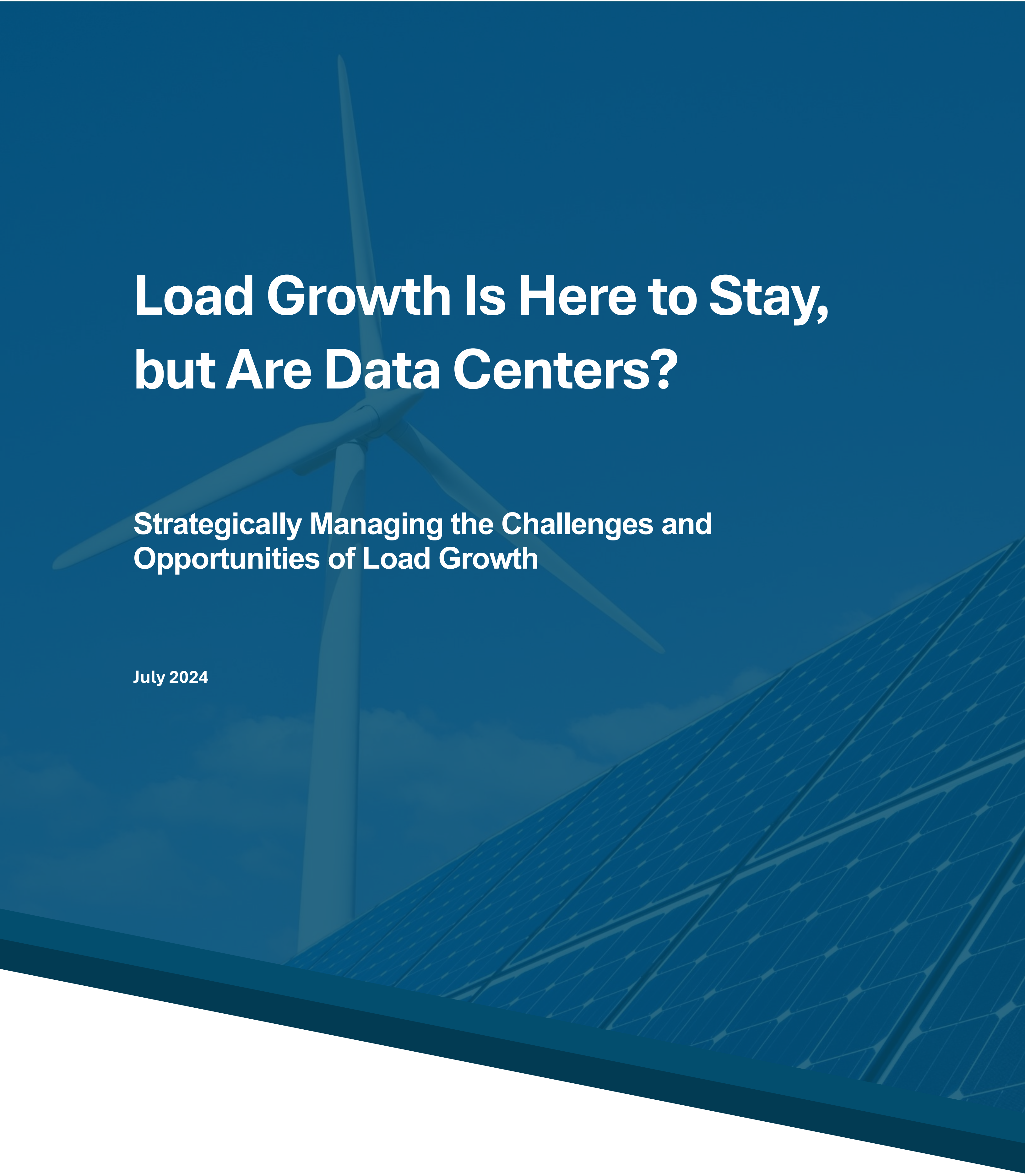


# Load Growth Is Here to Stay, but Are Data Centers?

## Strategically Managing the Challenges and Opportunities of Load Growth

July 2024

Energy+Environmental Economics

**This whitepaper is prepared by:**

Isabelle Riu, Senior Managing Consultant

Dieter Smiley, Senior Consultant

Stephen Bessasparis, Associate

Kushal Patel, Senior Partner

**Recommended Citation:**

I. Riu, D. Smiley, S. Bessasparis, K. Patel, "Load Growth Is Here to Stay, but Are Data Centers?: Strategically Managing the Challenges and Opportunities of Load Growth," Energy and Environmental Economics, Inc., July 2024. Available at: https://www.ethree.com/

# Executive Summary

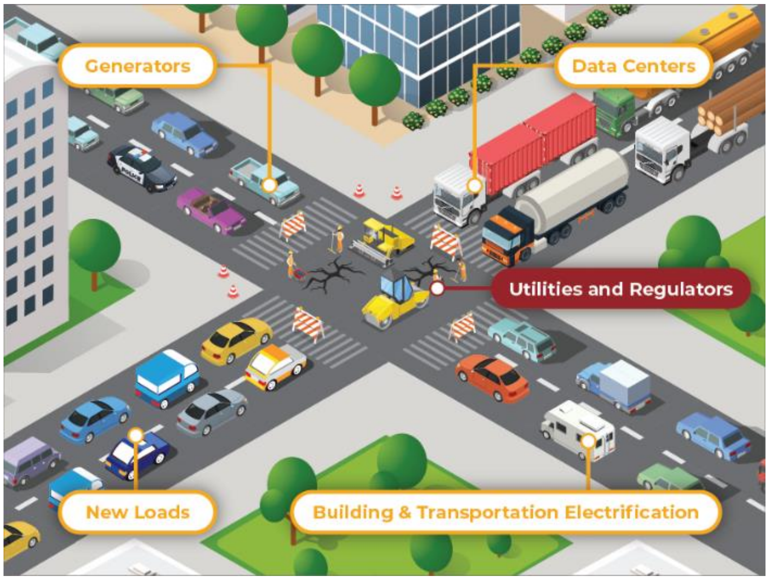
Until recently, the focus of the energy transition has primarily been on retiring legacy fossil generators and adding more renewables and energy storage that can sustain electrification-driven load growth in the longer-term. Now, rapid near-term load growth is underway, driven by large loads like data centers for artificial intelligence (AI) as well as a resurgence of U.S. industry due to industrial policy and manufacturing reshoring. This surge has surprised utilities and regulators across the country as they steer an aging grid through the challenges of an already ambitious energy transition. While the suddenness of these new large loads may seem unexpected, careful analysis highlights strategies to understand and mitigate risks as well as taking advantage of the opportunities they may enable.

Data center demand growth poses three primary challenges:

- Data centers are highly incentivized to interconnect as quickly as possible but face significant congestion and delays.
- Large new point loads can require substantial grid upgrades, forcing utilities to make potentially risky decisions about allocating scarce capital and managing ratepayer impacts.
- Data centers may consume large quantities of energy (both from existing and new electricity generators), which may challenge grid reliability if unmanaged.

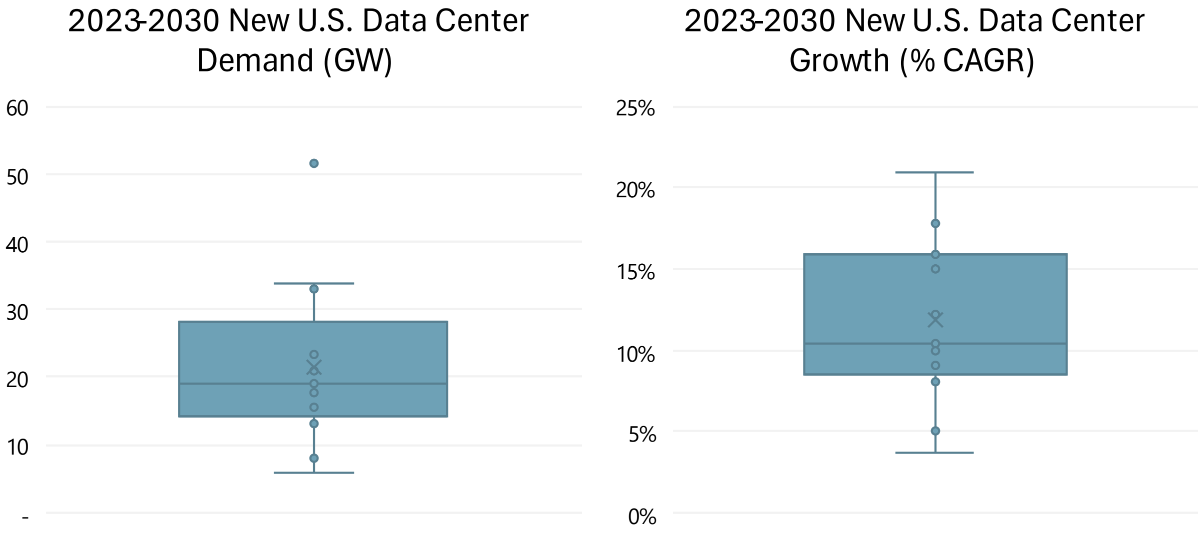
If the transforming grid is a traffic jam during highway construction, then data centers are a large convoy of trucks with urgent deliveries pulling into the on-ramp. This confluence of factors creates a gridlock, where utilities and regulators are overwhelmed working to modernize and decarbonize the grid, while managing queues of generators and new loads seeking interconnection, all bottlenecked at the same constraint. This may lead to suboptimal outcomes if grid decision makers only see limited near-term options such as delaying new large loads interconnections and/or delaying retirement of existing fossil fuel generators.

***Figure ES-1: "Gridlock"***



In this context, data center demand forecasts may be over-estimated, or "hyped." If this is the case, it would likely be from limits to interconnect the demand, not the volume of near-term demand itself. AI is the much publicized and discussed cause of recent data center growth, given the level of large investments being made by several technology companies. As can be seen in Figure ES-2, many data center demand forecasts reflect large growth over the next several years albeit over a wide range. This range reflects several uncertainties such the fundamental demand for more computing power (or "compute") as well as supply of data centers which can be constrained by available power.

***Figure ES-2: Range of Select Projections for U.S. Data Center Growth***[1]

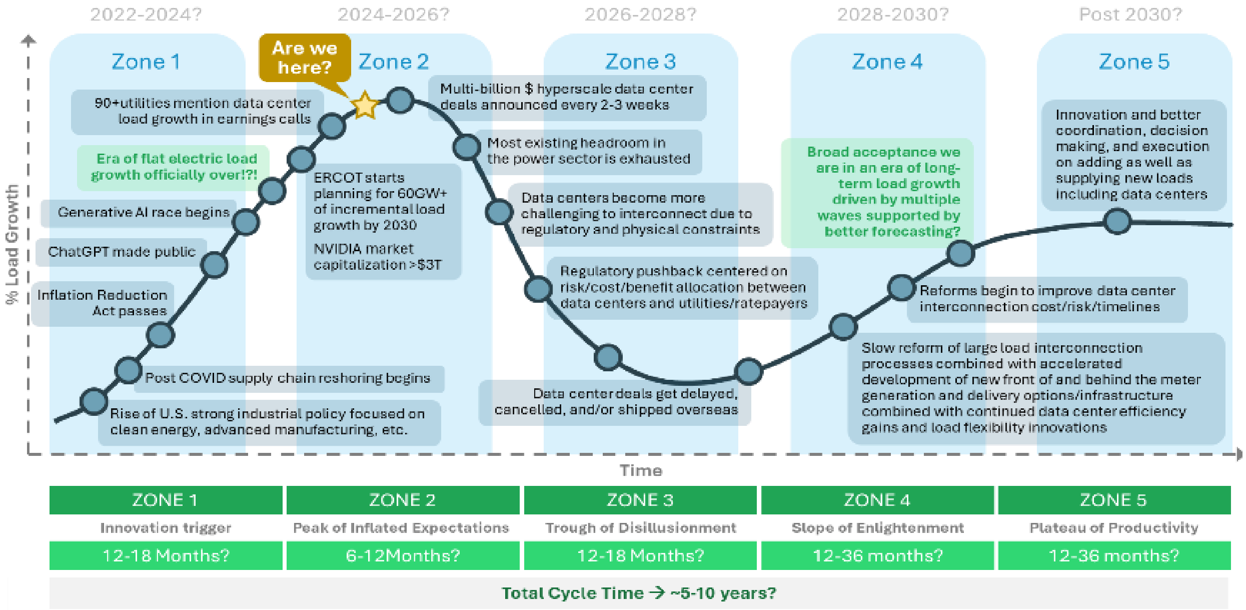


Even if the promise of AI falls short, general computing load is still likely to grow, due to factors like population growth and demographic shifts to more tech savvy generations, along with increased digitization of the economy. Figure ES-3 shows one way data center demand may change and evolve assuming we are currently near the top of a "hype" cycle.

---

[1] Projections from JLL, McKinsey, EPRI, IEA, BCG, Mordor and Goldman Sachs (total n = 13). E3 estimates data center capacity from energy estimates using an assumed 86% data center load factor and, as needed, linearly extrapolates projections to estimate changes from 2023 to 2030. BCG's "US Data Center Power Outlook" report issued in July 2024 provides its more updated view, projecting new data center demand growth ranging from 60 to 90 GW in 2023-2030. More detail provided in Appendix 1.

