Kimmeridge

Evolution of the APES: AI, Power, and Energy Supply

Table of Contents

1 Executive Summary

2 A Note on Our Methodology

2 Incremental Data Center Capacity Could Reach 55 GW by 2030

5 Incremental Gas Demand Could Reach 6 Bcf/d by 2030

6 Understanding and Estimating Key Gas Demand Sensitivities

10 Natural Gas Infrastructure Limitations

11 Hyperscaler Load Growth Accelerating Carbon Offset Demand

13 Conclusions

15 Appendix: Terminology and Our Methodology

Executive Summary

Since the launch of ChatGPT in November 2022, the AI hype cycle has been inescapable. By early 2024, many energy investors shifted their focus to the rapid expansion of AI-driven data center buildout and its potential impact on U.S. power consumption. A wave of top-down reports has since emerged, extrapolating load growth scenarios and market effects. However, these analyses often lack critical foundational insights and fail to provide actionable investment conclusions.

Kimmeridge, in collaboration with a third-party data center market expert, conducted an extensive analysis to model this megatrend with the highest possible accuracy. Leveraging our in-house data science capabilities and energy market expertise, we developed a bottom-up, regional U.S. power supply and demand model through 2030. Our analysis sought to answer key questions, including:

- What is the most accurate methodology for forecasting data center load growth?
- How large will the load be, when will it materialize, and where will it be concentrated?
- How will the U.S. power grid accommodate this growth?
- How will the generation mix, renewable queues, transmission, and congestion evolve?
- Will there be an incremental draw on natural gas, and can the infrastructure support it?
- What are the most influential variables affecting load growth and natural gas demand?
- What cascading effects and other key beneficiaries will emerge from this megatrend?

We began this study over a year ago, continuously tracking power market data and data center development. While we've revised our projections upward, our fundamental outlook remains unchanged. Recent DeepSeek news has reignited discussions, prompting speculation around muted AI-driven power demand. These developments were both anticipated inevitabilities and already incorporated in our analysis. Data centers currently under development are unlikely to deviate from our forecast within the study period.

Our analysis led to several key conclusions. First, we believe many power system operators are underestimating the scale of upcoming load growth. While constraints—primarily access to power—will limit expansion, our base case projects 55 GW of incremental wholesale, cloud, and AI data center demand by 2030. Including other base load growth, total projected demand reaches ~1,000 TWh, equivalent to adding Japan's entire electricity consumption to the U.S. grid within six years.

Second, despite an expected 300 GW of new solar, wind, and storage capacity coming online over the next five years, natural gas demand will increase. This demand will be highly regional, driven by concentrated load growth and infrastructure constraints. Our base case forecasts an additional 6 Bcf/d of natural gas demand by 2030, with even higher consumption possible if coal retirements proceed as planned. This increase is nearly equivalent to the total production of Expand Energy, the largest U.S. natural gas producer.[1] When combined with an additional 17 Bcf/d draw from under-construction LNG facilities, existing natural gas infrastructure will struggle to meet 2030 demand.

[1] Source: Expand Energy Corporation. "Expand Energy Corporation Reports Fourth Quarter and Full-Year 2024 Results." ***Expand Energy Corporation***. (2025).

After decades of efficiency gains suppressing power demand growth, the AI-driven shift will disrupt the established order of U.S. power markets from the past 25 years. This transformation will bring both unintended consequences and asymmetric opportunities. The proliferation of AI is profound, and energy consumption is poised to be one of its greatest beneficiaries. The era of stranded electrons and molecules will soon be a relic of the past. This analysis highlights the key variables shaping the outcomes of this generational secular shift.

A Note on Our Methodology

For data center demand modeling, we adopted a site-specific, bottom-up approach as the most accurate methodology. This involved geospatially mapping each data center under construction or development, assigning precise latitude and longitude coordinates, and determining or deducing its projected gross power capacity. This granular approach provides the most accurate picture of data center power demand and, critically, allows for integration with nodal-level power grid data. A full explanation of our methodology can be found in the appendix.

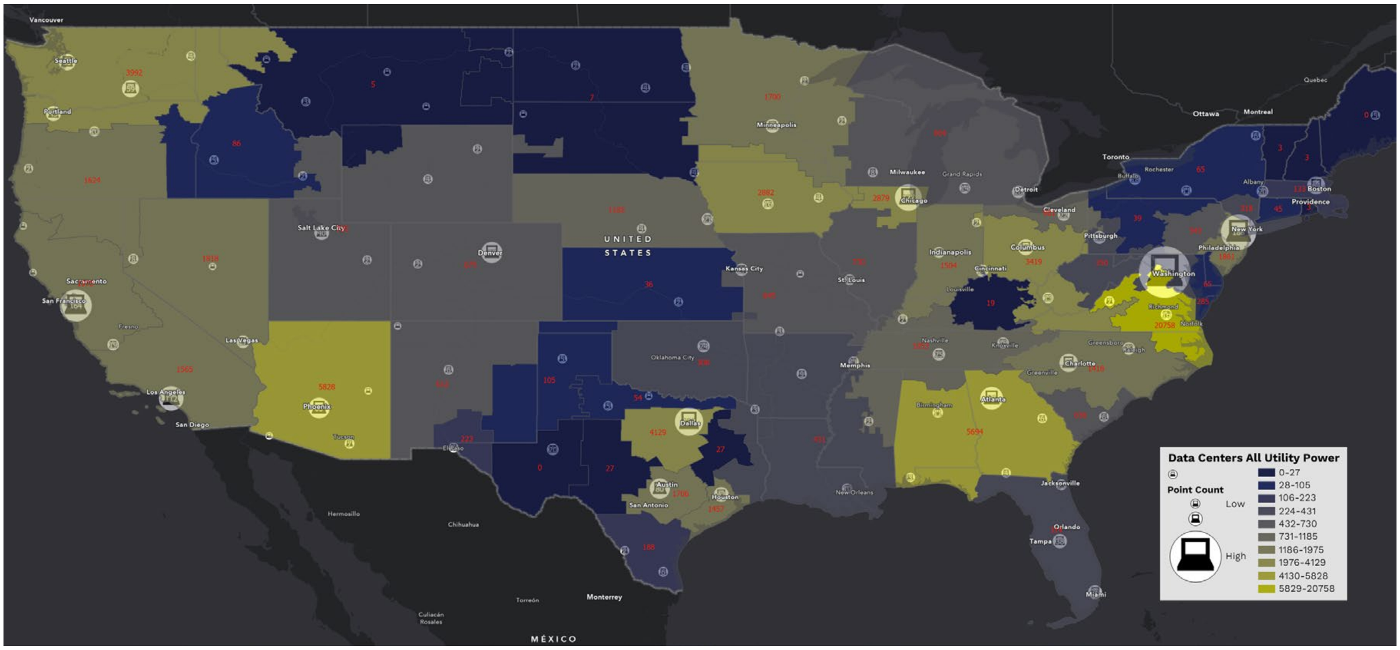
Incremental Data Center Capacity Could Reach 55 GW by 2030

Our base case forecast projects an additional 55 GW of cloud and AI data center capacity in the U.S. by 2030, bringing total capacity to ~80 GW, up from ~25 GW at the end of 2024. However, this growth will not be evenly distributed across major markets.

The U.S. power grid is managed by Regional Transmission Organizations (RTOs) and Independent System Operators (ISOs), which coordinate, control, and monitor single- or multi-state electric grids. North America has nine ISOs, each responsible for regional electricity distribution. Among them, PJM—originally formed by Pennsylvania, New Jersey, and Maryland but now overseeing the grid for all or parts of 13 states and the District of Columbia—is particularly significant as it includes Ashburn, Virginia, commonly known as Data Center Alley, a major global data center hub near Dulles Airport. Outside of ISOs, power grids are managed by interconnected regulated utilities.

