF-Background-for-Circulation-March-28-2025.pdf.
[23] "Data Center Tiers: Everything You Need to Know," Sunbird DCIM (Sunbird Software, Inc.), September 26, 2022, https://www.sunbirddcim.com/blog/data-center-tiers-everything-you-need-know
[24] Rod Walton, "Microsoft Replaces Diesel Generators with Saft Batteries at Swedish Data Center," *Microgrid Knowledge*, October 4, 2023, https://www.microgridknowledge.com/energy-storage-microgrid/article/33012640/microsoft-replaces-diesel-generators-with-saft-batteries-at-swedish-data-center/
[25] "Energy Storage for a Better World," Form Energy (website), Form Energy, Inc., accessed September 23, 2025, https://formenergy.com/
[26] Jeffrey Rissman and Eric Gimon, *Thermal Batteries: Decarbonizing U.S. Industry While Supporting a High-Renewables Grid* (Energy Innovation Policy & Technology LLC, July 12, 2023), https://energyinnovation.org/report/thermal-batteries-decarbonizing-u-s-industry-while-supporting-a-high-renewables-grid/.
[27] Eric Gimon, Mark Ahlstrom, and Mike O'Boyle, *Energy Parks: A New Strategy to Meet Rising Electricity Demand* (Energy Innovation, December 9, 2024),

https://energyinnovation.org/report/energy-parks-a-new-strategy-to-meet-rising-electricity-demand/.

[28] Brian Potter, "How to Build an AI Data Center," *Construction Physics*, June 10, 2024, https://www.construction-physics.com/p/how-to-build-an-ai-data-center

[29] Douglas Donnellan, Andy Lawrence, Daniel Bizo, Peter Judge, John O'Brien, Jacqueline Davis, Max Smolaks, Jabari Williams-George, and Rose Weinschenk, *Uptime Institute Global Data Center Survey 2024* (Uptime Institute Intelligence, July 2024), https://datacenter.uptimeinstitute.com/rs/711-RIA-145/images/2024.GlobalDataCenterSurvey.Report.pdf

[30] Vertiv, *Vertiv Investor Event* (Atlanta, GA, November 18, 2024), https://s23.q4cdn.com/959471387/files/doc_presentations/Vertiv-2024-Investor-Event-Presentation.pdf.

[31] Dan Swinhoe, "Vantage Plots 1.4GW Data Center Campus in Texas," *Data Center Dynamics*, August 19, 2025, https://www.datacenterdynamics.com/en/news/vantage-plots-14gw-data-center-campus-in-texas/

[32] Electric Reliability Council of Texas (ERCOT), *System Planning and Weatherization Update* (Board of Directors meeting presentation, September 15, 2025), https://www.ercot.com/files/docs/2025/09/15/16.2-System-Planning-and-Weatherization-Update.pdf.

[33] Electric Reliability Council of Texas (ERCOT), *System Planning and Weatherization Update* (Board of Directors meeting presentation), September 15, 2025, https://www.ercot.com/files/docs/2025/09/15/16.2-System-Planning-and-Weatherization-Update.pdf.

[34] North American Electric Reliability Corporation, *State of Reliability Overview: Assessment of 2023 Bulk Power System Performance*, 2024, https://www.nerc.com/pa/RAPA/PA/Performance%20Analysis%20DL/NERC_SOR_2024_Overview.pdf.

[35] U.S. Department of Energy, Oak Ridge National Laboratory, and Alcoa, *Alcoa & ORNL Demand Response White Paper*, January 2009, https://www.energy.gov/sites/prod/files/2015/04/f21/Alcoa%20%26%20ORNL%20Demand%20Response%20White%20Paper%20-%20Jan.%202009.pdf

[36] U.S. Environmental Protection Agency, *Utility CHP Standby Rates*, webinar presentation, June 2018, https://www.epa.gov/sites/production/files/2018-06/documents/utility_chp_standby_rates.pdf

[37] Alcoa. *Alcoa to Close Warrick Smelter and Curtail Remaining Capacity at Pt. Comfort Refinery*, Press Release, January 7, 2016. Available at: https://news.alcoa.com/press-releases/press-release-details/2016/Alcoa-to-Close-Warrick-Smelter-and-Curtail-Remaining-Capacity-at-Pt.-Comfort-Refinery/default.aspx.