*Figure ES-3: Are we in a Power Sector Data Center Hype Cycle? Illustrative Visualization based on Gartner Hype Cycle*[2]

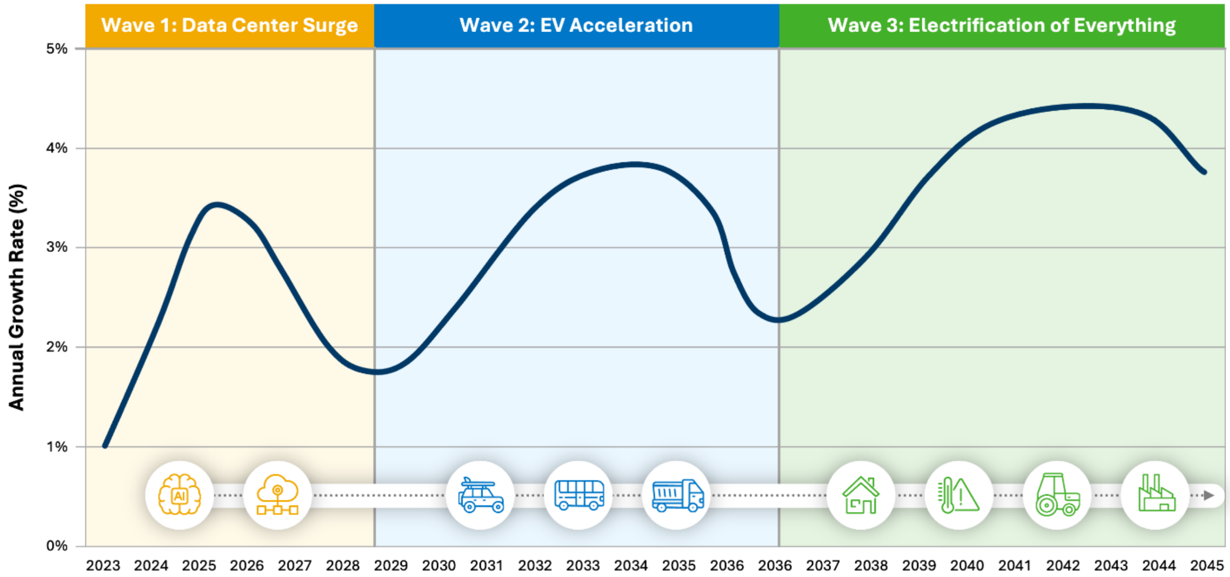


And if the near-term AI-driven load growth does flatten or even reverse, this is likely only the first wave of major U.S. load growth; the energy sector should prepare itself for the subsequent waves driven by strong industrial policies and electrification of transportation, buildings, and industry as seen in Figure ES-4. This is not the first time the U.S. power system has experienced this magnitude of demand growth, and we can learn from the past to make proactive decisions. A near-term rush of data center buildout and aggressive longer-term demand forecasts can put pressure on energy affordability and decarbonization efforts if not managed proactively. Establishing priorities is critical, and it will require all stakeholders to collaborate on demand- and supply-side solutions to avoid near-term unintended consequences and optimally capture long-term benefits. While data center load forecasts are inherently uncertain, uncertainty is no reason for paralysis nor a reason to avoid making proactive decisions.

---

[2] "Gartner Hype Cycle" Wikipedia.com. Accessed 21 June 2024. https://en.wikipedia.org/wiki/Gartner_hype_cycle

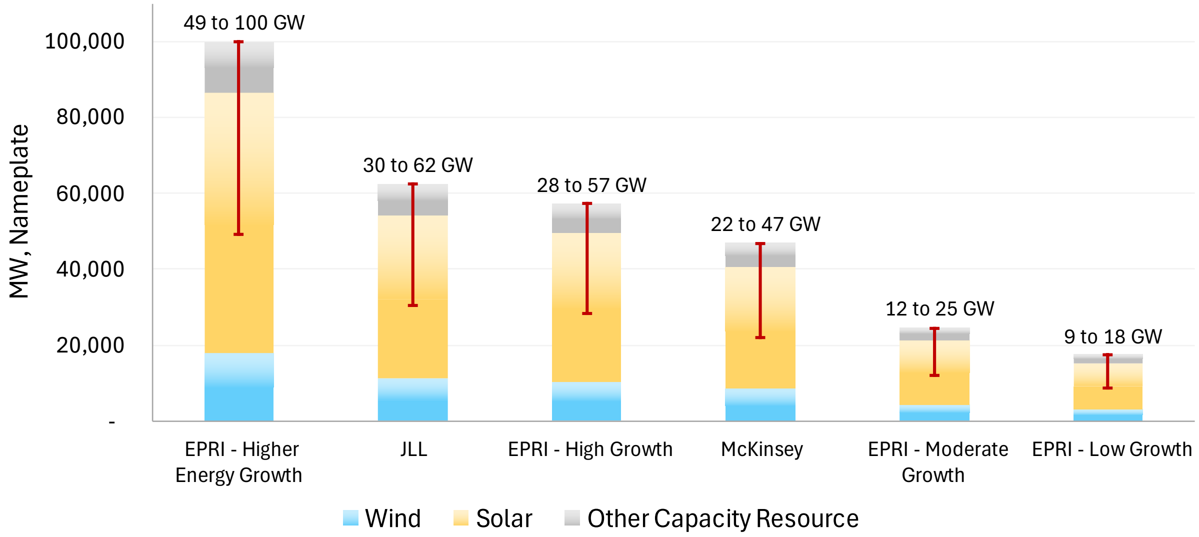
*Figure ES-4: Waves of Load Growth (Illustrative Load Growth)*



This paper seeks to ground the conversation around large load growth in some basic facts, offer historical context, and propose innovative ideas for large load developers, power industry planners, and investors to mitigate risks and take advantage of potential opportunities. We believe the decades-long work E3 has been doing on future load growth in the context of the energy transition, combined with a number of active engagements across our diverse client base, ranging from public sector regulators and agencies to utilities as well as private investors and developers, gives us a 360-degree understanding of the challenges, issues, and potential solution to help unjam the current gridlock between electric supply and demand.

There is still much uncertainty regarding the scope and scale of data center growth, but the key question should not be “How much will load grow?”, but instead, “Where and what kind of load growth can be accommodated?”. As we enter a “new build” era with multiple waves of load growth, planners must innovate and scale both demand and supply to navigate this evolving landscape effectively. Figure ES-5 shows initial, high-level E3 estimates on the level of new generation resources needed to meet the energy and grid reliability needs of data center demand, which can range from 20 GW to 100 GW of incremental new generation by 2030, reflecting the large uncertainty with demand and supply. It also includes error bars to indicate how uncertainty around potential energy efficiency improvements could impact builds, further emphasizing the need for adaptable resource planning.

*Figure ES-5: 2030 Resource Capacity with 75% Renewables to Meet Data Center Energy Demand with Varying Efficiency Improvements*[3]

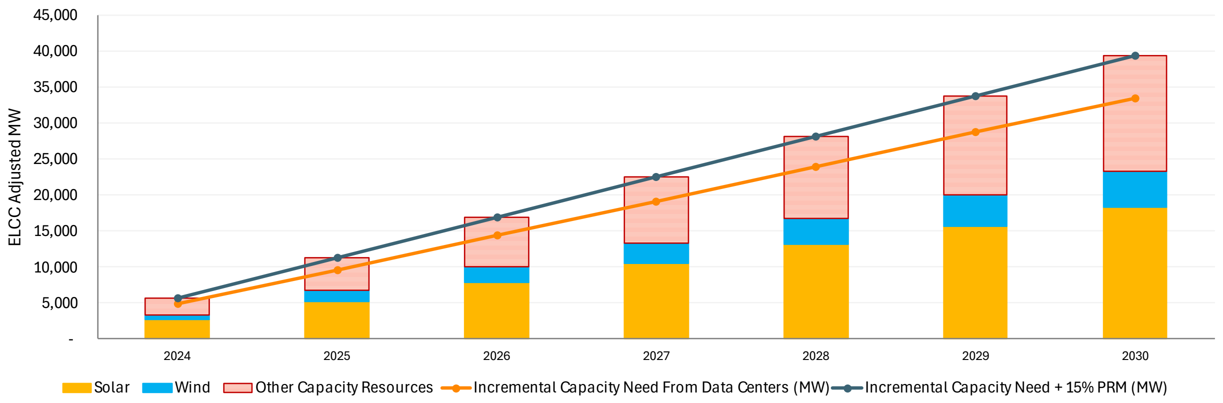


Electric grid planners and operators will need to ensure data center electricity needs are met reliably especially given many data centers need power supply at higher reliability standards than typical utility criteria, which means capacity resources will be required on top of resources that provide energy under average conditions to maintain service under peak (i.e. highest-need) conditions. Figure ES-6 illustrates E3's high-level analysis of what that need could be. We estimate that anywhere from 5 to 15 GW of additional capacity resources will be needed on top of assumed new renewables for reliability. These "other capacity" resources can take the form of currently commercial energy storage technologies, like lithium-ion batteries or pumped storage hydropower along with new peaking gas generation and customer demand response, as well as emerging technologies, such as long duration energy storage, low carbon fuels (such as hydrogen or renewable natural gas), enhanced geothermal systems, small modular nuclear reactors, and potentially others over time. Note that this analysis is for illustrative purposes using relatively simple heuristics and Appendix 1 provides additional detail on methodology.

---

[3] Uses EPRI Higher Energy Growth Scenario from "Powering Intelligence: Analyzing Artificial Intelligence and Data Center Energy Consumption." EPRI. 2024. https://www.wpr.org/wp-content/uploads/2024/06/3002028905_Powering-Intelligence_-Analyzing-Artificial-Intelligence-and-Data-Center-Energy-Consumption.pdf.

***Figure ES-6: Effective Capacity Contribution of Renewables and Other Capacity Resources to Meet Incremental Data Center Peak Demands***



Planning for load under uncertainty is nothing new, but the scale and speed of this load growth, combined with today's supply side constraints, is unprecedented. These unique circumstances require a new paradigm to avoid near-term unintended consequences and optimally capture long-term benefits. For example, data center load growth could be a positive for the industry if leveraged effectively. Well-resourced large baseload customers can help fund much-needed grid upgrades, support the adoption of emerging technologies, and drive new clean energy supply, potentially reducing costs for other customers and the system as a whole. For utilities and regulators, shifting away from traditional planning approaches to an integrated systems planning model would optimize existing resources, improve energy affordability, and support decarbonization efforts, all while enabling long-term strategic planning.

For more detailed information on proactive options and potential solutions see the "Options by Stakeholder" section of this paper. We provide a detailed set of options and their associated impacts on costs and risks, but this list is non-exhaustive. We expect each region in the U.S. to chart its own unique path in how best to manage near and longer-term load growth tailored to the local market structure and historical context. For example, large power users can manage energy needs through utility supply, self-generation, demand response, direct negotiation with generators, and infrastructure acquisition, while utilities and regulators can improve proactive planning, streamline interconnection, and implement cost-sharing and risk mitigation mechanisms to ensure grid reliability and affordability.

***Key Takeaway: Load growth is likely here to stay, even if the exact nature, timing, and scale is unclear. This means that utilities, regulators, and customers – both large and small – should proactively work together to realize the potential benefits and avoid the hazards of this new paradigm.***

*About Us:*

*E3 works on hundreds of projects a year exclusively in the energy sector for a diverse range of clients, ranging from public sector regulators and agencies, to utilities, to private investors and developers. We believe this broad work gives us a unique perspective on the challenges, issues, and potential solutions needed to address rapid load growth.*

*We have already incorporated data center impacts into E3's custom North American–wide PLEXOS market model to support investors, developers, utilities, and system operators. E3 has been working with a variety of clients on data center related issues such as supporting utilities on load forecasting, rate design, load interconnection process improvement, and resource planning related to data center growth. For big technology companies, data center companies, and various investors, E3 has advised and built in-house models to support both the siting and interconnecting data centers, procuring clean energy, and assessing power supply options including demand response.*

*For more insights into how E3 can support stakeholders across the industry on the impacts and opportunities presented by new large loads, email Kushal.Patel@ethree.com.*

The rest of this paper is organized into the following sections:

- Historical Context
- Scale and Shape of Demand
- Supply Challenges
- Options by Stakeholder
- Conclusion
- Appendix

# Historical Context

Load growth over the past ten years in the United States has been relatively flat, with a national peak power demand growth of only 0.5% annually.[4] However, in 2023 peak power demand growth sharply increased to 0.9%, driven by data centers and other large new loads.[5] Data centers were estimated to account for 4% of total US electricity consumption in 2023 and are expected to continue to grow, possibly up to 9.1% by 2030.[6] Grid planners are adjusting their forecasts accordingly. They have nearly doubled the U.S. 5-year load growth forecast (from 2.6% to 4.7%), and many expect a peak demand growth of 38 GW through 2028.[7] This would require rapid planning and buildout of new generation and transmission and could threaten the planned retirements of fossil fuel power plants if not executed quickly enough.

This cycle is not without precedent, however. The post-WWII era saw rapid load growth as economic prosperity and the population both surged, homes electrified, and new industrial manufacturing facilities centers grew out of wartime production. The subsequent decades had largely flat load, although there were pockets of regional high load growth driven by local manufacturing and/or population growth, such as in the Sunbelt, offset by de-industrialization and population loss in other regions. We are now seeing a return to a rapid growth era with the development of new digital industries along with advanced manufacturing and supply chain reshoring.

However, the landscape for growth is much different today, both in terms of the sheer volume of load growth being contemplated in absolute terms and today's more challenging environment to build large new infrastructure from a cost, regulatory, and timing perspective. Figure 1 shows the electricity usage growth rate averaged over 5-year periods to illustrate this historical context; grid planners and utilities have rapidly built out infrastructure in response to steep load growth in the past and need to revive these capabilities again.

---

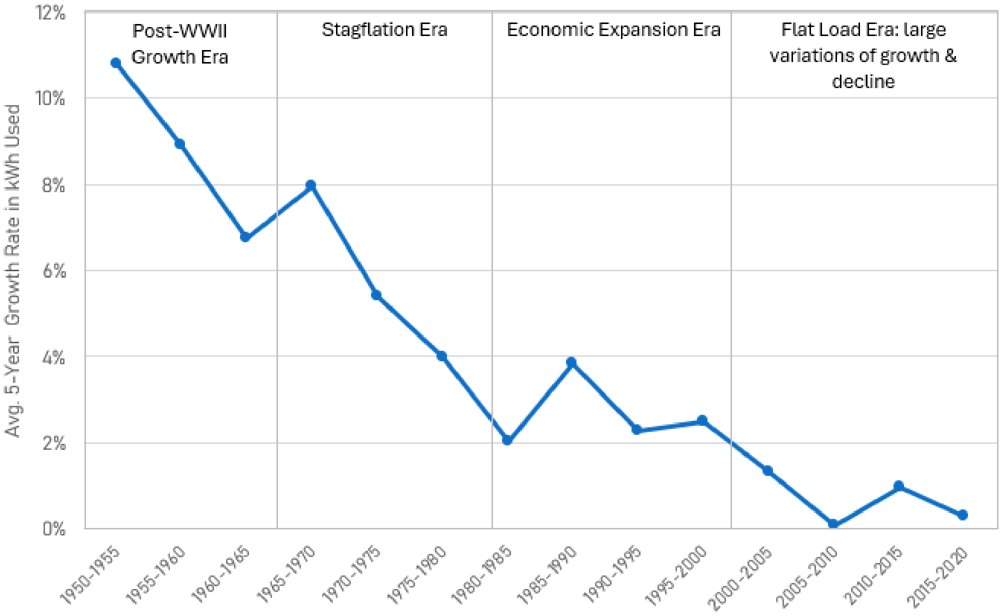
[4] John D. Wilson and Zach Zimmerman. "The Era of Flat Power Demand is Over." GridStrategies. December 2023. https://gridstrategiesllc.com/wp-content/uploads/2023/12/National-Load-Growth-Report-2023.pdf

[5] John D. Wilson and Zach Zimmerman. "The Era of Flat Power Demand is Over." GridStrategies. December 2023. https://gridstrategiesllc.com/wp-content/uploads/2023/12/National-Load-Growth-Report-2023.pdf

[6] "Powering Intelligence: Analyzing Artificial Intelligence and Data Center Energy Consumption." Electric Power Research Institute. 2024. https://www.epri.com/research/products/3002028905

[7] John D. Wilson and Zach Zimmerman. "The Era of Flat Power Demand is Over." GridStrategies. December 2023. https://gridstrategiesllc.com/wp-content/uploads/2023/12/National-Load-Growth-Report-2023.pdf

*Figure 1: 5-Year Avg. Growth Rate in Electricity Usage 1950-2020*[8]



The initial decades of digitally-driven load growth were not as large as initially anticipated due to efficiency gains and microchip development trends in line with Moore's Law, i.e. the observed doubling of transistors in an integrated circuit roughly every two years. Subsequently, the average power usage effectiveness[9], a measure of the inefficiency of transforming electrical energy into server processing time, in U.S. data centers decreased from an average of 2.5 in 2007 to about 1.5 in 2022. This has tempered a demand spike from data centers, i.e. more compute for less energy.