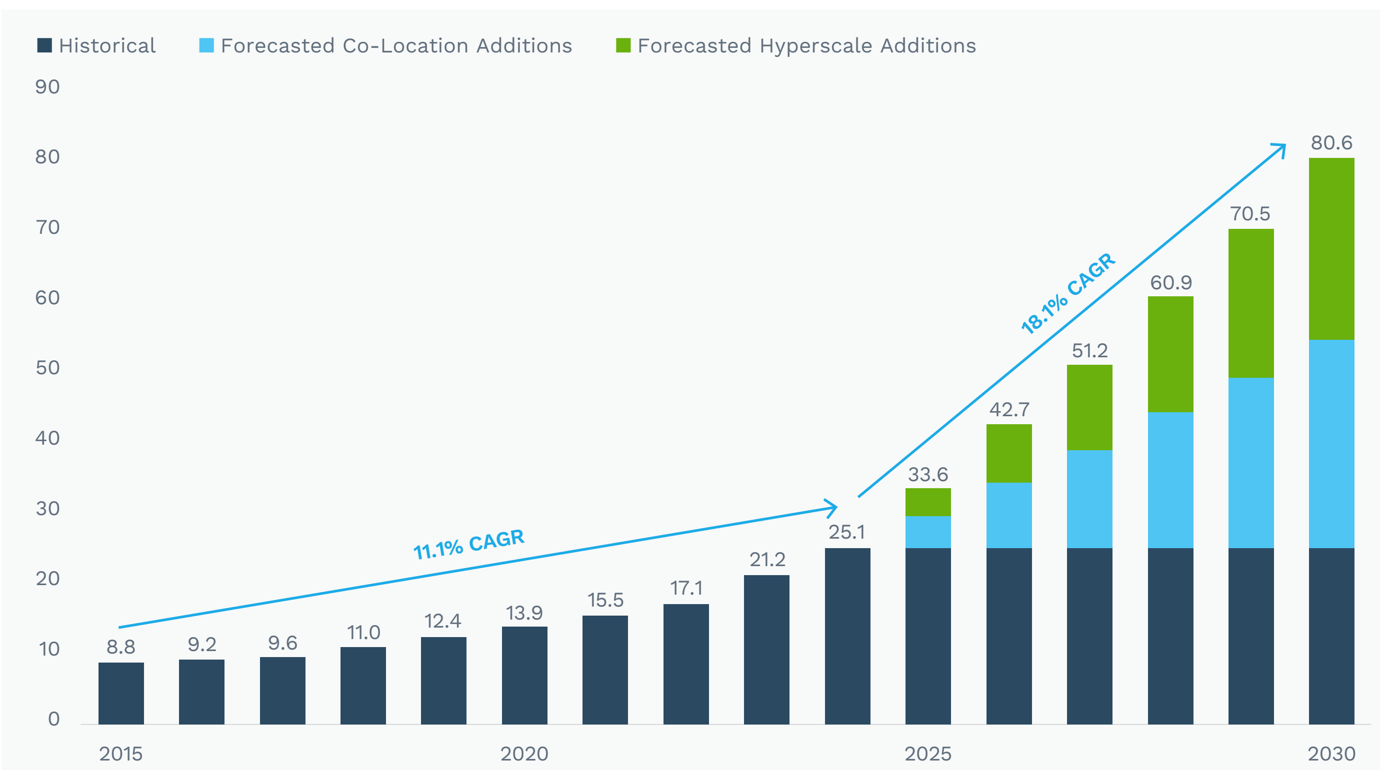
Our analysis indicates that PJM, ERCOT, and non-ISO regulated utilities in the Southeast will experience the most significant data center expansion. In PJM, the highest growth will be concentrated in two key clusters: Ashburn, Virginia, and the eastern side of Columbus, Ohio. Despite being the most mature data center market in the world, Ashburn continues to lead all U.S. metropolitan areas in growth. In ERCOT, data center development is marching west from Dallas-Fort Worth towards energy-dense West Texas. In the Southeast, expansion is largely occurring in the territories of regulated utilities Southern Company and Duke Energy.

FIGURE 1: Significant Data Center Load Growth Across the Southeast, Virginia, Ohio & Texas (MW)



Source: Kimmeridge Proprietary Analysis

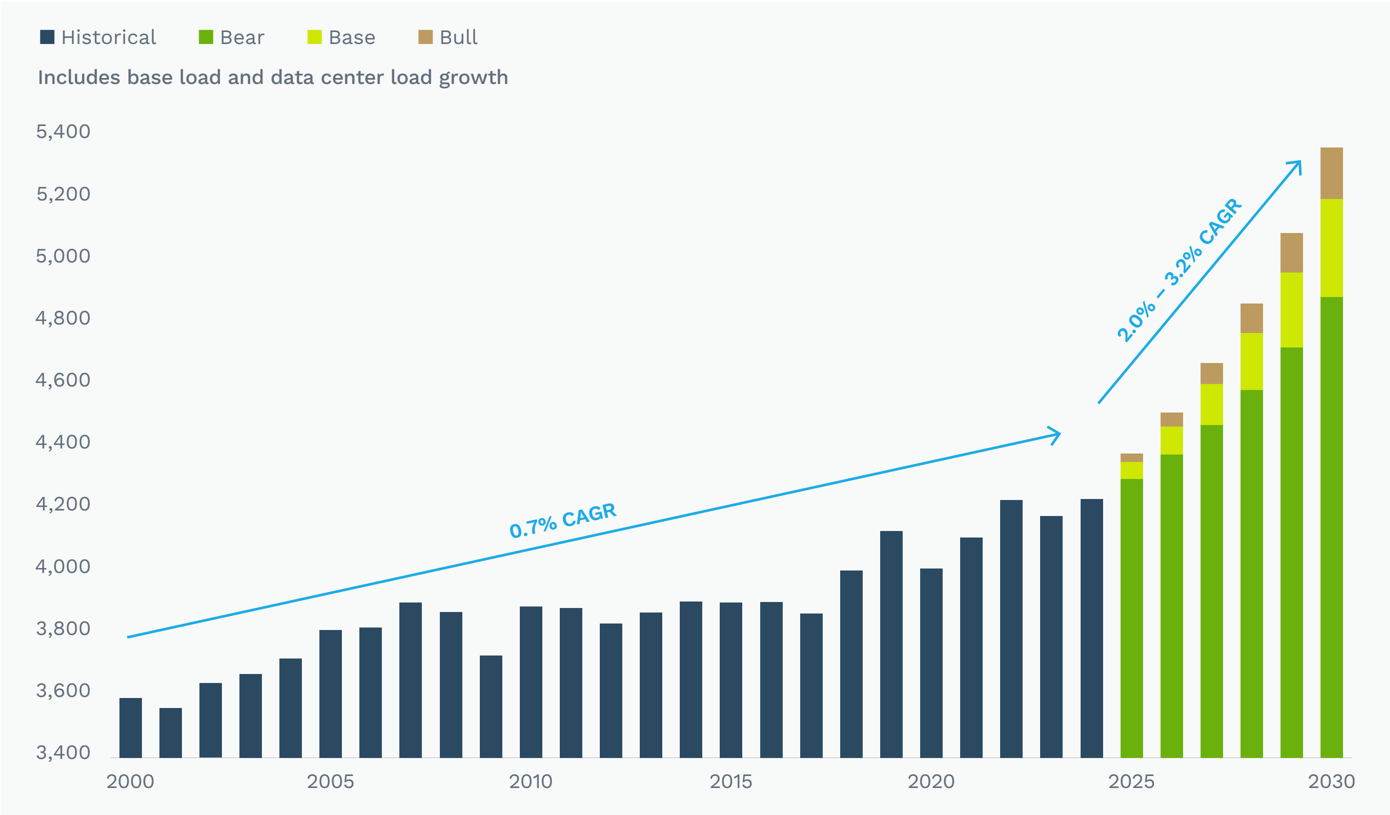
FIGURE 2: Historical and Projected U.S. Data Center Capacity (GW)