[38] Kaiser Aluminum Corporation, "Kaiser Aluminum Corporation Reports Second Quarter and First Half 2021 Financial Results," July 21, 2021, accessed October 18, 2025, https://web.archive.org/web/20211023002708/https://investors.kaiseraluminum.com/news-releases/news-release-details/kaiser-aluminum-corporation-reports-second-quarter-and-first-8.

[39] Eric Gimon, Michelle Solomon, and Mike O'Boyle, *Coal Cost Crossover 3.0: Local Renewables Plus Storage Create New Opportunities for Customer Savings and Community Reinvestment* (Energy Innovation, January 30, 2023), https://energyinnovation.org/report/the-coal-cost-crossover-3-0/.

[40] *"Co-Located Data Center, Nuclear Plant Power Deal Draws Rapid Ire From Industry,"* Concentric Energy Advisors, July 2, 2024. Retrieved from https://ceadvisors.com/co-located-data-center-nuclear-plant-power-deal-draws-rapid-ire-from-industry.
[41] Sonal Patel. "Talen, Amazon Launch $18B Nuclear PPA — A Grid-Connected IPP Model for the Data Center Era,." POWER Magazine, June 12, 2025. https://www.powermag.com/talen-amazon-launch-18b-nuclear-ppa-a-grid-connected-ipp-model-for-the-data-center-era/.
[42] Gimon, Ahlstrom, and O'Boyle, *Energy Parks*, https://energyinnovation.org/report/energy-parks-a-new-strategy-to-meet-rising-electricity-demand/.
[43] Brian Martucci. "Intersect Power, Google, TPG partner on up to $20B renewable energy push for data centers," Facilities Dive, December 18, 2024. Retrieved from https://www.facilitiesdive.com/news/intersect-power-google-tpg-partner-on-up-to-20b-renewable-energy-push-fo/735867/?utm_source=chatgpt.com.
[44] U.S. Energy Information Administration, "Solar, Battery Storage to Lead New U.S. Generating Capacity Additions in 2025," *Today in Energy*, February 24, 2025, accessed October 19, 2025, https://www.eia.gov/todayinenergy/detail.php?id=64586.
[45] Joseph Rand et al., *"Generation, Storage, and Hybrid Capacity in Interconnection Queues"*, Interactive Data Workbook, Lawrence Berkeley National Laboratory (2025), accessed October 19, 2025, https://emp.lbl.gov/generation-storage-and-hybrid-capacity
[46] "Data Centers Face Seven-Year Wait for Power Hookups in Virginia." Bloomberg News, August 29, 2024. https://www.bloomberg.com/news/articles/2024-08-29/data-centers-face-seven-year-wait-for-power-hookups-in-virginia.
[47] Natural Resources Defense Council, *At the Crossroads: A Better Path to Managing Data Center Load Growth*, September 15, 2025, https://www.nrdc.org/resources/crossroads-better-path-managing-data-center-load-growth.
[48] David E. Mills, Chair, Board of Managers, PJM Interconnection, "PJM Board Letter Re: Implementation of Critical Issue Fast Path Process for Large Load Additions," August 8, 2025, PJM Interconnection, 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
[49] PJM, *Conceptual Proposal—Non-Capacity-Backed Load (NCBL)*, August 18, 2025, https://www.pjm.com/-/media/DotCom/committees-groups/cifp-lla/2025/20250818/20250818-item-03---pjm-conceptual-proposal-and-request-for-member-feedback---presentation.pdf.
[50] "PJM Stakeholders Pan Proposal for Data Centers on Fast Track," Utility Dive, September 3, 2025, https://www.utilitydive.com/news/pjm-stakeholders-ncbl-data-center-fast-track/759096/.
[51] "PJM Revises Non-capacity Backed Load Proposal," RTO Insider, September 16, 2025, https://www.rtoinsider.com/115072-pjm-revises-non-capacity-backed-load-proposal/.
[52] Norris et al., *Rethinking Load Growth*.
[53] ICF. *Beneficial Electrification: Transforming the Energy System.* Accessed October 19, 2025. https://www.icf.com/insights/beneficial-electrification
[54] Brendan Pierpont, *Data Center Demand Flexibility* (San Francisco: Energy Innovation Policy & Technology LLC, March 11, 2025), https://energyinnovation.org/report/data-center-demand-flexibility/.