It is unclear if these trends will continue through the current phase of data center construction, but observers note that gains in data center efficiency have slowed in recent years.[10] This trend, combined with the unique demands of AI data centers, could lead to new demand significantly exceeding future efficiency gains and driving a stark increase in system-wide electrical load.

> ***The potential scale of this new load is bound by the system's ability to supply power and the ability of demand to effectively use said power.***

---

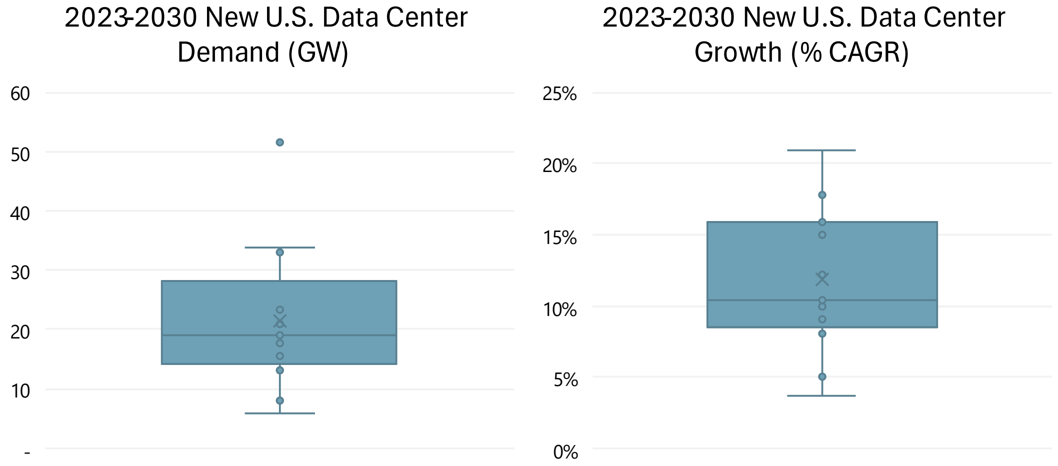
[8] "Monthly Energy Review." U.S. Energy Information Administration. May 2024. https://www.eia.gov/totalenergy/data/monthly/

[9] Power Usage Effectiveness (PUE) is a data center industry-preferred metric that represents the infrastructure energy efficiency for data centers. It divides the facility's total energy usage by the IT equipment's energy usage. A lower PUE indicates a more efficient data center using less energy to run secondary functions like cooling.

[10] Daniel Bizo. "Global PUEs – are they going anywhere?". 04 December 2023. https://journal.uptimeinstitute.com/global-pues-are-they-going-anywhere/

Estimates vary, but most agree the potential scale of data center demand is extremely large. Many observers forecast at least 15 GW to 30 GW of new data center demand will be added to the U.S. system by 2030, with a theoretical upper limit[11] of 70 GW based on worldwide microprocessor fabrication limitations.[12] The growth of AI is expected to be a major driver; AI has represented half of data center power demand growth since 2016, and this share is predicted to increase through 2030.[13]

*Figure 2: Range of Select Projections for U.S. Data Center Growth[14]*



AI energy demand can be categorized into the two major phases of an AI's lifetime: training and utilization (also known as inference).[15] During the training phase, the AI program is digesting vast amounts of data to build the associations needed for the model to work. This typically has consistently high power requirements. During the utilization phase, the completed model is responding to user queries and performing its actual task. The exact scale of this growth depends on several independent factors in the energy-to-AI value chain, which is a multi-step process of transforming energy into compute and ultimately into completed AI tasks with economic value. Forecasts are sensitive to changes in these factors, as a modification to a step in the process, such as an efficiency improvement, can have a significant impact on total energy demand.

---

[11] As NVIDIA's servers could be sold anywhere globally, the upper end of the forecast assumes the US has the lion's share of the growth.

[12] Alex de Vries. "The growing energy footprint of artificial intelligence." Cell: Joule. 10 October 2023. https://doi.org/10.1016/j.joule.2023.09.004A

[13] John D. Wilson and Zach Zimmerman. "The Era of Flat Power Demand is Over." GridStrategies. December 2023. https://gridstrategiesllc.com/wp-content/uploads/2023/12/National-Load-Growth-Report-2023.pdf

[14] Select projections from JLL, McKinsey, EPRI, IEA, BCG, Mordor and Goldman Sachs, including low, medium, and high scenarios (total n = 13). E3 estimates data center capacity from energy estimates or vice versa using an assumed 86% data center load factor and, where applicable, linearly extrapolates projections to estimate changes from 2023 to 2030. BCG's "US Data Center Power Outlook" report issued in July 2024 provides its more updated view, projecting new data center demand growth ranging from 60 to 90 GW 2023-2030. More detail provided in Appendix 1.

[15] Michael Copeland. "What's the Difference Between Deep Learning Training and Inference?". NVIDIA. 22 August 2016. https://blogs.nvidia.com/blog/difference-deep-learning-training-inference-ai/

## The Impact of Energy Efficiency on Demand Growth

Each of the steps in the energy-to-AI value chain has an associated transformation efficiency, with both physical factors (e.g., power plant heat rate) and financial factors (e.g., the cost of acquiring new data center capacity) acting on those efficiencies. For example, an increase in microchip energy or cooling efficiency would enable more compute to be extracted from an energy input, decreasing the energy needed to run the same amount of AI capacity. Conversely, creating a new AI model that can accomplish new, high-value tasks (e.g., interpreting radiological scans in healthcare or mass consumer adoption of AI assistants) would incentivize greater production of AI capacity with associated increases in energy demand.

Toggling just the variable of efficiency improvements can have significant impacts on the total energy needed to meet this new AI-driven demand. Suppose NVIDIA's recently announced Grace CPU Superchip, which reportedly consumes 50% less power than other chips of its type, becomes the new standard for efficiency in transforming energy into compute.[16] If compute is roughly half of a data center's energy consumption, this breakthrough may reduce 20 GW of anticipated new demand down to 15 GW and obviate the need for potentially one-third of new solar additions based on E3 analysis. Figure 3 shows a wide range of potential data center load growth trajectories (normalized to today's data center demand levels) across several high-level scenarios focused on efficiency improvements.[17]

> ***There could be a wide range of demand growth outcomes solely on the variable of efficiency improvements, with the low end still significant at 50% growth from today's level.***

***Figure 3: Illustrative Data Center Energy Demand Growth Under Efficiency Improvement Scenarios***

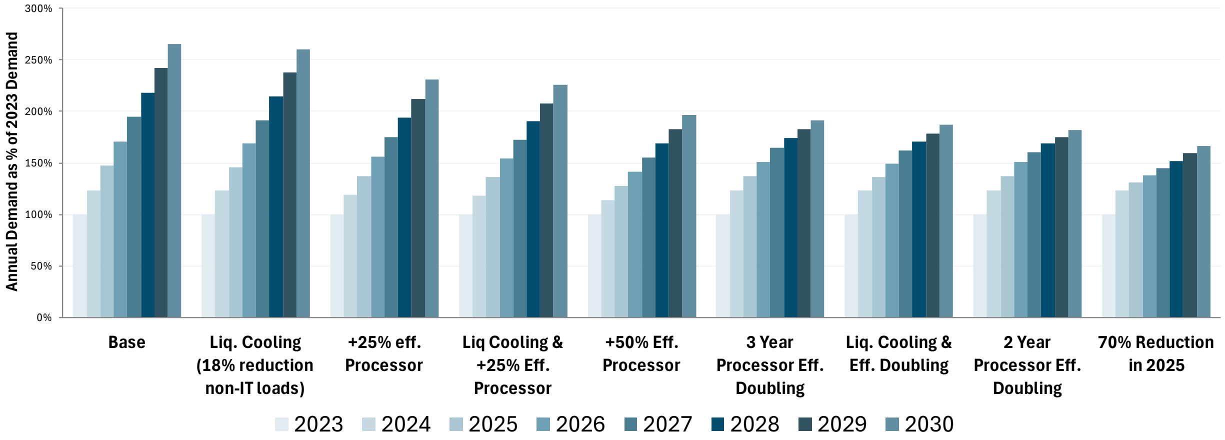
| Scenario | Description |
|---|---|
| **No Assumed Efficiency** | No incremental efficiency. Assumes energy demands consistent with EPRI's U.S.-wide Higher Energy Growth projections (cumulative 250 TWh of new energy demand by 2030) and power usage effectiveness (PUE) of 1.2 |
| **Liquid Cooling (10% Reduction)** | Liquid Cooling, rather than air cooling, is assumed to result in 10% facility wide energy reductions. Applied to all years |
| **+25% Efficiency Processor Improvement** | 25% improved efficiency is assumed for processing power. Applied to all years |

---

[16] Ivan Goldwasser. "Green Light: NVIDIA Grace CPU Paves Fast Lane to Energy-Efficient Computing for Every Data Center." NVIDIA. 21 March 2023. https://blogs.nvidia.com/blog/grace-cpu-energy-efficiency/https://blogs.nvidia.com/blog/grace-cpu-energy-efficiency/

[17] E3 modeled incremental efficiency gains on top of the projected energy amounts. As a result, any energy efficiency measures that may have already been included in the source material were not taken into account.

| | |
|---|---|
| **Liquid Cooling and +25% Efficiency Processor Improvement** | 25% improved efficiency is assumed for processing power and subsequent 10% efficiency gain is applied facility wide. Applied to all years |
| **+50% Efficiency Processor Improvement** | 50% improved efficiency is assumed for processing power. Applied to all years |
| **3 Year Processor Efficiency Doubling** | Processor efficiency is assumed to double every 3 years (first improvement in 2025) |
| **Liquid Cooling and Efficiency Doubling** | Processor efficiency is assumed to double every 3 years (first improvement in 2025) and subsequent 10% efficiency gain is applied facility-wide |
| **2 Year Processor Efficiency Doubling** | Processor efficiency is assumed to double every 2 years (first improvement in 2025) |
| **70% Reduction in 2025** | Facility wide energy reductions of 70% assumed beginning in 2025 |

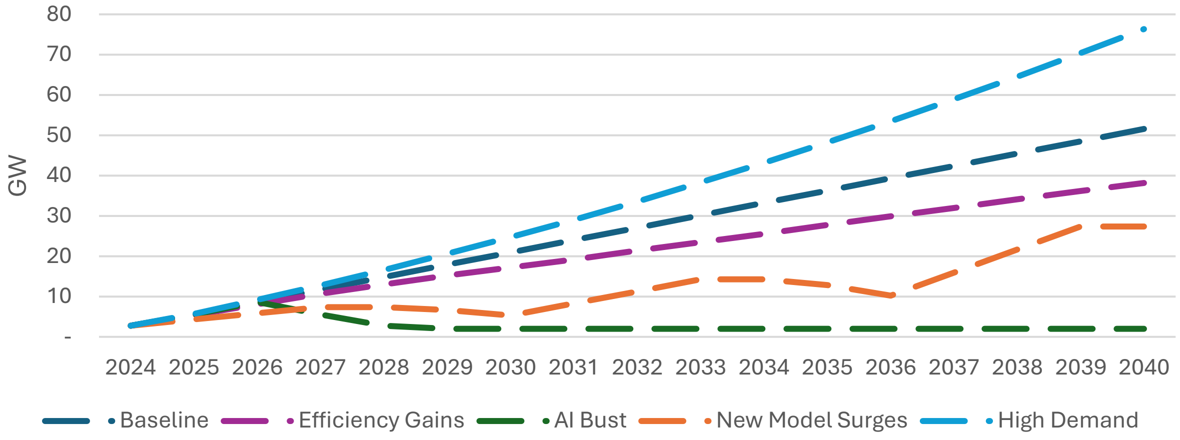


## The Impact of the AI Use Case on Demand Growth

This framing also illustrates the main incentive behind AI data center load growth: the value of the tasks AI can accomplish. In an efficient market, the cost of acquiring the fuel, energy, compute, and AI capacity to create an AI model should not exceed the value of the tasks it can accomplish. Theoretically, this limits the amount of energy to be dedicated to the AI. For example, consider the value of interpreting radiological scans. There are approximately 32,000 radiologists in the U.S. with a mean annual wage of $354,000. If AI replaced 10% of the value of their labor, then that AI would have a value of about $1.1 billion. The cost of the supporting inputs, including the infrastructure needed to produce and deliver the electrical energy, would be significantly less than this for the creation of that AI to make economic sense. Conversely, if AI demand flatlines or even declines due to fundamental challenges with the technology, such as a barrier to further AI model development, then flatline or declining growth could occur, especially if the near-term demand represents a “boom” or overinvestment cycle to "win" the AI race. Estimating the value of AI tasks and the costs of each step of the conversion process can provide a fundamentals-based estimate for the amount of energy dedicated to AI.

These different factors informing how AI is used in society could play out in a range of different growth trajectories as illustrated in Figure 4. For example, if there are high demand drivers consistent with discovering more profitable uses for AI, we would expect data center load growth to increase over time. In contrast, if market saturation decreases demand, marginal inputs become prohibitively expensive, or the load becomes increasingly flexible (by moving computing loads in time and/or utilizing data center on-site generation), we would expect load growth to slow over time. If a significant portion of load comes from the training phase of AI model development and new models are trained on a periodic basis, then we would expect periodic surges in demand to train new energy-intensive models followed by relative troughs as those models are utilized by consumers. Finally, if AI turns out to not have very many market applications and stops improving, then we would expect the load growth from AI data centers to drop back down to pre-2023 baselines. In short, there is a wide range of potential load growth shapes hinging on how AI evolves and is used, creating a large cone of uncertainty and it is important to consider these factors when developing data center load forecasts.