Source: Kimmeridge Proprietary Analysis

The addition of 55 GW of data center capacity translates to ~400 TWh of incremental data center-specific load growth by 2030, based on our base case utilization assumption. When factoring in base load growth, incremental power demand is projected to reach 1,000 TWh, which is equivalent to adding seven New York Cities to the U.S. power grid. Our load growth scenarios range from 700 TWh (bear case) to 1,200 TWh (bull case) by 2030, depending on data center construction timing, utilization rates, and broader base load growth dynamics.

FIGURE 3: Historical and Projected U.S. Power Consumption (TWh)



Source: Kimmeridge Proprietary Analysis

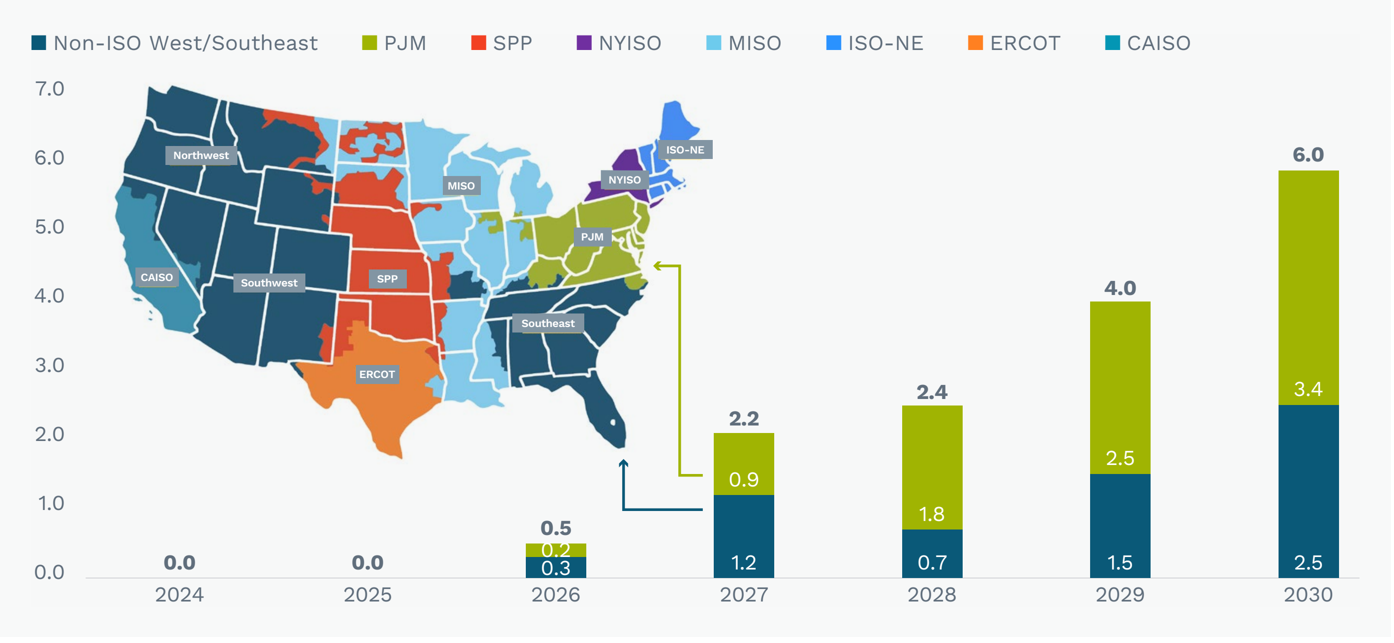
Our base load growth forecasts are modeled at the state level, incorporating historical trends, population shifts, and, most critically, manufacturing, industrial, and crypto-related demand. While power consumption has remained relatively flat in most states over the past two decades, secular trends such as electrification, onshoring, and energy transition-driven manufacturing are now driving demand growth.

Certain states are projected to see 1% to 3% annual base load growth, a significant impact given the current U.S. base load of ~4,200 TWh annually. Historically, load growth has averaged ~0.7% per year since 2000, but our model forecasts a 3–5x acceleration over the next six years. Meeting this surging demand will require a mix of growing renewable generation and slowing coal retirements, but the key swing generation capacity will come from natural gas. With abundant domestic gas resources, the U.S. is well-positioned to supply the flexible, dispatchable power needed to support this rapid growth.

Incremental Gas Demand Could Reach 6 Bcf/d by 2030

While many factors influence load growth, even more impact natural gas demand, leading to greater variability in potential outcomes. Power demand scenarios show a maximum variance from base of 30%, ranging from 700 TWh (bear case) to 1,000 TWh (base case). In contrast, gas demand scenarios exhibit a maximum variance of 100%, from 6 Bcf/d (base case) to nearly 12 Bcf/d (bull case). The two largest factors driving gas demand uncertainty are the percentage of queued renewable projects successfully brought online and the pace of coal retirements.

FIGURE 4: Base Case Projected Incremental Natural Gas Draw by Region (Bcf/d)



Source: Kimmeridge Proprietary Analysis

A conservative base case for coal retirements and renewable additions supports a high confidence forecast of 6 Bcf/d of incremental gas demand by 2030. However, this demand will be highly regional, primarily concentrated in PJM and the Southeast.

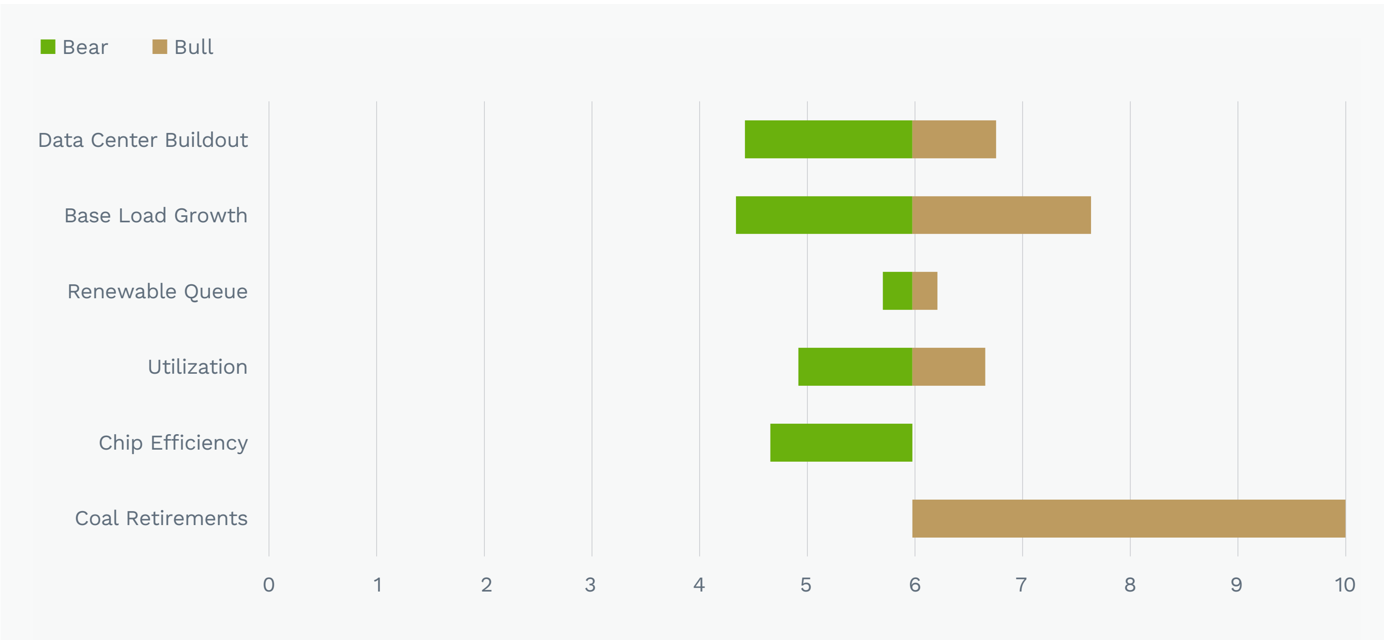
Despite significant data center and manufacturing growth in Texas and parts of the Midwest, our model does not forecast incremental gas demand in ERCOT, MISO, or SPP. This is because projected renewable and storage additions in these regions are expected to be sufficient to meet new data center load growth. While data centers require a continuous 24/7 power supply and wind and solar generation is intermittent, incremental gas demand is assessed on an annual basis. On days when renewables fall short, gas generation will temporarily increase. However, the overall annual gas consumption due to incremental data center demand is not expected to rise, as growing periods of idle gas-fired generation from increased renewables will offset the period of higher gas burn.