[55] Mehmet Turker Takci et al., "Data Centres as a Source of Flexibility for Power Systems" *Energy Reports* 13 (June 2025): 3661-3671, https://doi.org/10.1016/j.egyr.2025.03.020.
[56] Travis Kavulla, "A Basic Premise of Network Economics & Systems," LinkedIn, posted October 2025, accessed October 13, 2025, https://www.linkedin.com/posts/travis-kavulla-a199994_a-basic-premise-of-network-economics-systems-activity-7378929612451446784-CZFW/.
[57] Tim McLaughlin, "Big Tech's Data Center Boom Poses New Risk to US Grid Operators," *Reuters*, March 19, 2025, https://www.reuters.com/technology/big-techs-data-center-boom-poses-new-risk-us-grid-operators-2025-03-19/.
[58] Ian Walch, *A Transformative Approach to Bridging the Data Center Power Gap* (7×24 Exchange International, 2024), https://www.7x24exchange.org/a-transformative-approach-to-bridging-the-data-center-power-gap/.
[59] Ariel Wittenberg, "How Come I Can't Breathe?: Musk's Data Company Draws a Line over 35 Gas Turbines, No Air Pollution Permits," E&E News, May 1, 2025, https://www.eenews.net/articles/elon-musks-xai-in-memphis-35-gas-turbines-no-air-pollution-permits/.
[60] "Vantage Loses Appeal to Overturn Data Center Planning Permission Denial in Dublin." *Data Center Dynamics*, April 14, 2025. https://www.datacenterdynamics.com/en/news/vantage-loses-appeal-to-overturn-data-center-planning-permission-denial-in-dublin/.
[61] Institute for Energy Economics & Financial Analysis (IEEFA). *Data Centers Drive Buildout of Gas Power Plants and Pipelines in the Southeast*. January 10, 2025. https://ieefa.org/sites/default/files/2025-01/UPDATED-REVIEWED-Southeast%20Gas%20Infrastructure%20and%20Data%20Cente.pdf.
[62] Clean Air Task Force. *Optimizing Grid Infrastructure and Proactive Planning to Support Load Growth and Public Policy Goals.* July 22 2025. https://www.catf.us/resource/optimizing-grid-infrastructure-proactive-planning-support-load-growth-public-policy-goals/.
[63] GridLab, *Practical Guidance and Considerations for Large Load Interconnections* (San Francisco: GridLab, March 2025), https://gridlab.org/wp-content/uploads/2025/03/GridLab-Report-Large-Loads-Interim-Report.pdf
[64] Natural Resources Defense Council, *At the Crossroads: A Better Path to Managing Data Center Load Growth*, September 15, 2025, https://www.nrdc.org/resources/crossroads-better-path-managing-data-center-load-growth.
[65] Bipartisan Policy Center, *Electricity Demand Growth and Data Centers: A Guide for States.* February 5, 2025. Retrieved from https://bipartisanpolicy.org/report/electricity-demand-growth-and-data-centers/?utm_source=chatgpt.com
[66] Regulatory Assistance Project and Energy Regulators Regional Association, *Navigating Power Grid Scarcity in the Age of Renewable Energy — Policy and Regulatory Context and Tools*, October 14, 2024, https://www.raponline.org/wp-content/uploads/2024/10/RAP-ERRA-Power-grid-report-October-2024.pdf.
[67] Gimon, O'Boyle, and Solomon, *Meeting Growing Electricity Demand Without Gas*, March 2024.
[68] Brian Martucci, "DeepSeek called a net positive for data centers despite overcapacity worries." Utility Dive, February 20, 2025. https://www.utilitydive.com/news/deepseek-called-a-net-positive-for-data-centers-despite-overcapacity-worrie/740635/.

[69] Voltus, Inc., “Voltus Launches ‘Bring Your Own Capacity’ Product to Support Data Center Growth and Grid Resiliency,” September 30 2025, Voltus, Inc., https://www.voltus.co/press/bring-your-own-capacity-data-centers.
[70] Rewiring America, *Homegrown Energy*, September 2025.
[71] GridLab. “*Surplus Interconnection: Barriers, Opportunities, and Strategic Solutions*.” February 21, 2025. GridLab. https://surplusinterconnection.s3.us-east-1.amazonaws.com/2025-02-21_GridLab_Surplus_Interconnection_Barriers_Report.pdf.
[72] Alex Engel, Udal Varadarajan, and David Posner, *How “Power Couples” Can Help the United States Win the Global AI Race* (RMI, February 20, 2025), https://rmi.org/how-power-couples-can-help-the-united-states-win-the-global-ai-race.