***Figure 4: Projected AI Load Growth Under Various AI Growth Scenarios***

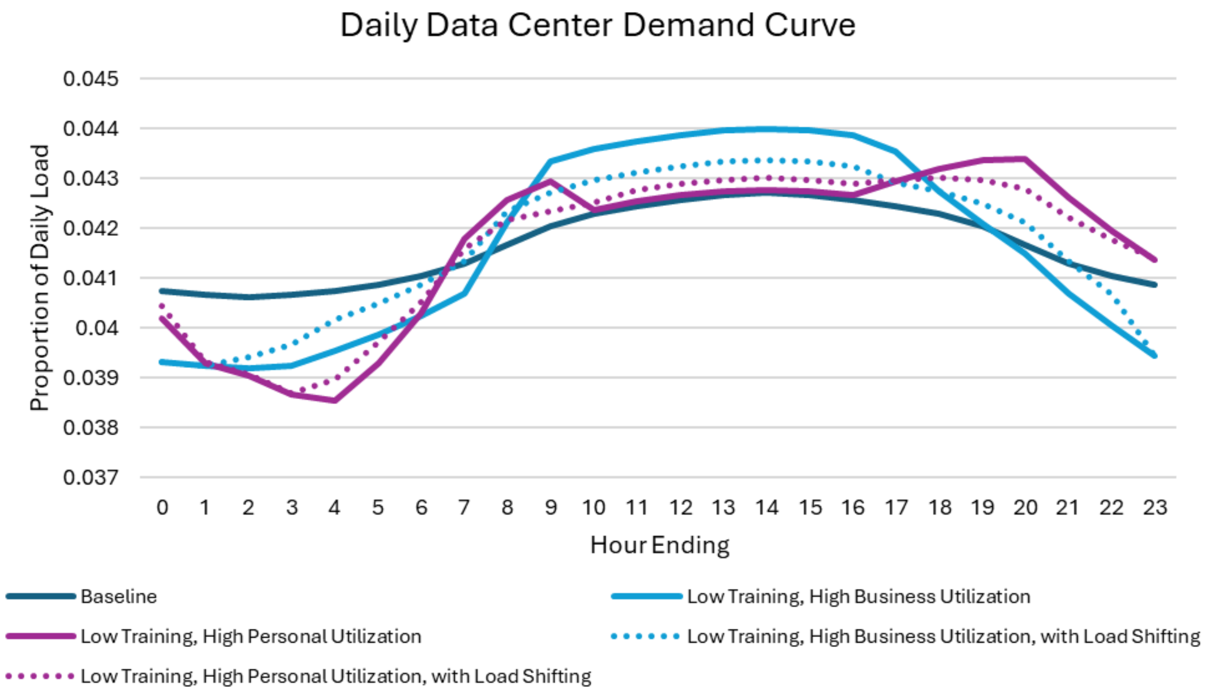


## The Role of AI's Daily Use Case on Demand Growth

The use case and level of demand for AI may also affect the shape of the daily data center load curve. Data centers are capital intensive, from building sites to buying servers, which incentivizes the facilities to run at high utilization, which is also aligned to the underlying business need. Currently, data centers, which are high load factor i.e. mostly baseload facilities, have a relatively flat shape, reflecting baseload computing needs, but also have some seasonal variation due to significant weather-dependent cooling needs. This load shape could continue if new AI models are in constant development and therefore the majority of AI data center load is dedicated to model training.

However, if usage overtakes training as the dominant load source, then the daily peak would be more dependent on usage time and type. If AI is primarily a business tool, then peak demand may mostly coincide with business hours. If AI is mostly a personal tool, then twin morning and evening peaks may be more likely, resembling today's residential load shapes. Figure 5 illustrates these possible load shapes. There may be additional possible load shapes that reflect having more flexibility around AI computing, such as batching AI queries and running them flexibly during the day, moving other computing loads optimized around variables such as clean energy availability, and utilizing on-site behind-the-meter generation. Similar to the overall load growth trajectories to 2040, there is substantial uncertainty around future load shape, but it is unlikely to be truly flat, which has significant ramifications for grid planners and system operators on how to serve this demand.

***Figure 5: Projected AI Daily Load Curve Under Various Usage Scenarios***



Geographically, this demand is and will likely continue to be highly uneven. In 2023, 80% of national data center load was concentrated in 15 states which can lead to localized grid stress.[18] These clusters occur because developers are attracted to areas that have large population centers, strong internet connections, low electricity and land costs, potentially strong economic development policy incentives, skilled labor forces, and/or low disaster risk. But if primary markets saturate, development prices increase, and local community pushback grows, then new builds may transfer to other markets. These new markets may eventually see a significant proportion of their electricity

---

[18] "Powering Intelligence: Analyzing Artificial Intelligence and Data Center Energy Consumption." EPRI. 2024. https://www.wpr.org/wp-content/uploads/2024/06/3002028905_Powering-Intelligence_-Analyzing-Artificial-Intelligence-and-Data-Center-Energy-Consumption.pdf

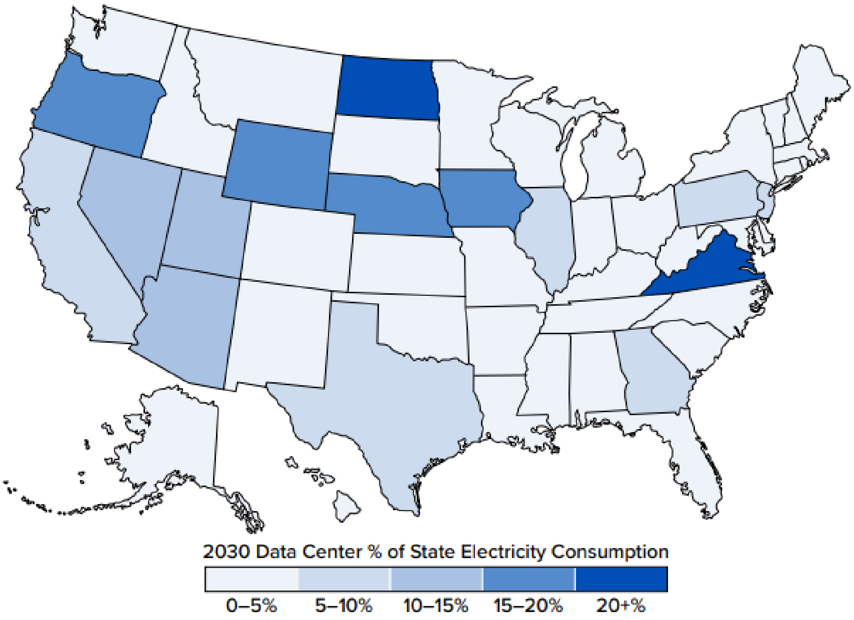
generation consumed by data centers, especially if the markets already had relatively low electricity consumption levels.

---

> ***Data centers tend to operate in discrete geographic markets; the eight primary markets contain five times the data center volume as the eight secondary markets, and the top market, Northern Virginia, contains half of all primary market data center capacity.*** [19]

---

### *Figure 6: EPRI's Projected Data Center Share of Electricity Consumption in 2030*[20]



[21]

In sum, data center demand is forecasted to be extremely large, but the exact scale, shape and geographic distribution of the growth are uncertain and depend on a number of variables. Planning for load under uncertainty is nothing new but the scale and speed of this new demand class combined with historic supply side constraints are unprecedented and have exacerbated the power system's gridlock.

---

[19] "North America Data Center Trends H2 2023." 06 March 2024. CBRE. https://www.cbre.com/insights/reports/north-america-data-center-trends-h2-2023

[20] "Powering Intelligence: Analyzing Artificial Intelligence and Data Center Energy Consumption." EPRI. 2024. https://www.wpr.org/wp-content/uploads/2024/06/3002028905_Powering-Intelligence_-Analyzing-Artificial-Intelligence-and-Data-Center-Energy-Consumption.pdf

[21] "Powering Intelligence: Analyzing Artificial Intelligence and Data Center Energy Consumption." EPRI. 2024. https://www.wpr.org/wp-content/uploads/2024/06/3002028905_Powering-Intelligence_-Analyzing-Artificial-Intelligence-and-Data-Center-Energy-Consumption.pdf

While the U.S. generation system is robust, new data center load is growing during a time of major transition. E3 has developed high level analyses to illustrate how this new demand could be met amidst today's supply challenges by contextualizing the load against existing grid capacity and illustrating potential generation buildouts, reflecting historic thermal retirements, varying renewable energy goals, and reliability constraints. Appendix 1 provides additional detail on methodology but the purpose of these simplified calculations is to illustrate the potential scale of buildout needed and to underscore important planning considerations to serve this demand reliably and cleanly.

## Utilizing Existing Grid Headroom

A key metric for the system's ability to absorb new load is the system's projected headroom. In this paper, headroom is defined as the difference between the grid's hypothetical generation potential and grid demand. It exists for a variety of reasons, but mostly reflects the margin needed to provide electricity at least cost while maintaining reliability.[22]

From a power perspective, the grid has a peaking headroom of approximately 100 GW. The 20 GW of new data center load alone would consume a significant portion of this headroom, but with baseline growth, the expected total additions of 38 GW lay an even heavier burden on the system.[23] This would occur during a time when firmer thermal generation is being retired in favor of cleaner, but more intermittent renewable generation. How exactly the headroom need is changing to reflect adding more intermittent renewables and batteries (which have lower reliability value compared to nameplate capacity) combined with lower levels of fossil fuel generators (which usually having higher reliability value compared to nameplate capacity) in the face of more extreme weather events is outside the scope of this paper but represents another important variable. Figure 7 and Figure 8 illustrate key scenarios, and additional sensitivities and details can be found in the appendix. Note that this analysis is for illustrative purposes using relatively simple heuristics based on E3 work to show potential impacts to headroom.

---

[22] Our calculations assume thermal fossil plants have an average forced outage rate of 10%, and therefore with sufficient demand could operate at 90% capacity factor.

[23] John D. Wilson and Zach Zimmerman. "The Era of Flat Power Demand is Over." GridStrategies. December 2023. https://gridstrategiesllc.com/wp-content/uploads/2023/12/National-Load-Growth-Report-2023.pdf

***Figure 7: EIA Forecast Power Headroom by 2030 Under a Static 15% Planning Reserve Margin (PRM)***

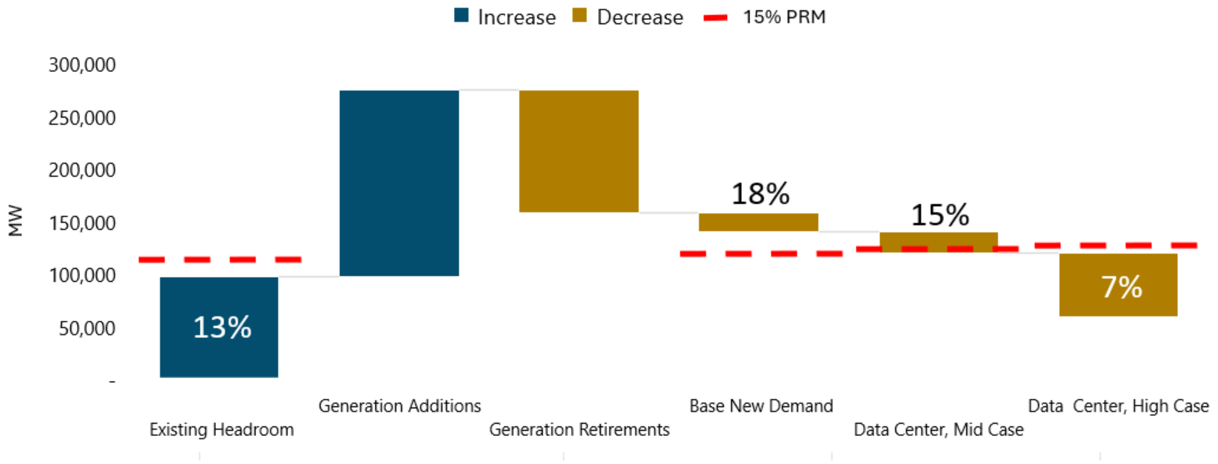
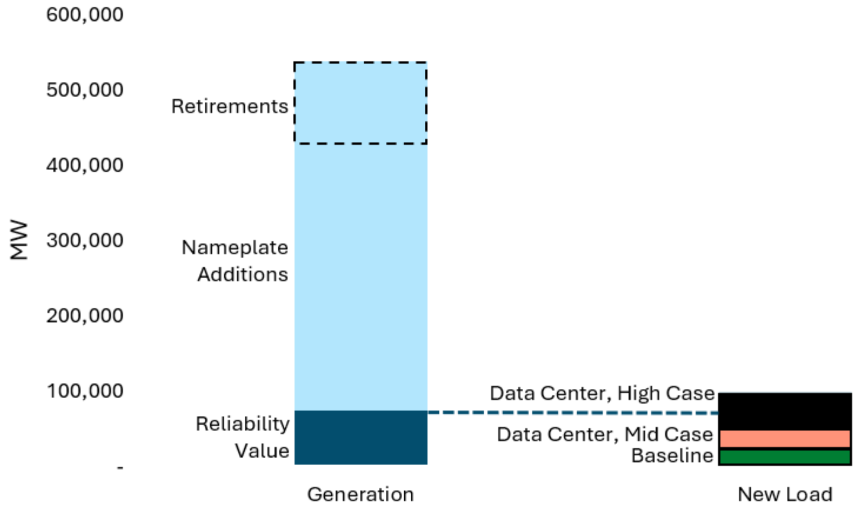


***Figure 8: EIA Forecast Generator Additions and New Load by 2030***



Planned additions will exceed planned load and increase overall headroom, but this is contingent upon sufficient system energy storage being able to shift intermittent renewable generation to peak times. Without long duration energy storage to enable a clean and reliable stream of power from these renewable additions and/or breakthroughs in clean firmer generation technologies (e.g., small modular nuclear reactors, advanced geothermal, and others), finding firmer and relatively cheap power for data centers that are reluctant to curtail operations during peak reliability periods may become complex. This increases the likelihood that data center load will need to be met with additional renewable generation with short-duration batteries. If this generation is not present, data centers may opt to build generation on-site or grid planners may need to delay fossil fuel generator retirements.

To meet this demand cleanly without reducing existing headroom, the system would need to add about 57 GW of solar and 15 GW of wind of nameplate capacity along with 10 GW of effective capacity from other technologies like energy storage or gas (this analysis uses basic calculations outlined in more detail in Appendix 1 and is meant for illustrative purposes). Meeting 75% of this energy demand with renewables would require 42 GW of solar and 11 GW of wind nameplate capacity with almost 14 GW of other effective capacity. While 61 GW of wind and 66 GW of solar were added to the grid in the past six years, thermal plant retirements exceeded thermal additions as aging fossil plants continue to shut down**.**

> ***In 2023, 16 GW of capacity was retired from the grid – more than the lower estimate of data center load additions by 2030.***

The overall conclusion is that headroom will likely be challenged by 2030, but there is a wide range of potential outcomes given 1) the uncertainty around future demand; 2) the uncertainty on the ability to keep pace with that demand with new generation additions that will predominantly be intermittent renewables and energy storage; and 3) the uncertainty around retiring existing generation (e.g. coal) as scheduled or even under an accelerated schedule.