Analyzing daily capacity factors presents a complex challenge. On average, natural gas plants—both single-cycle and combined-cycle—operate at a 42% capacity factor nationally, suggesting sufficient excess capacity to absorb future demand growth. However, during peak summer and winter periods, many facilities run at 100% capacity for multiple days, straining generation resources. In such instances, existing capacity may be insufficient to support new peak loads. We continue to refine our supply and demand model to better capture regional daily capacity imbalances.

Understanding and Estimating Key Gas Demand Sensitivities

Recognizing the inherent uncertainty in six-year macro forecasts, it is critical to identify the key drivers that influence potential outcomes. To assess this, we conducted a single-variable sensitivity analysis, testing each variable across high-confidence intervals while holding all other factors constant to base case assumptions.

FIGURE 5: Natural Gas Demand Single Variable Sensitivity Analysis (Bcf/d)



Source: Kimmeridge Proprietary Analysis

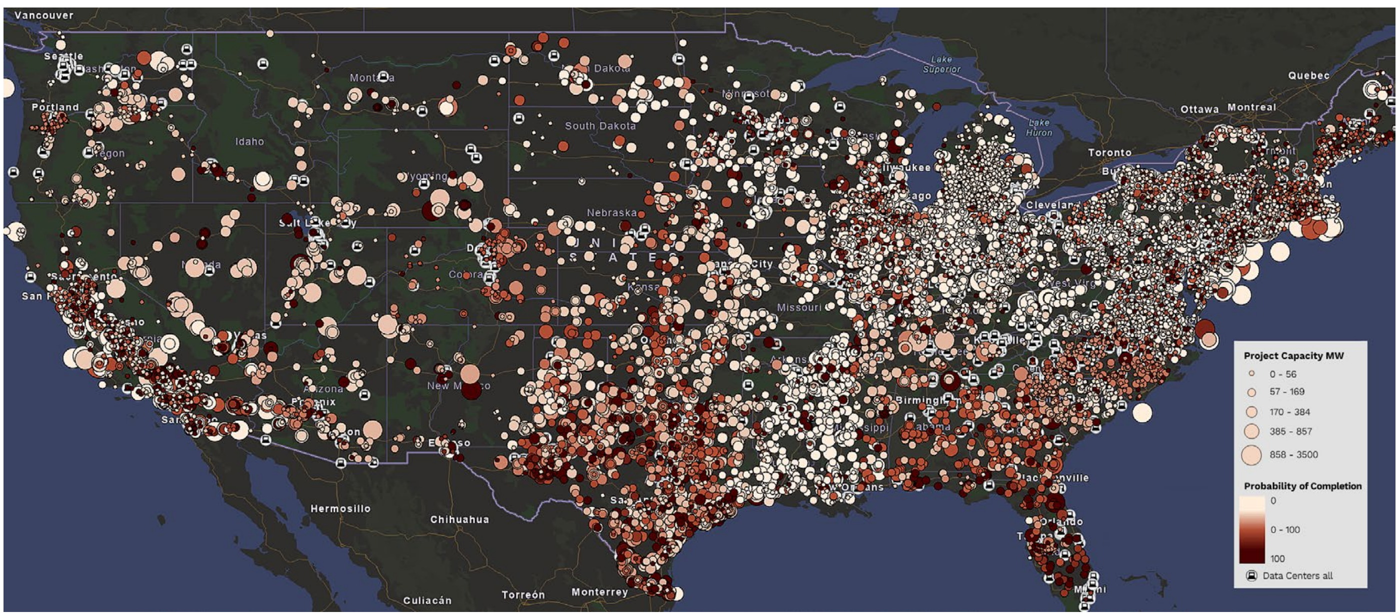
Data center capacity additions range from 40 GW (bear case) to 65 GW (bull case), with a base case of 55 GW. Given the significant impact of base load growth on this analysis, we conducted an in-depth evaluation to develop a differentiated and well-informed outlook. Several secular trends are driving base load growth in the U.S., including electrification, onshoring of manufacturing and supply chains, crypto mining, and expansion of renewable and energy transition-related industries—the latter heavily incentivized by the Inflation Reduction Act. State-level and site-specific assessments of crypto and manufacturing loads further informed our base load growth projections, which range from -0.3% to 3.0% annually, depending on the region.

The Renewable Queue

The U.S. interconnection queue currently holds over 2,000 GW of solar, wind, and battery storage projects, nearly double the 1,300 GW of installed capacity today. While this capacity could, in theory, support all data center expansion and broader load growth, only a small fraction of these projects will ultimately gain interconnection approval and be brought online. As a result, the renewable queue completion rate remains the single most significant variable shaping power markets and natural gas demand.

Developers overload queues with applications far exceeding what the grid can realistically accommodate. Historically, only 8% of projects receive approval,[2] and while this rate may increase due to rising power demand and regulatory reforms, transmission constraints will remain a major bottleneck, severely limiting total capacity additions.