## Potential New Generation Buildouts

E3 performed analyses to illustrate the potential new resource builds required to meet incremental annual energy demands driven by data center development as informed by growth projections from Jones Lang Lasalle Incorporated (JLL),[24] McKinsey & Company (McKinsey),[25] and Electric Power Research Institute (EPRI).[26] Using simplified assumptions[27] to account for a mix of new renewable resources meeting 75% of incremental energy demands, E3 demonstrates that by 2030, required new builds could range from 20 GW to 100 GW of additional generation capacity, translating to 3 to 17 GW of new generator additions per year. Over the past six years, a record high deployment of renewables has been achieved, with an average of 21 GW of wind and solar being added annually. Our estimated range of required generators highlights the significant uncertainty in projected energy demand and the challenges of scaling supply to accommodate new data center growth, other load

---

[24] Kari Beets. “North America Data Center Report.” JLL. 28 February 2024. https://www.us.jll.com/en/trends-and-insights/research/na-data-center-report

[25] Srini Bangalore, Arjita Bhan, Andrea Del Miglio, Pankaj Sachdeva, Vijay Sarma, Raman Sharma, and Bhargs Srivathsan. “Investing in the rising data center economy.” McKinsey & Company. 17 January 2023. https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/investing-in-the-rising-data-center-economy

[26] “Powering Intelligence: Analyzing Artificial Intelligence and Data Center Energy Consumption.” EPRI. 2024. https://www.wpr.org/wp-content/uploads/2024/06/3002028905_Powering-Intelligence_-Analyzing-Artificial-Intelligence-and-Data-Center-Energy-Consumption.pdf

[27] New renewables generation is comprised of 70% solar and 30% wind, with respective capacity factors of 22% and 36%; generic gas generation assumes a 54% capacity factor. For public reports that provide data center projections in terms of capacity (MW), energy demand is estimated assuming consumption profiles have a load factor of approximately 86%. More information in Appendix.

growth and replacing retiring fossil fuel units. Note these illustrative estimates depicted in Figure 9 only consider balancing energy demands with the generation potential of new resource builds.

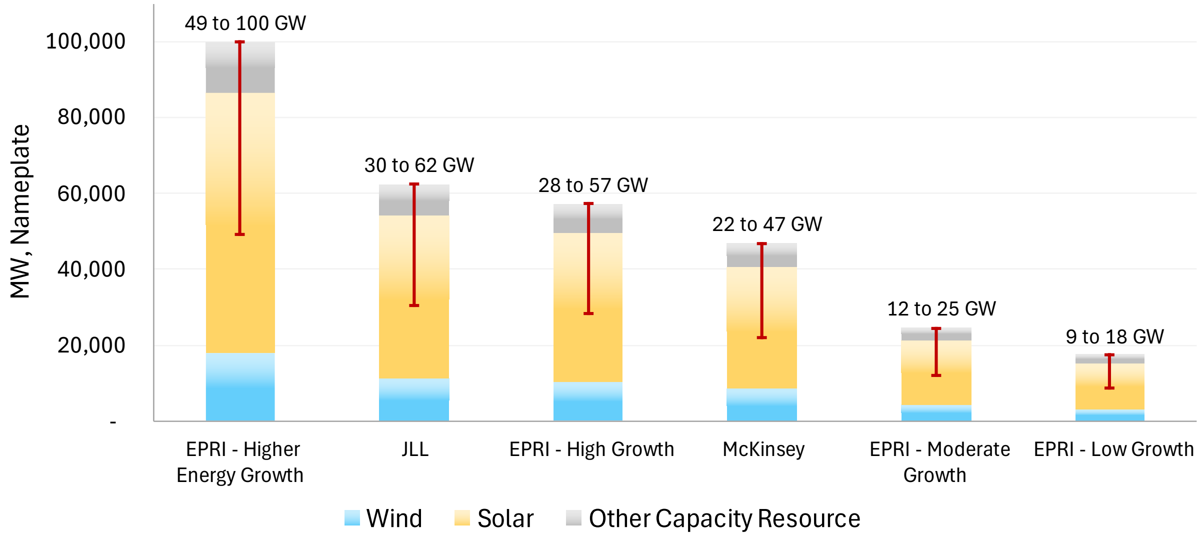
E3 estimated new resource builds under a range of sensitivities, examining lower energy demands from incremental energy efficiency gains to computing and cooling data center operations. Figure 9 shows errors bars to indicate the range of uncertainty based on E3's hypothetical energy efficiency scenarios outlined in Figure 3. The efficiency analysis indicates that significant advances in data center operations could result in a large range of builds, further emphasizing the need for adaptable resource planning.

---

***The sensitivities applied are illustrative but exhibit the additional uncertainty in forecasting load requirements of data centers in a sector uniquely sensitive to hardware and software technology improvements as well as rapidly shifting business models and demand drivers vs. other more "traditional" industries.***

---

### *Figure 9: 2030 Resource Capacity with 75% Renewables to Meet Data Center Energy Demand with Varying Efficiency Improvements*[28]



In addition to ensuring a sufficient annual energy supply, maintaining grid reliability at all times remains a crucial factor in grid operations as well as in resource and reliability planning. Addressing resource adequacy and reliability across various markets may necessitate different combinations of transmission, distribution, and generation builds and capacities, especially firmer resources, to meet projected growth in annual energy consumption and peak demand. A significant increase in

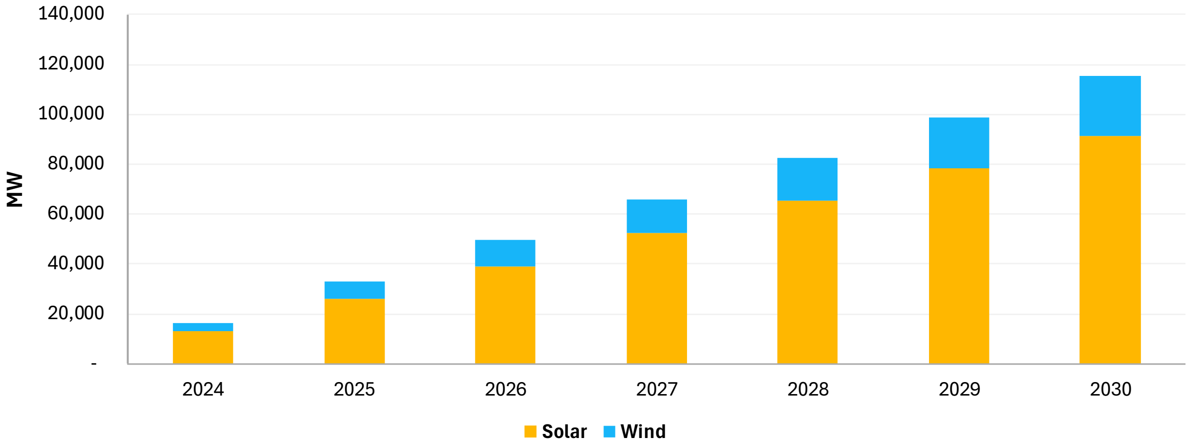
---

[28] Uses EPRI Higher Energy Growth Scenario

large high load factor demands, such as those from data centers, could further intensify build requirements, underscoring the need for proactive and comprehensive resource planning processes.

Building from the energy supply analysis above, E3 demonstrates the Effective Load Carrying Capability (ELCC)[29] (i.e. reliability or capacity value) of renewables needed to meet 100% of projected annual energy demands under EPRI's "Higher Energy Growth" scenario. This analysis highlights the potential magnitude and mix of resources required to maintain acceptable levels of reliability. To meet 100% of the incremental energy demand for data centers, 115 GW of nameplate renewables capacity would need to be built by 2030 as illustrated in Figure 10. However, to ensure reliable service, an additional 15 GW of additional firm capacity would still be necessary.

*Figure 10: Renewables Nameplate Capacity to Meet 100% of Incremental Data Center Energy (EPRI - Higher Energy Growth)*

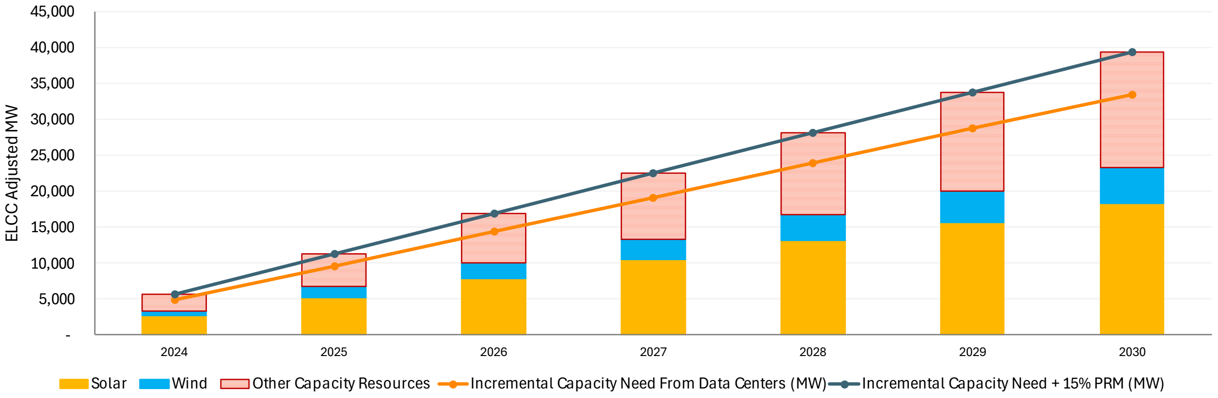


Using static ELCC assumptions for solar and wind to estimate each technology's contribution to grid reliability, E3 estimated that the effective capacity contribution of renewables in 2030 would be nearly 23 GW, as shown in Figure 11. From EPRI's projected energy demands, E3 estimates the capacity needs of new data centers assuming an 86% load factor and a 15% planning reserve margin. The remaining 16 GW gap to meet estimated capacity requirements indicates the need to consider other capacity resources in planning efforts to maintain grid safety and reliability, whether that be energy storage, geothermal, nuclear, demand response, or thermal resource options.

---

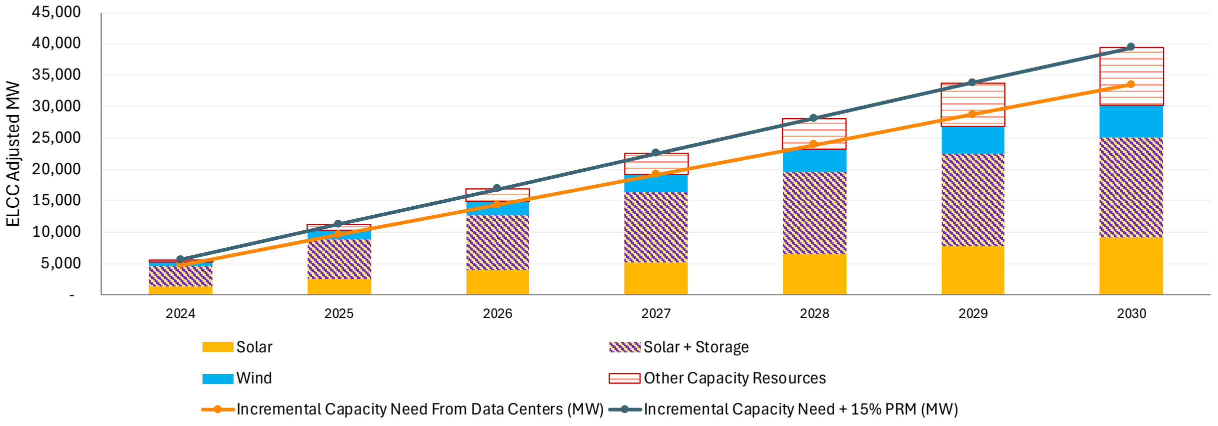
[29] See here for more information on ELCC: [https://www.ethree.com/wp-content/uploads/2020/08/E3-Practical-Application-of-ELCC.pdf](https://www.ethree.com/wp-content/uploads/2020/08/E3-Practical-Application-of-ELCC.pdf)

### *Figure 11: Effective Capacity Contribution of Renewables and Other Capacity Resources to Meet Incremental Data Center Peak Demands*



If half of the built solar nameplate capacity in Figure 11 is assumed to be paired with short duration storage,[30] renewables become much more effective in their reliability contributions, diminishing the need for incremental capacity resources. As illustrated in Figure 12, the effective capacity need from other capacity resource declines more than 40% from Figure 11 when half of the solar resources are assumed to be paired with short-duration storage.

### *Figure 12: Effective Capacity Contribution of Renewables with Storage and Other Capacity Resources to Meet Incremental Data Center Peak Demands.*



Transmission and distribution infrastructure limitations may further complicate efforts to meet this new load. According to the Department of Energy's National Transmission Needs Study, a quintupling of transmission capacity is needed to meet a high load growth future by 2035.[31] But supply chain delays and multi-year planning processes continue to slow the deployment of new infrastructure, with transformer lead times increasing from 10 to 16 weeks pre-pandemic to 48 to 62

---

[30] Solar with storage is assumed to have an ELCC that declines linearly from 0.5 in 2024 to 0.35 in 2030.

[31] "National Transmission Needs Study." U.S. Department of Energy. October 2023. https://www.energy.gov/sites/default/files/2023-12/National%20Transmission%20Needs%20Study%20-%20Final_2023.12.1.pdf

weeks or more.[32] Generators waiting to come online to meet this demand are facing increasingly long interconnection timelines, growing from an average of 2.1 years in 2000-2010 to 3.7 years in 2011-2021; ultimately, 72% of projects withdraw.[33]

Critical to these supply challenges is the significant difference in development timelines between data centers and electric infrastructure. Data centers can be developed and connected in one to two years while new generation and transmission can take four years or more.[34] Given these long timelines to develop more supply, grid operators and utilities have postponed the retirement of fossil fuel power plants for reliability.[35] [36] These rollbacks clash with the customers' environmental goals and investments, and they threaten state and utility emission reduction targets. For example, delaying a coal plant retirement by just one year would have significant carbon emissions ramifications. A one-gigawatt coal plant with the average 42.1% capacity factor[37] and 2,300 lbs/MWh CO2 emission rate[38] would emit 3.8 million metric tons of CO2 in that year – equal to deploying nearly three gigawatts of utility-scale solar from an average avoided emissions perspective.[39] This example illustrates how any delay in coal retirements must be avoided from a carbon reduction perspective as it dwarfs most solar or wind additions.

Moreover, these operating challenges and large required investments pose major customer affordability, safety, and reliability concerns. While the rapid load growth from data centers is one cause of these challenges, strategically leveraging the resources of data center investors can also help derisk and finance many long-needed grid upgrades if there is an ability to appropriately balance costs, risk, and timing by all the stakeholders which is non-trivial.

---

[32] Susan Partain. "Guessing Game; how Uncertainty in the Supply Chain is Affecting Utilities." American Public Power Association. 15 February 2023. https://www.publicpower.org/periodical/article/guessing-game-how-uncertainty-supply-chain-affecting-utilities

[33] Joseph Rand, Ryan Wiser, Will Gorman, Dev Millstein, Joachim Seel, Seongeun Jeong, Dana Robson. "Queued Up; Characteristics of Power Plants Seeking Transmission Interconnection." Lawrence Berkeley National Laboratory. April 2022. https://eta-publications.lbl.gov/sites/default/files/queued_up_2021_04-13-2022.pdf

[34] "Powering Intelligence: Analyzing Artificial Intelligence and Data Center Energy Consumption." EPRI. 2024. https://www.wpr.org/wp-content/uploads/2024/06/3002028905_Powering-Intelligence_-Analyzing-Artificial-Intelligence-and-Data-Center-Energy-Consumption.pdf

[35] Brand Plumer and Nadja Popovich. "A New Surge in Power Use Is Threatening U.S. Climate Goals." The New York Times. 14 March 2024. https://www.nytimes.com/interactive/2024/03/13/climate/electric-power-climate-change.html

[36] Nicole Jao. "US grid operator PJM asks Talen Energy to postpone fossil fuel plant retirements." Reuters. 11 January 2024. https://www.reuters.com/business/energy/us-grid-operator-pjm-asks-talen-energy-postpone-fossil-fuel-plant-retirements-2024-01-10/

[37] "Monthly Energy Review." U.S. Energy Information Administration. May 2024. https://www.eia.gov/totalenergy/data/monthly/

[38] "How much carbon dioxide is produced per kilowatthour of U.S. electricity generation?". U.S. Energy Information Administration. 07 December 2023. https://www.eia.gov/tools/faqs/faq.php?id=74&t=11

[39] "AVERT v4.3 Avoided Emission Rates 2017-2023 (April 2024) xlsx." U.S. Environmental Protection Agency. https://www.epa.gov/avert/avoided-emission-rates-generated-avert

These unprecedented circumstances require a new paradigm—one that prioritizes proactive planning, collaborative partnerships, and innovative solutions. By embracing this approach, both power system planners and customers can navigate the challenges posed by large load growth while advancing towards a more sustainable, strategic, and reliable energy future. Listed below are avenues stakeholders can pursue in the face of these challenges and become critical parts of the solution.

## What can large load customers do, besides taking a scattershot approach to development leading to potential abandonment of sites as they wait for interconnection?

Large load customers prioritize fast interconnections and relatively low-cost reliable electricity. If waiting in interconnection queues is unacceptable, these customers have several options to minimize their missed opportunities, risks, and revenue losses, and even become a grid asset.

- Historically, data centers have been passive loads, drawing continuous power from the grid. They generally have backup diesel or natural gas generators to maintain uptime during grid outages but rarely use them as they are located in highly reliable areas. Other technologies can provide clean back up and serve as grid assets, such as on-side battery storage, and can help manage peak demand more sustainably and affordably while helping to integrate renewables. Viewing back up power as a strategic asset deployment could transform the industry from being a "sink" (inflexible flat load, demanding specific power level at specific time frame) to a partner or potential "source" in a sustainable future. In exchange for the project providing valuable grid services, utilities can more effectively and quickly enable access to power.
- Data centers can commit to flexible load plans to accommodate grid limitations and avoid lengthy upgrade timelines. For example, facilities can leverage the temporal and spatial flexibility of certain AI workloads (e.g., model training) and schedule batch data processing to optimize power usage around renewable energy availability and total system load.
- Collaborating directly with utilities and/or market operators (depending on the region) can drive innovative solutions. For example, coordinating during peak periods can position data centers as large-scale, flexible grid assets. While this might under-utilize rapidly depreciating high-cost servers, it may be a worthwhile trade-off if it enables a faster interconnection, which may have significant strategic business value to data centers, in addition to providing compensation for their flexibility and a reduced carbon footprint.[40] [41]

---

[40] "Powering Intelligence: Analyzing Artificial Intelligence and Data Center Energy Consumption." EPRI. 2024. https://www.wpr.org/wp-content/uploads/2024/06/3002028905_Powering-Intelligence_-Analyzing-Artificial-Intelligence-and-Data-Center-Energy-Consumption.pdf

[41] Data centers should also continue to work on in reducing overall energy footprint by investing in energy efficient computational hardware (which represents 40-50% of data center energy consumption) and cooling technology (30-40%). For example, a recent study examining a shift from 100% air cooling to 25% air and 75% liquid cooling observed a 15.5% decrease in the data center's energy usage.

- Data centers can also leverage their relative flexibility in siting their loads, which is a new advantage as fossil fuel generation has historically been more geographically flexible than load.[42] While still considering proximity to fiber networks and low natural disaster risk, data centers are generally able to seek out areas with low-cost land, renewables, access to water, and sufficient grid capacity to meet their needs. Data center planners should expand their search areas beyond the traditional primary markets to take advantage of these new opportunities and could acquire existing infrastructure or support new transmission lines to minimize interconnection time.

Table 1 outlines options for large load customers along with some examples in action.

*Table 1: Summary of Options for Large Load Customers*

| Option | Examples in action |
|---|---|
| **Self-generate** to bridge the gap until full service is available or use for flexibility (e.g., interruption or demand response) | Enchanted Rock has its "Bridge-to-Grid" offering, building microgrids for facilities awaiting firm grid connection. Once interconnected, the resource can provide flexible capacity back to the grid and serve as backup power.[43] |
| **Leverage flexible load** via batch processing, task shifting, demand response, interruptible service tariffs | Google can shift compute tasks based on clean energy availability[44] and participates in demand response.[45] |
| **Work with energy suppliers directly** to bypass/minimize interconnection timelines | Microsoft deal with Brookfield for 10.5 GW of renewables.[46] |
| **Work with specific utilities on innovative solutions** | Amazon, Google, Microsoft and Nucor signed MOUs with Duke Energy to develop "Accelerating Clean Energy" tariffs that would lower the costs of investing in clean energy technologies through early commitments, facilitate beneficial on-site generation and participation in load flexibility programs.[47] Arizona Public Service has an Extra High Load Factor rate for customers that can demonstrate 50% of its annual energy consumption within APS is carbon free.[48] |

---

[42] It is easier for a planner to decide the location of a natural gas power plant than move an entire city as an extreme example. However, current renewable generation is more geographically constrained than data centers.

[43] "Enchanted Rock Bridge-to-Grid Solution Addresses Power Demand Growth from AI and Electrification." Enchanted Rock. 16 May 2024. https://enchantedrock.com/enchanted-rock-bridge-to-grid-solution-addresses-power-demand-growth-from-ai-and-electrification/

[44] Ross Koningstein. "We now do more computing where there's cleaner energy." Google: The Keyword. 18 May 2021. https://blog.google/outreach-initiatives/sustainability/carbon-aware-computing-location/

[45] Varun Mehra and Raiden Hasegawa. "Supporting power grids with demand response at Google data centers." Google: Cloud. 03 October 2023. https://cloud.google.com/blog/products/infrastructure/using-demand-response-to-reduce-data-center-power-consumption

[46] "Brookfield and Microsoft Collaborating to Deliver Over 10.5 GW of New Renewable Power Capacity Globally." Brookfield. 01 May 2024. https://bep.brookfield.com/press-releases/bep/brookfield-and-microsoft-collaborating-deliver-over-105-gw-new-renewable-power

[47] "Responding to growing demand, Duke Energy, Amazon, Google, Microsoft and Nucore execute agreements to accelerate clean energy options." Duke Energy: News Center. 29 May 2024. https://news.duke-energy.com/releases/responding-to-growing-demand-duke-energy-amazon-google-microsoft-and-nucor-execute-agreements-to-accelerate-clean-energy-options

[48] Arizona Public Service Company. Extra High Load Factor Rate Schedule. Accessed July 9, 2024, from https://www.aps.com/-/media/APS/APSCOM-PDFs/Utility/Regulatory-and-Legal/Regulatory-Plan-Details-Tariffs/Business/Business-NonResidential-Plans/ExtraHighLoadFactor.ashx?la=en

| | |
|---|---|
| **Acquire infrastructure** that already has interconnection (e.g., former industrial site, underperforming load) | Amazon buys nuclear-powered data center from Talen.[49] Skybox Datacenters converted a vacant Prologis-owned distribution center in IL into a 30 MW data center. QTS Realty Trust purchased 400 acres previously planned as a $1.5B logistic park in AZ.[50] |
| **Support new transmission lines** looking to interconnect low-cost renewables | None public yet; emerging as existing transmission/generation headroom becomes constrained |

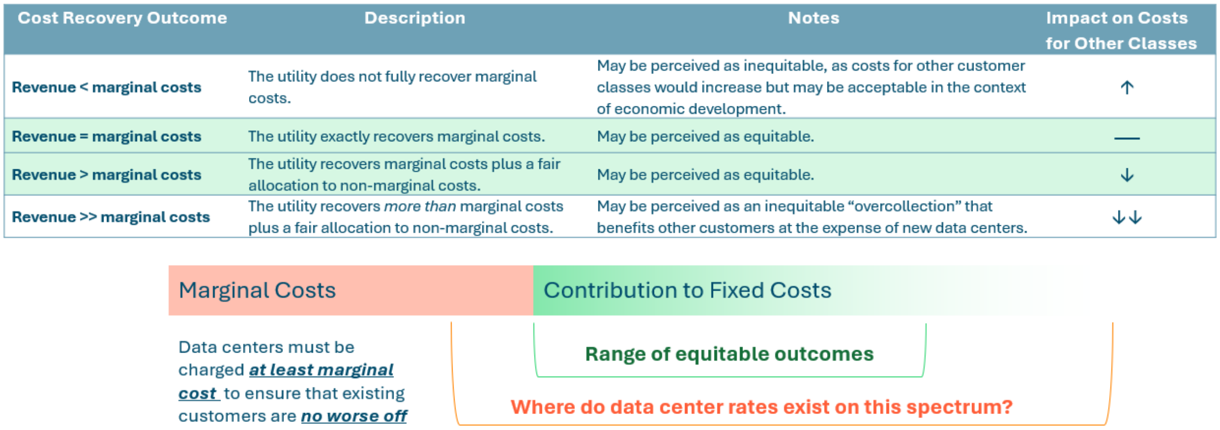
## What can utilities, system planners, and regulators do?

Given the scale of these projects and the relative novelty of these innovative opportunities, regulators need to develop the proper structure to appropriately value and compensate for these resources. More broadly, as these stakeholders think about proactive grid planning decisions, it is important to distinguish different categories of investments.

- Growth-related investments that work to support near-term large loads, medium term electric vehicle growth, and longer-term building and industrial electrification
- Investments with potential stranding and/or underutilization that may not benefit additional current and future customers, but ways to minimize risk
- Investments that spread too much risk and/or cost to utility ratepayers that may not directly benefit from the investments

The options below can help maximize the value of the growth-related investments, suggest structures to minimize risk in utility investments, and manage potential ratepayer impacts. These options can have neutral to positive impacts on costs for other customers across utility rate classes as outlined in Figure 13. Accordingly, determining the appropriate level of investment to collect from large loads versus other customers will be key to ensuring an long-term equitable outcome.

***Figure 13: Potential Spectrum of Utility Rates for New Data Center Customers***



| Cost Recovery Outcome | Description | Notes | Impact on Costs for Other Classes |
|---|---|---|---|
| **Revenue < marginal costs** | The utility does not fully recover marginal costs. | May be perceived as inequitable, as costs for other customer classes would increase but may be acceptable in the context of economic development. | ↑ |
| **Revenue = marginal costs** | The utility exactly recovers marginal costs. | May be perceived as equitable. | — |
| **Revenue > marginal costs** | The utility recovers marginal costs plus a fair allocation to non-marginal costs. | May be perceived as equitable. | ↓ |
| **Revenue >> marginal costs** | The utility recovers *more than* marginal costs plus a fair allocation to non-marginal costs. | May be perceived as an inequitable "overcollection" that benefits other customers at the expense of new data centers. | ↓↓ |

---

[49] "Amazon buys nuclear-powered data center from Talen." Nuclear Newswire. 07 March 2024. https://www.ans.org/news/article-5842/amazon-buys-nuclearpowered-data-center-from-talen/

[50] Dan Rabb. "Industrial Sites Start Flipping to Data Centers Amid Fears Of Logistics Slowdown." Skybox Data Centers. 22 July 2022. https://www.skyboxdatacenters.com/news/industrial-sites-start-flipping-to-data-centers-amid-fears-of-logistics-slowdown

Table 2 below provides further detail on potential options and their cost and risk impacts.