FIGURE 6: Proprietary Risk-Weighted Renewable Generation Queue

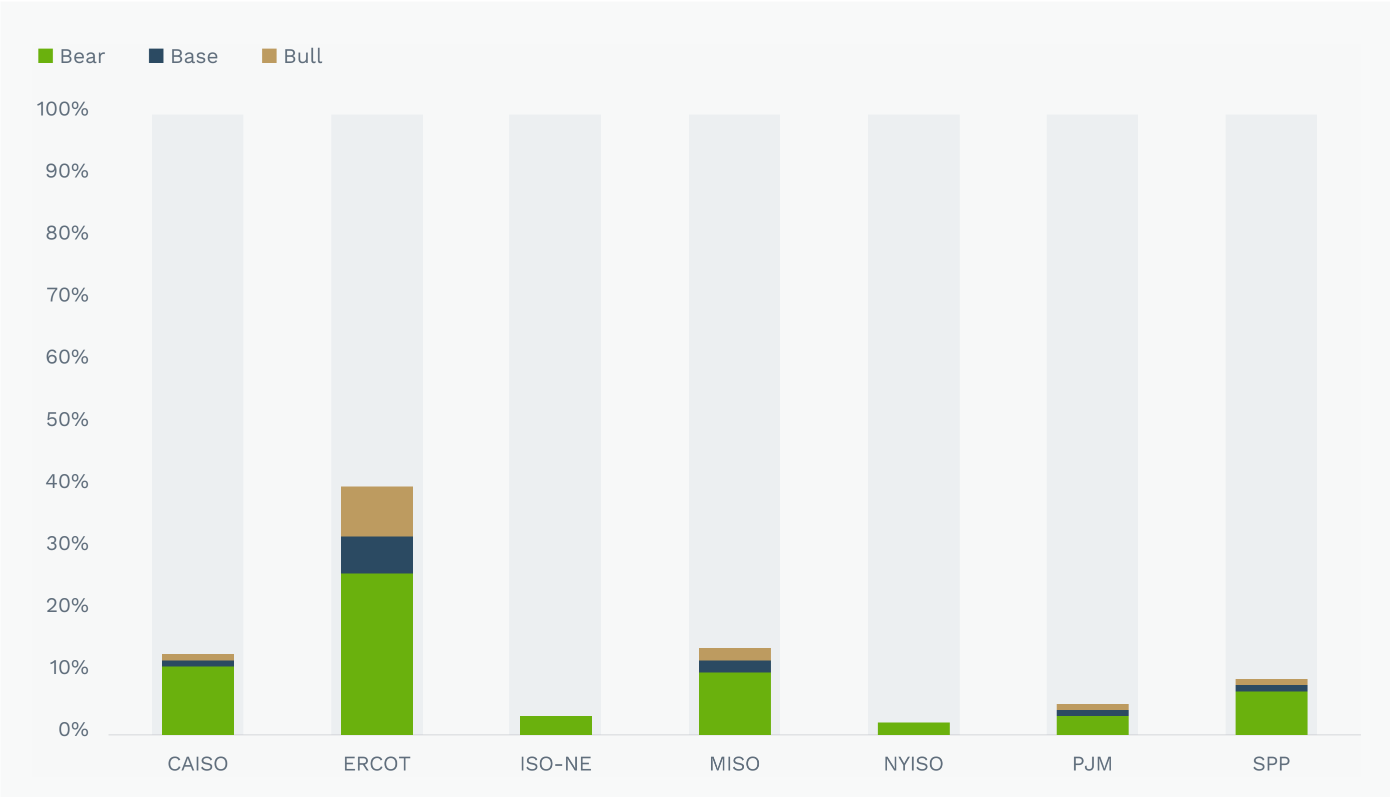


Source: Kimmeridge Proprietary Analysis

By analyzing historical projects by ISO, we identified the key factors that determine whether a project gains interconnection approval or fails. Using these insights, we developed a predictive model to assess the likelihood of success for projects currently in the queue. Our model forecasts that ~4% of projects in PJM, 12% in CAISO, 32% in ERCOT, and 11% in MISO will ultimately reach operational status, then further sensitized for bear and bull renewable penetration scenarios in each system. Our base case forecast predicts an aggregate 300 GW of solar, wind, and storage capacity will become operational by 2030.

[2] Source: Kimmeridge proprietary renewable queue analysis.

FIGURE 7: Probability Intervals of Renewable Queued Capacity by ISO



Source: Kimmeridge Proprietary Analysis

Utilization vs. Chip Efficiency

In this demand model, data center utilization and chip efficiency are interrelated variables. Recently, U.S. data center utilization has averaged ~65%, with a trend toward 70%. Meanwhile, chip efficiency—measured by the number of compute operations for a given unit of energy—has been improving at a rate exceeding Moore's Law.[3]

Bearish power forecasts often assume that efficiency gains in GPU design will significantly curb future data center power demand. However, these projections contradict empirical evidence and ignore Jevons' Paradox, where greater efficiency drives expanded use cases and increased demand. A modern example is internet energy consumption—while energy intensity per unit has declined by 15% per year since inception, data volumes have increased by 30% annually, resulting in a net 15% annual increase in total energy consumption.[4]

An analysis of compute-level energy consumption since the rise of deep learning in the 2010s shows that AI compute across model architectures has benefited from more powerful hardware, delivering higher floating-point operations per second ("FLOPs") per watt.[5] However, models have also increased FLOP consumption per unit of input, striving for greater accuracy and multimodality.

[3] Source: Huang's Law.

[4] Source: West, Rob. "Jevons Paradox: what evidence for energy savings?" ***Thunder Said Energy*** (2023).

[5] Source: Hobbhahn, Marius, and Tamay Besiroglu. "Trends in GPU price-performance." ***Epoch AI*** (2022).

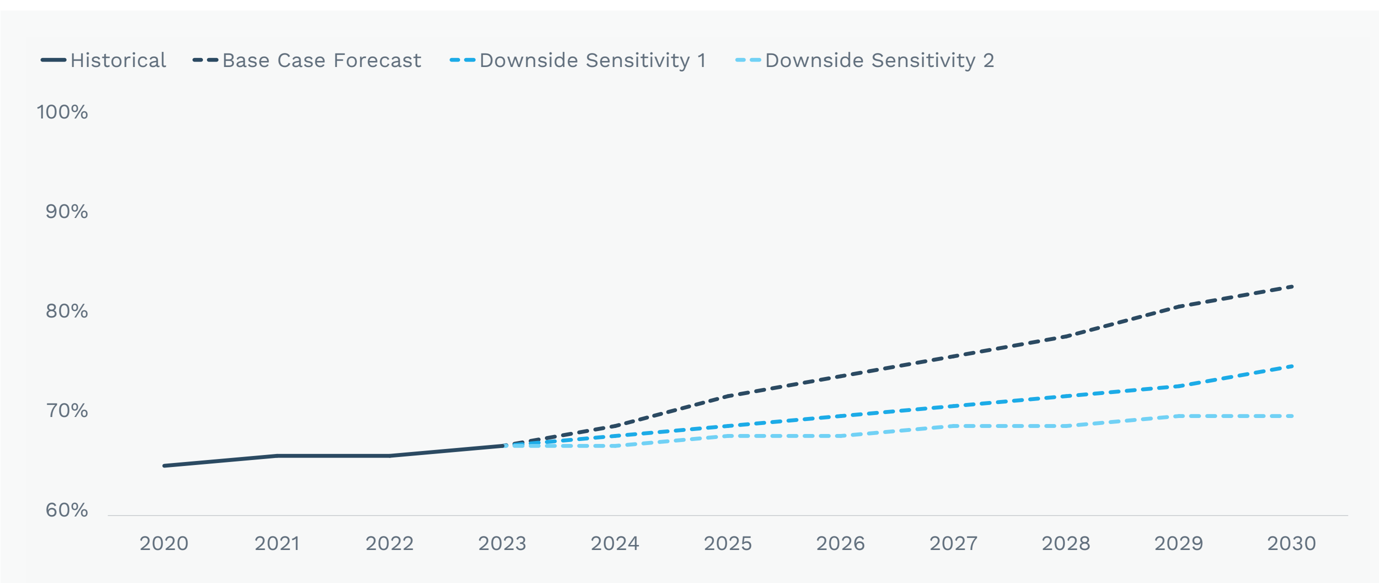
While hardware efficiency continues to improve, AI models are scaling faster than efficiency gains can offset per-unit energy demand, making the net impact on power demand uncertain. Beyond chip-level advancements in materials engineering, additional efficiency gains will come from more sophisticated model architectures, reducing reliance on brute-force compute scaling to achieve higher performance. DeepSeek exemplifies this trend, leveraging innovative engineering optimizations, many of which are refined versions of existing techniques, to rapidly close the gap with OpenAI.