## *Table 2: Summary of Options for Utilities, System Planners and Regulators*

| Goal | Option | Examples in action | Cost / Risk Impacts |
|---|---|---|---|
| **Increase Proactive Planning** | Identify and share data on where there is existing headroom and bring large loads into the long-term planning process. | California utilities: supported the combined IRP and transmission planning process, in which local network constraints are directly incorporated into the capacity expansion modeling framework | Identifying areas where investments to serve loads in the short-term would also support anticipated electrification in the long-term would yield savings. |
| **Reform Interconnection Process from “First Come, First Serve” to Reduce Timeliness** | Create a “fast track” by requiring upfront payment / long-term commitments to reduce timelines for credible (vs. speculative) loads. | Emerging | Collecting upfront payment would help the utility de-risk and improve cost-sharing. |
| | Allocate capacity via a competitive economic process e.g., auction or a “first ready, first serve” process. | | Using an economic mechanism should increase efficiency and revenue. |
| **Facilitate Large Loads Developing Own Resources and Leveraging Flexibility** | Design tariffs that allow for flexibility and innovation on the load side such as on-site generation or being compensated for taking interruptible/non-firm service as well as promoting emerging generation technologies. | Duke Energy’s proposed Accelerating Clean Energy tariffs would facilitate large customers’ on-site generation, participation in load flexibility programs, and investments in clean energy.[51]<br>NV Energy’s Clean Energy Transition Tariff supporting large customers like Google’s desire to adopt new clean energy technologies like enhanced geothermal.[52] | Flexible load management can provide cost savings to the utility and customer.<br>De-risking as an early adopter emerging technologies for the clean energy transition like long duration energy storage, advanced nuclear, enhanced geothermal, carbon capture, and other technologies. |
| | Utilities could off-take or manage the unused energy or purchase in the future. | Omaha Public Power District to access 600 MW of wind capacity from NextEra’s facility, which is part of Google’s clean energy portfolio.[53] | Innovative procurement could help utilities meet regulatory requirements and ensure reliability more cost-effectively. |
| **Implement Cost-Sharing Mechanisms to Avoid Inequitable Cost-Shift** | Require large load customers to provide upfront investment or other risk mitigants such as long-term commitments. | AEP Ohio’s proposed data center rate category would require 10-year contracts and a minimum demand charge payment based on 90% of contract capacity, up from 60%.[54] Duke Energy proposed a rate structure that would have a “minimum take” clause.[55] | Requiring upfront and/or additional investment could help fairly allocate risk and costs. |
| | Design large load tariffs based on incremental cost of service. | Emerging and there may already be examples from previous eras of growth. | Designing rates that more closely reflect cost structures would help minimize cost-shift. |

---

[51] “Responding to growing demand, Duke Energy, Amazon, Google, Microsoft and Nucore execute agreements to accelerate clean energy options.” Duke Energy: News Center. 29 May 2024. https://news.duke-energy.com/releases/responding-to-growing-demand-duke-energy-amazon-google-microsoft-and-nucor-execute-agreements-to-accelerate-clean-energy-options

[52] Amanda Peterson Corio and Briana Kobor. “How we’re working with utilities to create a new model for clean energy.” Google: The Keyword. 11 June 2024. https://blog.google/outreach-initiatives/sustainability/google-clean-energy-partnership/

[53] “OPPD Welcomes Clean Capacity Collaborations as part of its Reliable Growth Plans.” Omaha Public Power District. 14 May 2024. https://www.oppd.com/news-resources/news-releases/2024/may/oppd-welcomes-clean-capacity-collaboration-as-part-of-its-reliable-growth-plans/

[54] Ethan Howland. “AEP Ohio proposes data center, crypto financial requirements amid 30 GW in service inquiries.” UtilityDive.com. 15 May 2024. https://www.utilitydive.com/news/aep-ohio-data-center-crypto-rates-puc/716150/

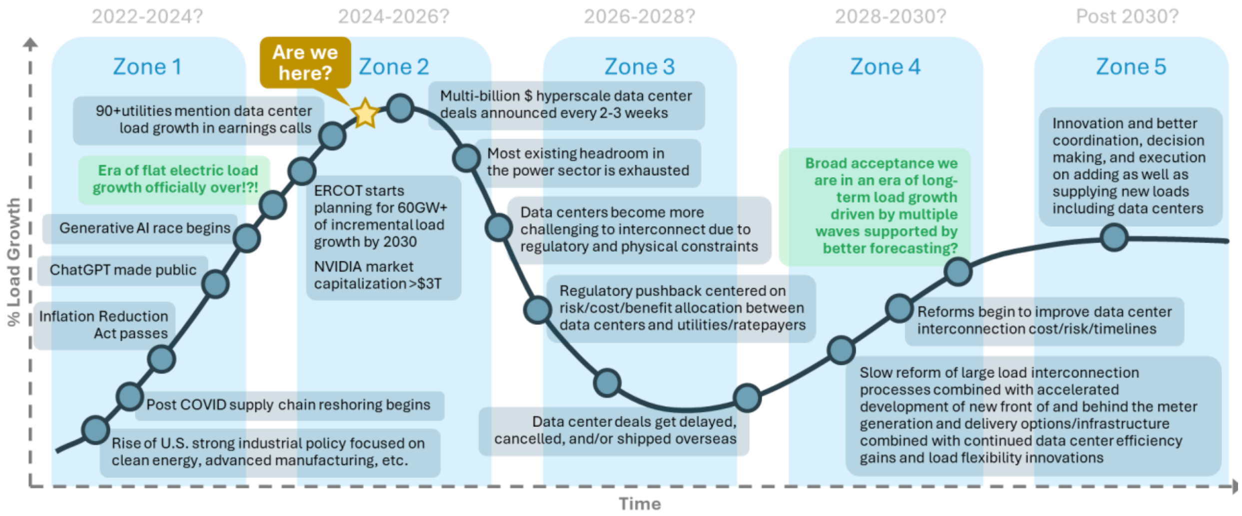
[55] Laila Kearney. “Duke Energy seeks take-or-pay power contracts for data centers.” Reuters. 07 May 2024. https://www.reuters.com/business/energy/duke-energy-seeks-take-or-pay-power-contracts-data-centers-2024-05-07

## E3's View

E3 works on hundreds of projects a year exclusively in the electric and gas sectors for a diverse range of clients, including state and federal agencies, utilities and market operators, regulators, and private industry, developers and investors. We believe this gives us a nuanced perspective on the challenges, issues, and potential solutions around near to longer-term load growth. The many questions we have been getting from our clients on the topic of data center load growth along with our current work on the topic motivated the creation of this whitepaper.

There is still much uncertainty at the macro and micro level regarding the scope and scale of data center growth. However, the key question should not be "How much will load grow?", but instead "Where and what kind of load growth can be accommodated in different jurisdictions?". As we crest the initial wave of a potential "hype" cycle shown in Figure 14 and growth expectations may reach their peak, stakeholders must begin to take hard looks at their abilities to meet different load scenarios and their options for doing so. Growth will occur, and its potential may ultimately be shaped by the energy sector's ability to accommodate that growth.

***Figure 14: Are we in a Power Sector Data Center Hype Cycle? Illustrative Visualization based on Gartner Hype Cycle*** [56]



| ZONE 1 | ZONE 2 | ZONE 3 | ZONE 4 | ZONE 5 |
|---|---|---|---|---|
| Innovation trigger | Peak of Inflated Expectations | Trough of Disillusionment | Slope of Enlightenment | Plateau of Productivity |
| 12-18 Months? | 6-12Months? | 12-18 Months? | 12-36 months? | 12-36 months? |

**Total Cycle Time → ~5-10 years?**

---

[56] "Gartner Hype Cycle" Wikipedia.com. Accessed 21 June 2024. https://en.wikipedia.org/wiki/Gartner_hype_cycle

Amidst this uncertainty, a coordinated approach with a clear understanding of cost and risk sharing and mitigation is essential. The integrated systems planning[57] approach with a different market construct is well-suited to matching supply and demand in systems with hard constraints on both. This would entail transitioning away from traditional planning approaches, which focus on incremental growth and a serial one-off or limited duration perspective, into a longer term and collaborative planning and execution model.

This type of approach would help realize optimal existing headroom allocation, increase energy affordability, and aid decarbonization efforts, all while enabling long-term strategic planning. As we enter a “new build” era with multiple waves of load growth, planners may need to adapt to operating in a constrained environment for the foreseeable future where scalable and innovative solutions on both the demand and supply side will be required.

Data center load growth could be a positive for the industry if leveraged effectively. Well-resourced customers with high load factors can help fund much-needed transmission grid upgrades and drive new clean energy supply. This could lead to cost savings for other customers if higher incremental sales are used to pay down the fixed costs (both existing and incremental) of the system. Furthermore, these customers can mitigate the risks and costs of bringing emerging technology generators (e.g. small modular reactors, advanced geothermal) to market and accelerate their adoption by becoming early customers for their power as well as anchoring other needed grid investments.

The surge in new electricity demand poses challenges, but also offers a unique opportunity to accelerate the transition to a cleaner, more reliable, and affordable energy future. The traditional paradigm of short-term incremental planning and reactive infrastructure development is no longer sufficient, especially in the face of continual waves of new load growth. A new approach is needed, one that embraces proactive planning, collaborative partnerships, and innovative solutions. Market participants and power planners should adapt and work together to develop a comprehensive and sustainable approach to meeting the nation's growing energy needs. The growth of data centers is not just a challenge to be overcome; it is an opportunity to build a better electric system for all.

---

[57] Integrated system planning (ISP) utilizes a cohesive integrated set of data, processes, and models to integrate generation and customer resource planning with transmission and distribution grid planning. This integration is critical to making decisions that balance making the right investments, in the right places, at the right times. As technology and participatory models advance, customers will increasingly be involved participants in this transition, requiring utilities and regulators to know when to support customer investments or behavioral changes over their own historical “wires” investments, while ensuring sufficient reliability and operational control over the broad and powerful set of emerging customer resources.

## Appendix 1: Demand and Supply Forecasting Methodology and Sensitivities

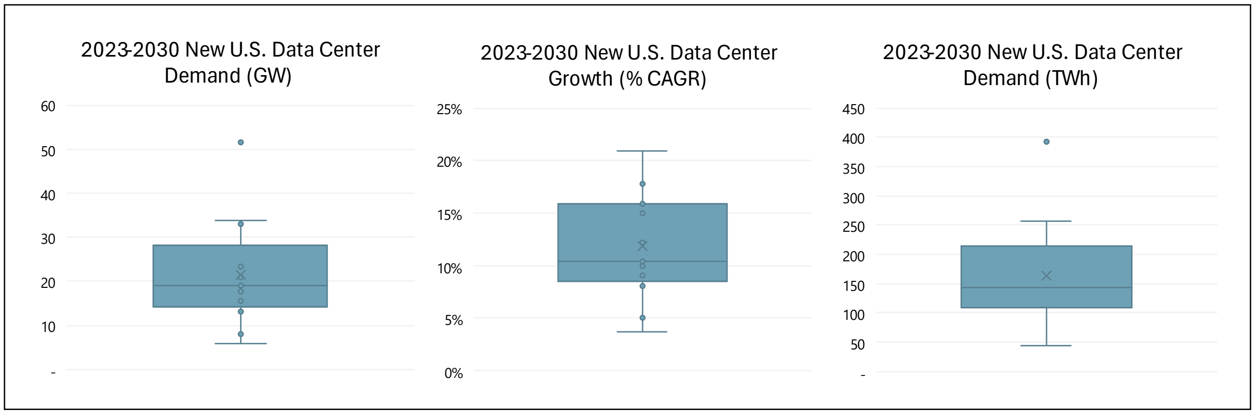
E3 used published data center projections to analyze and illustrate the potential new resource builds required to meet incremental annual energy demands driven by new data center development. E3 estimated new resource builds under a range of sensitivities examining lower energy demands from assumed incremental energy efficiency gains to computing and cooling data center operations.

***Key assumptions for the analysis include:***

- Load profile of a data center has an 86% load factor (average MWh/peak)
- 1.2 Power Usage Effectiveness (PUE) of new data centers is used as the starting point for further energy efficiency improvements
- 75% or 100% of new data center energy demand is met by renewables (wind and solar) in any year with gas meeting the remainder
- Any new renewables generation is comprised of 70% solar and 30% wind, with respective capacity factors of 22% and 36%; other capacity generation assumes a 54% capacity factor
- For the effective capacity analysis, E3 makes simplified assumptions of Effective Load Carrying Capability (ELCC) values (*Note: the use of ELCC values in this analysis is for illustrative purposes only and does not provide a comprehensive assessment for true reliability planning*)
    - Solar, Wind, and Firm Capacity resources assume respective ELCC values of 0.2, 0.21, and 0.95 for all years
    - Solar + Storage resources assume an ELCC value of 0.5 in 2024, declining linearly to 0.35 in 2030;
    - assumes half of solar nameplate capacity is paired with short-duration battery storage

### *E3's Data Center Demand Projections Using Various Public Sources*

The estimates of new data center demand, compound annual growth rate (CAGR), and energy demand are based on a review of data center projections from JLL[58], McKinsey[59], EPRI[60], IEA[61], BCG[62], Mordor[63] and Goldman Sachs[64] illustrated below. The box-and-whisker plots compile 13 projections, including four sensitivities from EPRI (Low, Moderate, High, and Higher scenarios) and three from Goldman Sachs (Bear, Base, Bull scenarios). Sources provide projections in terms of forecasted data center capacity (e.g. MW) or energy demand (e.g. MWh). Where applicable, E3 estimates data center capacity from energy estimates or vice-versa using an assumed 86% load factor.



---

[58] Kari Beets. "North America Data Center Report." JLL. 28 February 2024. https://www.us.jll.com/en/trends-and-insights/research/na-data-center-report

[59] Srini Bangalore, Arjita Bhan, Andrea Del Miglio, Pankaj Sachdeva, Vijay Sarma, Raman Sharma, and Bhargs Srivathsan. "Investing in the rising data center economy." McKinsey & Company. 17 January 2023. https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/investing-in-the-rising-data-center-economy

[60] "Powering Intelligence: Analyzing Artificial Intelligence and Data Center Energy Consumption." EPRI. 2024. https://www.wpr.org/wp-content/uploads/2024/06/3002028905_Powering-Intelligence_-Analyzing-Artificial-Intelligence-and-Data-Center-Energy-Consumption.pdf

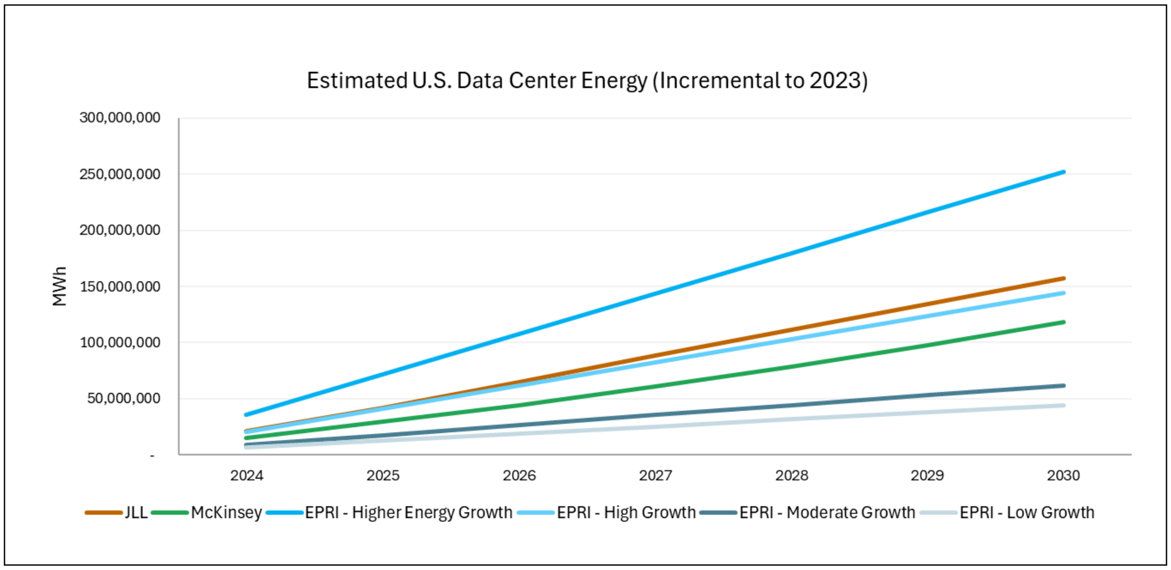
[61] "Electricity 2024: Analysis and forecast to 2026." IEA. 19 January 2024. https://iea.blob.core.windows.net/assets/6b2fd954-2017-408e-bf08-952fdd62118a/Electricity2024-Analysisandforecastto2026.pdf