At the same time, substitution effects must be considered. For example, GPUs consume more power than central processing units (CPUs), but their superior speed lowers the unit cost per computation. Additionally, the widespread adoption of AI-driven energy efficiency technologies could create a positive feedback loop, mitigating some power demand growth over time.

Recent developments surrounding DeepSeek do not alter our power demand outlook, as our analysis already accounted for both continued model architecture efficiencies and the expansion of AI beyond hyperscalers. The increased democratization of AI will introduce more market entrants competing for power, particularly those who lacked the capital or access to NVIDIA chips previously required for AI compute expansion.

Looking ahead, inference queries are projected to grow by as much as 1,000x by 2030,[6] as the computational burden shifts from training to inference. While efficiency gains may eventually lead to lower utilization rates, such shifts will only materialize beyond 2030 and remain irrelevant to our forecast period. The AI arms race is still in its infancy, and power consumption is set to grow significantly beyond 2030, underpinned by hyperscalers' 2025 capital expenditure commitments. While the magnitude of efficiency-driven impacts remains to be seen, we expect a continued upward trajectory in power demand, even if efficiency gains moderate the rate of growth. Since our data center power demand is modeled at the site level's gross power capacity, any efficiency-driven impacts are already reflected in the model through moderated utilization rates over time.

FIGURE 8: Historical and Forecast Average Data Center Utilization



Source: Kimmeridge Proprietary Analysis

[6] Source: West, Rob. "Energy intensity of AI: chomping at the bit?" ***Thunder Said Energy*** (2024).

Coal Retirements

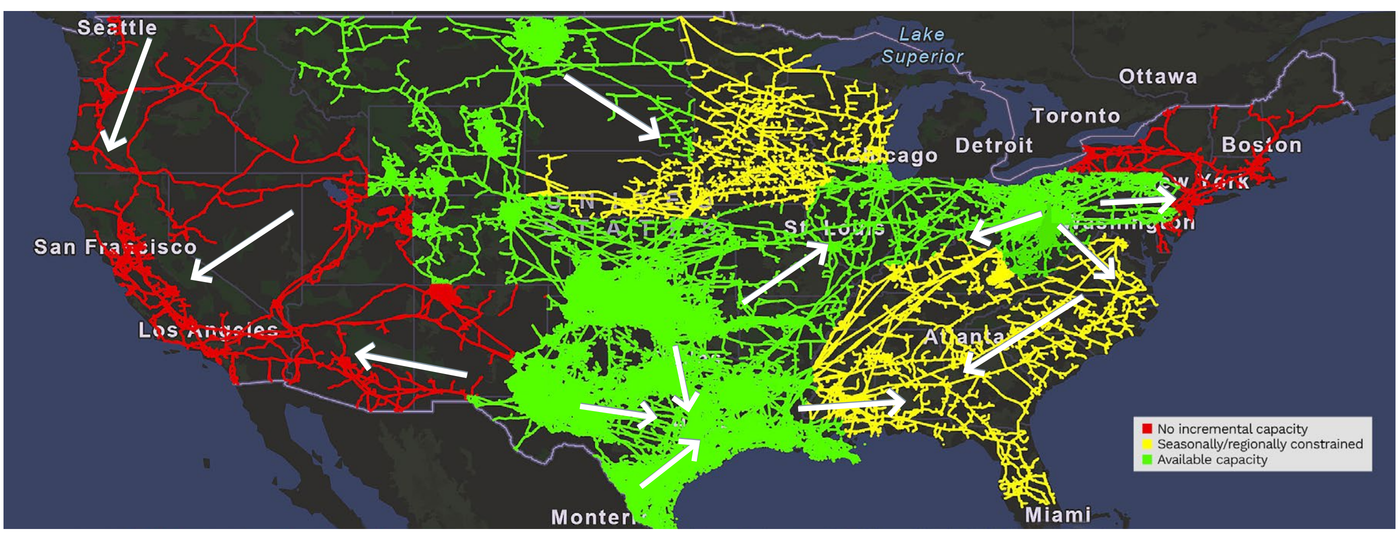
Coal retirements are a major driver of incremental natural gas demand, with every 10 GW of retired coal capacity adding approximately 2 Bcf/d of additional gas draw. However, aggressive coal retirement forecasts are now highly unlikely, given the paradigm shift in AI-driven load growth. Our analysis adjusts for a fraction of the quoted coal retirement capacity, leading to the clear conclusion that retirements will be delayed as system operators prioritize grid stability and economic incentives keep coal plants online.

Our base case assumes that only those coal plants that are already scheduled for decommissioning will be retired. A review of these facilities shows average plant lifespans of 50 to 75 years, reinforcing a conservative outlook on retirements. Consequently, natural gas demand in our model is only adjusted to the upside. If utility-published coal retirement forecasts were to materialize, the resulting gas demand would be unachievable. In fact, retiring just half of the forecasted capacity in PJM and the Southeast alone would require an additional 10 Bcf/d of gas, underscoring the mismatch between projected retirements and realistic energy supply needs.

Natural Gas Infrastructure Limitations

U.S. natural gas resources are concentrated in a few key basins, but despite an extensive pipeline network, many long-haul transmission lines are already at full capacity. Expanding this infrastructure is increasingly difficult and often infeasible due to regulatory hurdles, environmental opposition, and permitting constraints.

FIGURE 9: U.S. Natural Gas Infrastructure Available Capacity



Source: Kimmeridge Proprietary Analysis. Note: pipeline capacities simplified for illustrative purposes.

In Eastern Ohio, within PJM, gas demand is rising as data centers and semiconductor manufacturing expand east of Columbus. The region benefits from incremental gas transport capacity and its proximity to the Appalachian Basin. While additional gas supply is accessible, Combined Cycle Gas Turbine (CCGT) capacity factors have risen ~20% over the past four years, signaling potential tightening in natural gas generation capacity.

In Northern Virginia, expanding natural gas supply to Data Center Alley remains challenging despite its proximity to Appalachia and multiple major gas transmission systems. The region relies on two key pipelines: William's Transco and TC Energy's Columbia Gas. Columbia Gas transports gas west into demand centers with some annual excess capacity but faces seasonal constraints. Transco moves gas north or south from the southern Virginia border, and while the newly operational Mountain Valley Pipeline (MVP) adds 2 Bcf/d into Transco, northbound flows are already at full capacity, preventing Ashburn from accessing additional supply. Meeting Data Center Alley's growing demand will require pipeline expansions, but regulatory hurdles in the Northeast make future additions uncertain.

Emerging data center hotspots across the Southeast will have access to additional natural gas but must compete with LNG exports. The region depends on two major pipelines, Williams' Transco and Kinder Morgan's Southern Natural Gas, which source gas from the Haynesville, Texas, and Appalachia. While MVP's 2 Bcf/d is increasing southbound supply via Transco, competition between regulated utilities and LNG buyers is already evident—Transco Zone 4 saw the highest U.S. gas prices in summer 2022, reaching a $5.50/MMcf premium over Henry Hub.[7]

All long-haul natural gas transmission infrastructure west of the Rockies and the New Mexico Delaware Basin is fully utilized, leaving no capacity for additional volumes. Phoenix, a fast-growing data center hub with rising gas demand, exemplifies this challenge. Our generation capacity analysis shows that both existing capacity and planned power additions will fall short of meeting projected load growth. With gas infrastructure constraints limiting new supply, securing additional natural gas to support demand will not be possible without additional pipeline capacity.

Hyperscaler Load Growth Accelerating Carbon Offset Demand

The most significant secondary impact of surging data center load is the growing demand for carbon offsets. Hyperscaler power consumption is accelerating, driving more natural gas burns, delaying coal retirements, and intensifying the challenge of meeting 2030 Net Zero commitments. The trend is already evident—Microsoft's emissions have risen 29% since 2020,[8] while Google's have increased 48%since 2019, including a 13% year-over-year jump.[9] As data center expansion outpaces dispatchable clean energy deployment, demand for carbon offsets will continue to climb.

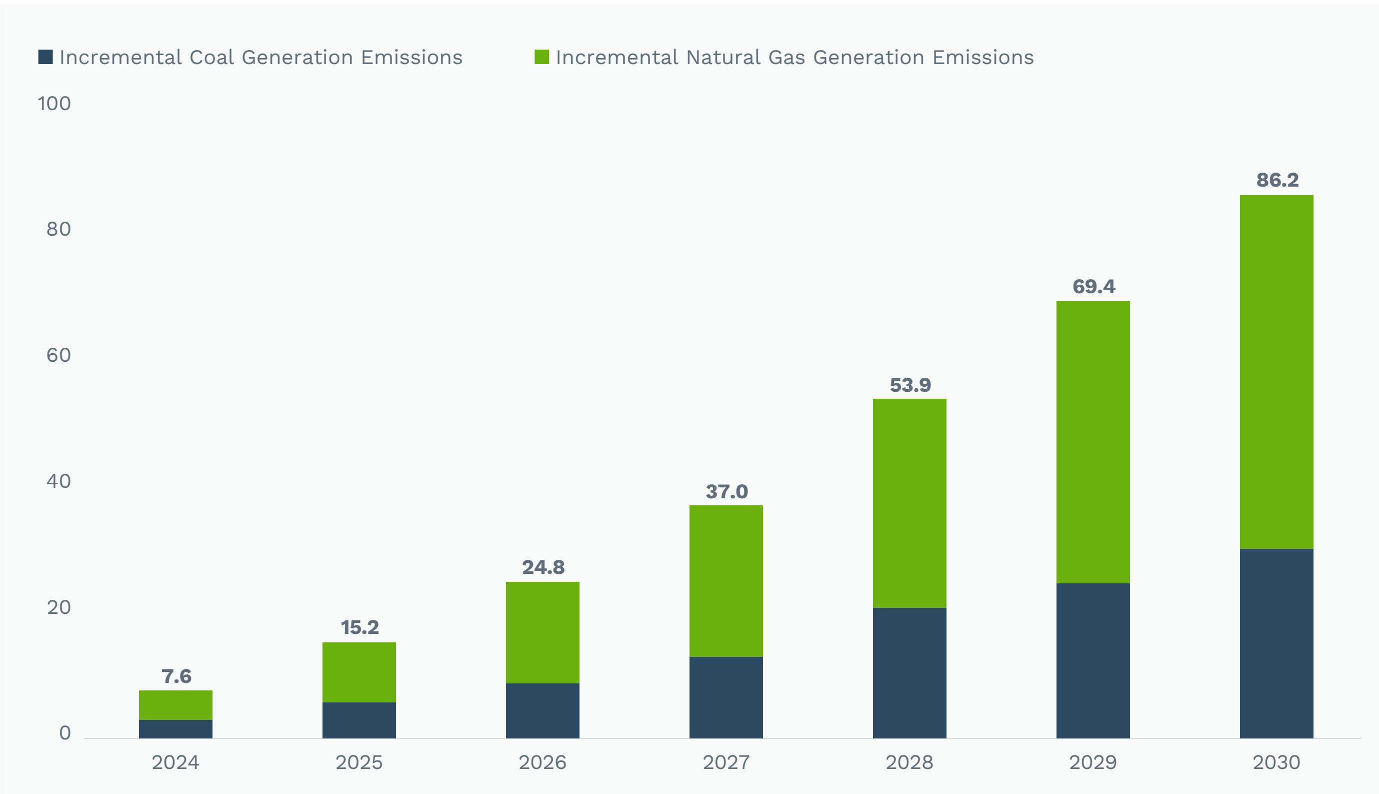
[7] Source: Natural Gas Intelligence. ***Transco Zone 4 Daily Prices.*** (2025).

[8] Source: Microsoft. ***2024 Environmental Sustainability Report.*** (2025).

[9] Source: Google. ***2024 Environmental Report.*** (2025).

Using site-specific forecasts for hyperscale compute load and system-level projections for generation mix, we calculated the incremental carbon offset demand through 2030 for hyperscalers with 2030 Net Zero targets. This analysis accounts for hyperscale compute demand from both self-built facilities and as co-location tenants but does not include new renewable PPAs or behind-the-meter clean generation initiatives that hyperscalers are actively pursuing. By 2030, cumulative emissions are projected to rise by 300 million tonnes of CO_2, incremental to 2023 baselines.

FIGURE 10: Implied Incremental CO_2 Emissions From Hyperscalers (Million Tonnes)



Source: EIA, US electricity net generation and resulting CO_2 emissions by fuel. Assumes 1.044 and 0.440 CO_2 metric tonnes per MWh for coal and natural gas, respectively; Kimmeridge proprietary analysis

Conclusions

1. Data center load growth is real, and many system operators, until recently, were likely underestimating its scale. Recent revisions to load growth by ISOs like ERCOT are predicting more load than can reasonably be built. Growth in some regions will be constrained, primarily by limited power availability. Historically, paradigm shifts have often outpaced even the most aggressive forecasts, and after decades of stagnating power demand, utilities and system operators may be ill-equipped to handle this surge. Regulatory barriers and transmission constraints will further hamper the grid's ability to keep pace with this mega trend.

2. We forecast 55 GW of additional data center capacity in the U.S. by 2030, an increase from the 50 GW projected during our initial findings in mid 2024. The 400 TWh of incremental load implied by our data center utilization assumptions aligns near the median of third-party research from industry experts, global consulting firms, and bank equity research. Given the accelerating pace of AI-driven demand, we expect these forecasts to be continuously revised upward.

3. The incremental draw on natural gas will be highly regional, driven primarily by renewable generation additions, available natural gas infrastructure capacity, and coal retirements. While renewable queue completion rates are expected to rise as load growth pressures regulatory reform, site and transmission constraints will limit actual renewable deployments to a fraction of the total queue. Similarly, natural gas infrastructure expansions and debottlenecking will increase supply to many regions, but regulatory barriers will materially impede volume growth. In regions with excess natural gas supply, average gas generation capacity factors have already risen significantly, and generation capacity itself will soon become a key constraint.

4. Coal retirements will face significant delays. Even before the current wave of load growth, the combination of thermal generation retirements and greatly renewable penetration had already raised grid stability concerns. Current coal retirement forecasts severely overestimate both the timing and scale of facilities expected to go offline. Recent capacity auction results reinforce the trend of economic incentives favoring grid stability, prioritizing generation supply retention over retirements. With coal suppliers recontracting future shipments for export, in anticipation of domestic plant closures facilities remain online longer than expected, the U.S. could face a structural coal supply shortage.

5. PJM will consume more natural gas as data center demand continues to grow, renewable additions are limited, and existing gas resources and infrastructure support sustained load growth. Despite being the most mature data center market globally, Ashburn is projected to see the highest data center additions of any U.S. metropolitan region. However, natural gas infrastructure constraints complicate Ashburn's long-term energy supply outlook. Meanwhile, regions within PJM with excess natural gas supply, particularly Eastern Ohio, are experiencing a significant rise in gas generation transactions and capacity expansions, reinforcing the region's role as a key supplier for growing power needs.

6. A string of data center developments is underway across the Southeast, following the path of Williams' Transco pipeline. These facilities, combined with large-scale manufacturing growth in the region, will drive significant incremental natural gas demand. In many regions, regulated utilities will face direct competition with LNG facilities for Gulf Coast gas supply. While the Mountain Valley Pipeline's additional 2 Bcf/d will help bolster Southeast gas availability, demand pressure will persist as LNG facilities ramp up consumption, adding an incremental 17 Bcf/d of gas demand by 2030. Gulf Coast gas supply from Texas and the Haynesville will see significant demand pulls into the region.

7. The Western U.S. data center buildout is expected to slow, as the region lacks the necessary natural gas infrastructure to support rising demand. Natural gas pipelines from the Rockies and Permian Basin are already at capacity, with no planned expansions to accommodate additional load growth. While the Southwest benefits from abundant solar resources, solving renewable intermittency will require major investments in battery storage. Without significant solar-plus-storage additions, regions could face power shortfalls. In the Pacific Northwest, all available excess hydropower capacity is already spoken for.

8. Different regions will meet future data center power needs through varied approaches. In ERCOT, a high renewable queue completion rate and immense queue are expected to exceed the state's significant data center and base load growth. In NYISO and ISO-NE, given minimal data center additions and flat-to-modest base load growth, forecasted renewable and storage additions should be sufficient to meet future demand. Lastly, in MISO and SPP, a mix of excess current generation and renewable capacity growth is projected to outpace data center and base load needs. However, large-scale coal retirements could disrupt power balances within these ISOs. Additionally, while annual power availability may appear sufficient, daily capacity factors in MISO and SPP could indicate potential regional shortages during peak summer and winter demand periods.

9. Extensive discussions with data center developers—both co-location providers and self-owned hyperscalers—reveal two recurring themes shaping development decisions: grid resiliency and time to power availability. These priorities have renewed interest in underutilized natural gas plants and potential nuclear recommissioning as interim solutions. While behind-the-meter solutions may emerge as secondary power sources—for backup or expansion—the need for eventual grid connectivity remains a limiting factor for off-grid facilities. Additionally, the concept of inference-only facilities built away from compute consumer hubs, where higher latency is offset by access to excess or stranded gas and power, currently remains a theoretical inevitability rather than an accepted near-term strategy.

10. The rapid expansion of data centers will drive an unprecedented need for more electrons and more molecules, with a premium on clean versions of each. AI-driven power demand is accelerating faster than grid capacity expansions, making every energy source critical to sustaining growth. Renewable deployment will need to scale at an unprecedented pace, but natural gas will remain the key swing fuel, filling gaps where clean energy falls short. At the same time, delayed coal retirements, rising LNG exports, and regulatory hurdles will strain regional fuel supplies. With net zero commitments looming, hyperscalers will increasingly turn to carbon offsets to bridge the gap between demand growth and clean energy availability. The scale of this transformation means no single solution is sufficient—meeting future data center power needs will require more of everything.

Appendix: Terminology and Our Methodology

Terminology

In this analysis, load forecasts focus exclusively on the wholesale data center market, encompassing wholesale, cloud, and AI compute demand from both co-location and hyperscale facilities. Co-location facilities are owned by data center providers, who lease space—along with power, security, cooling, and network connectivity—to third-party servers. These developers and owners are actively pre-leasing future rack space worldwide. Hyperscalers—including Amazon's AWS, Apple, Google, Meta, Microsoft's Azure, and Oracle—operate their own self-built data centers while also representing a significant portion of the co-location tenant market. To avoid double-counting demand, our analysis defines hyperscale demand strictly as power consumption from self-built and operated hyperscale data centers, excluding their footprint within co-location facilities.

Enterprise data centers and retail co-location facilities are excluded from our demand forecasts. A typical example of an enterprise data center is a Visa or MasterCard facility that processes credit card transactions. Much of this compute workload is shifting to the cloud, and residual enterprise data center power demand is accounted for within our base load assumptions.

Demand Model Methodology

Alternative bottom-up and top-down approaches—such as GPU sales or server import forecasts—are fundamentally flawed because they overlook the longest lead-time constraints in data center development: permitting, construction, and interconnection queues for power.

For example, the widely used NVIDIA GPU sales forecast methodology lacks reliable assumptions on regional sales distribution and relies on speculative estimates of NVIDIA's market share by decade's end. Similarly, the server import forecast methodology fails to account for load limitations at both existing and future data centers. While server racks have a 5–6-year useful life and are continuously replaced by newer power-hungry new generations, current data center infrastructure and substations cannot support a 30x increase in power draw per server.

Most critically, non-site-specific methods ignore the regional nature of the power grid, where transmission and congestion constraints significantly impact electricity distribution over short distances. To accurately forecast how the power grid will be impacted, the exact location of the data centers is required.

In collaboration with a data center market expert, a location-specific dataset was developed and refined using a combination of direct engagement with data center developers, public disclosure scrapes, permitting reviews, low-earth orbit observations, and on-site inspections. For co-location facilities, initial forecasting leveraged the future leasing market, where operators actively pre-sell rack space in upcoming data centers as early as possible. Sales teams with pipelines of prospective facilities are in-market securing leases for future square footage. Given the long lead times in planning, development, and construction, facility providers align in-service dates with GPU deliveries extending beyond the decade, allowing for the long-term forecasting required in this analysis.

Boots-on-the-ground research in every major data center market facilitated address-specific gross power demand projections through 2030. While leasing pipelines may sometimes be inflated by developers overselling rack space, local market experts risk-weight forecasts site by site, applying regional expertise and professional judgment. These estimates are further validated through permitting data and direct observation, ensuring accuracy against expected in-service dates.

The hyperscale self-built market is intentionally opaque. These companies are not pre-selling rack space and are often reluctant to disclose precise facility locations or power consumption due to security and environmental reporting concerns. Hyperscale development plans were estimated through a series of deductions, starting with public disclosure scrapes from corporate announcements and regional reports on job creation and tax revenues. Further refinements to confirm specific addresses were made by way of public zoning permits, low-earth orbit imagery, and physical site observations. Once latitudes and longitudes were verified, planned gross power capacity was either discovered or inferred through public permitting data.

Since utilities rarely disclose facility power requests, we leveraged public construction permits and combustion generator permits to estimate power demand. Given that construction permits disclose planned square footage, we applied company-specific kW per square foot averages to derive total power capacity estimates. This methodology remains conservative, as power density in hyperscale facilities continues to increase. Regardless of a facility's intended clean energy footprint, every site incorporates diesel or gas generators as backup power. Although combustion generator permits are often filed later in development, they provide crucial insight into Uninterruptible Power System capacity, further refining our estimates of gross power needs.

Supply Model Methodology

For the power supply side of our model, our data science team analyzed an extensive dataset, combining publicly available and regulated developer-level grid data. This provided granular, nodal-level insights into current and historical generation, transmission, congestion, curtailment, outages, and pricing across all U.S. system operators and balancing authorities.

We developed a proprietary model to assess queued generation, retirements, and future generation mix, identifying when and where transmission and congestion constraints will restrict the ability to distribute power. All data was geospatially mapped to align with forecasted data center demand.

Kimmeridge

New York

15 Little West 12th Street,
4th Floor
New York, NY 10014

Denver

1401 Lawrence Street,
17th Floor
Denver, CO 80202

+1-646-424-4317
info@kimmeridge.com
www.kimmeridge.com