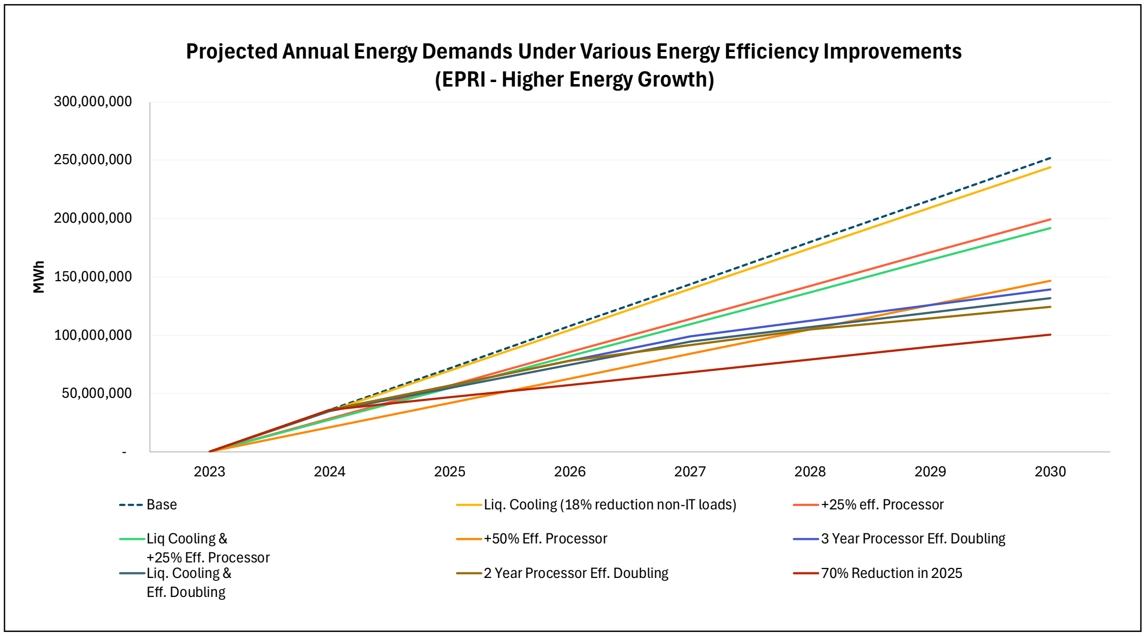
[62] "The Impact of GenAI in Electricity." Boston Consulting Group. 2024. https://www.linkedin.com/posts/bcg-on-energy_the-impact-of-genai-in-electricity-activity-7112787574032674816-uDEX/

BCG's "US Data Center Power Outlook" report issued in July 2024 provides its more updated view, projecting new data center demand growth ranging from 60 to 90 GW in 2023-2030.

[63] "United States Data Center Market Size & Share Analysis – Growth Trends & Forecasts Up to 2029." Mordor Intelligence. https://www.mordorintelligence.com/industry-reports/united-states-data-center-market

[64] "Generational Growth: AI, data centers and the coming US power demand surge." The Goldman Sachs Group, Inc. 28 April 2024. https://www.goldmansachs.com/intelligence/pages/gs-research/generational-growth-ai-data-centers-and-the-coming-us-power-surge/report.pdf.





### *Energy Efficiency Demand Sensitivities*

Assuming a new interconnected data center has a power usage effectiveness (PUE) of 1.2, this means the facility's total energy use is divided into 83% for IT demand and 17% for non-IT demand (such as cooling). E3 applied various improvements to IT power consumption, non-IT power consumption, or both.

Processor efficiency is assumed to impact all IT energy loads. Therefore, a 25% gain in processor efficiency reduces the IT load share of the facility-wide demand from approximately 83% to 63%. When combined with non-IT loads, which maintain their 17% share of the base case facility-wide energy demands, the new facility-wide demand is estimated to be around 80% of the base case (63% + 17%).

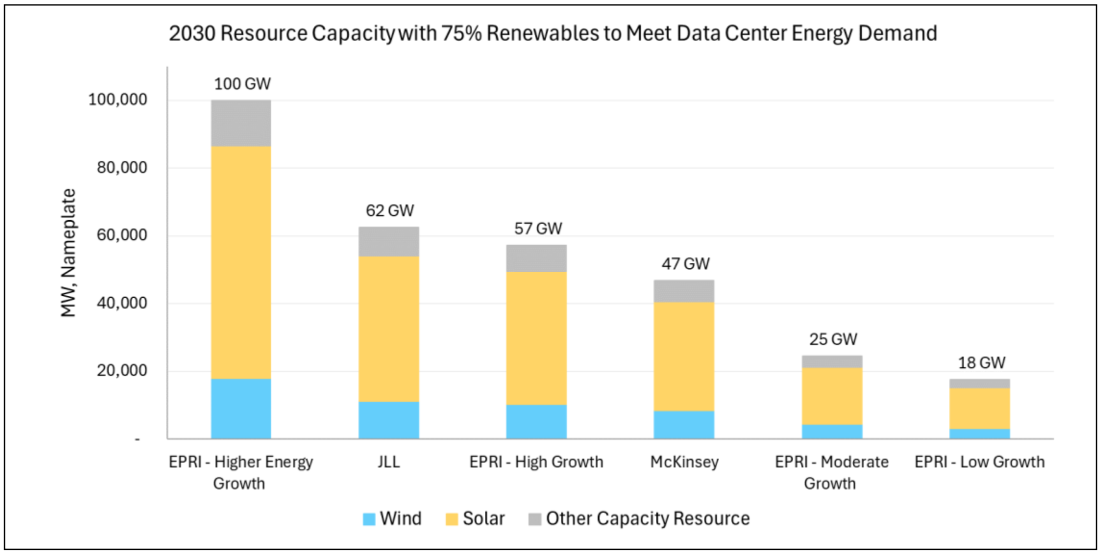
Interactions between IT and non-IT energy demands resulting from improvements to processor efficiency are not explicitly accounted for. This analysis also assumes that the end-use computing demands remain constant despite improved efficiency. This means the base case energy use serves a fixed amount of computing tasks and the possibility that enhanced IT efficiency could lead to induced computing and power demand, as facility operators seek to maintain high utilization, is not considered.

Under these hypothetical efficiency improvements and assumed schedules, facility-wide power demand of newly interconnected data centers are reduced relative to the base case (100%) as shown in the following table. Not all energy efficiency scenarios are depicted in the resource builds.

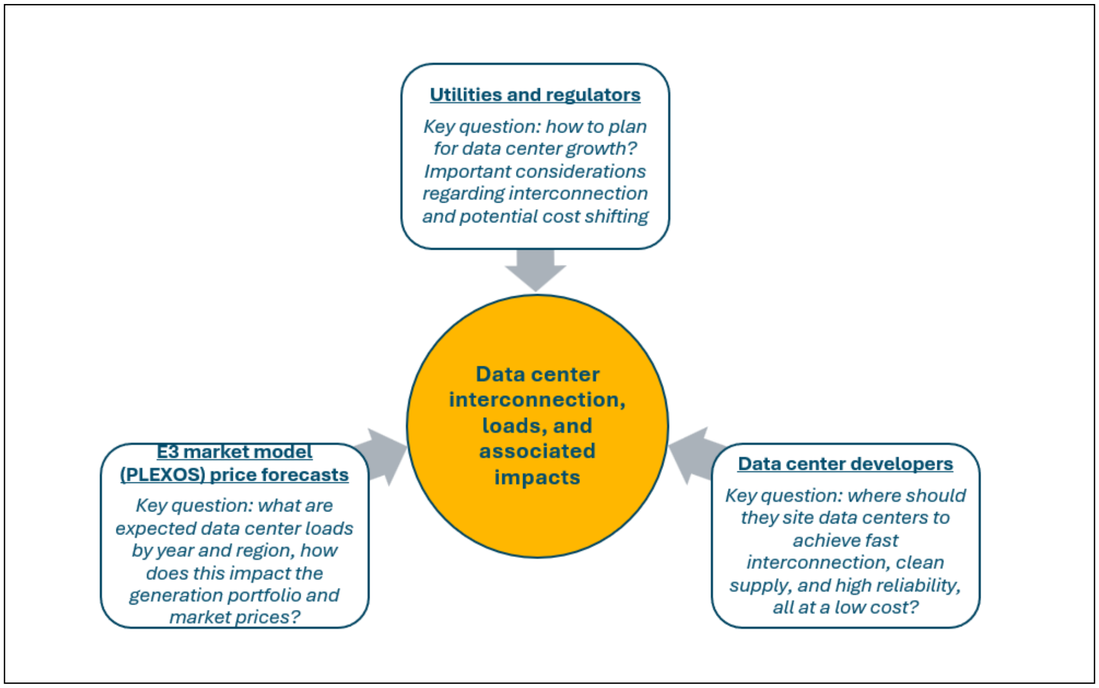
### Energy Efficiency Demand Sensitivities: New Data Center Facility-Wide Energy Consumption Relative to Base Case

| | 2024 | 2025 | 2026 | 2027 | 2028 | 2029 | 2030 |
|---|---|---|---|---|---|---|---|
| **Base** | 100% | 100% | 100% | 100% | 100% | 100% | 100% |
| **Liq. Cooling (18% reduction non-IT loads)** | 97% | 97% | 97% | 97% | 97% | 97% | 97% |
| **+25% eff. Processor** | 79% | 79% | 79% | 79% | 79% | 79% | 79% |
| **Liq Cooling & +25% Eff. Processor** | 76% | 76% | 76% | 76% | 76% | 76% | 76% |
| **+50% Eff. Processor** | 58% | 58% | 58% | 58% | 58% | 58% | 58% |
| **3 Year Processor Eff. Doubling** | 100% | 58% | 58% | 58% | 38% | 38% | 38% |
| **Liq. Cooling & Eff. Doubling** | 97% | 55% | 55% | 55% | 35% | 35% | 35% |
| **2 Year Processor Eff. Doubling** | 100% | 58% | 58% | 38% | 38% | 27% | 27% |
| **70% Reduction in 2025** | 100% | 30% | 30% | 30% | 30% | 30% | 30% |

The figure below depicts the estimated nameplate capacity of resources necessary to provide sufficient annual energy demand across various data center growth projections. Figure 9 in the report shows the same builds but also indicates the range of build uncertainty accounting for the following hypothetical energy efficiency improvement scenarios: base, liquid cooling, +25% efficient processor, liquid cooling & +25% efficient processor, +50% efficient processor, and 2-year processor efficiency doubling.



## Appendix 2: E3's Load Forecasting Approach Currently Incorporated in Our U.S. Wide PLEXOS Market Model



**Utilities and regulators**

*Key question: how to plan for data center growth? Important considerations regarding interconnection and potential cost shifting*

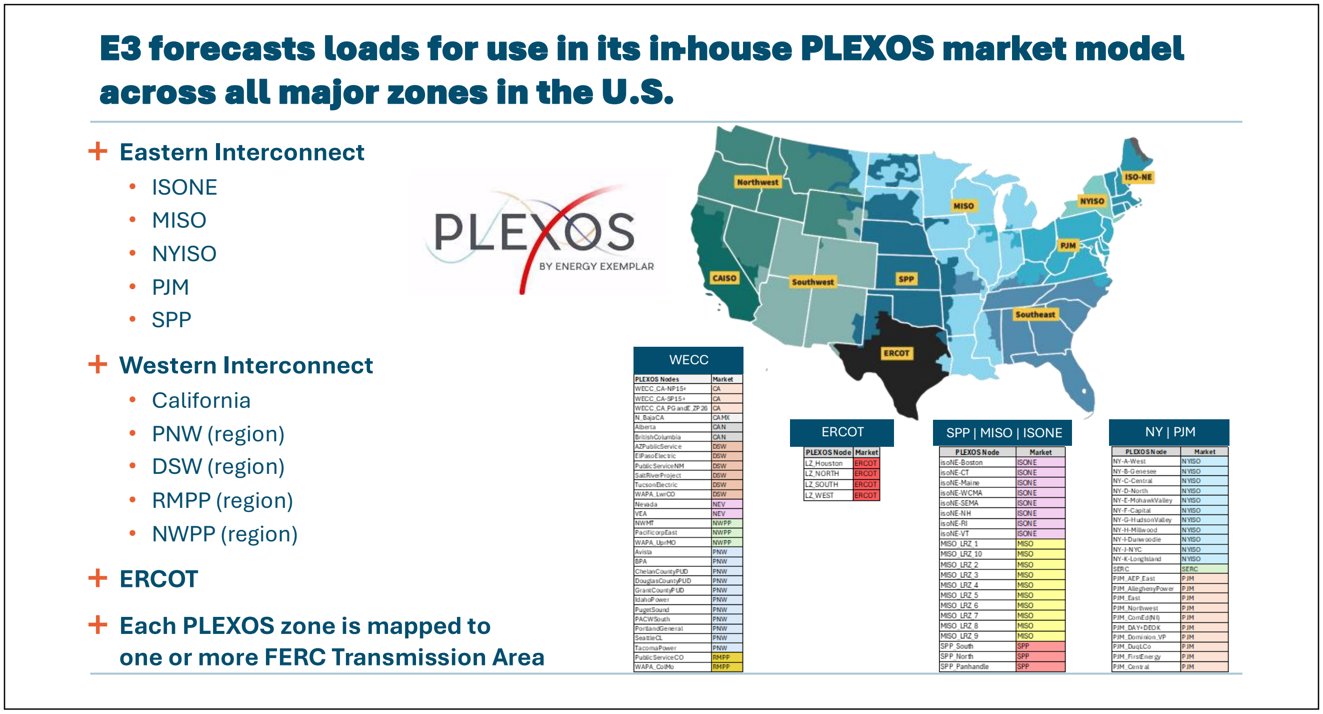
**Data center interconnection, loads, and associated impacts**

**E3 market model (PLEXOS) price forecasts**

*Key question: what are expected data center loads by year and region, how does this impact the generation portfolio and market prices?*

**Data center developers**

*Key question: where should they site data centers to achieve fast interconnection, clean supply, and high reliability, all at a low cost?*



### E3 forecasts loads for use in its in-house PLEXOS market model across all major zones in the U.S.

- **Eastern Interconnect**
  - ISONE
  - MISO
  - NYISO
  - PJM
  - SPP
- **Western Interconnect**
  - California
  - PNW (region)
  - DSW (region)
  - RMPP (region)
  - NWPP (region)
- **ERCOT**
- **Each PLEXOS zone is mapped to one or more FERC Transmission Area**

PLEXOS BY ENERGY EXEMPLAR

**WECC**

| PLEXOS Nodes | Market |
|---|---|
| WECC_CA-NP15+ | CA |
| WECC_CA-SP15+ | CA |
| WECC_CA_PGandE_ZP26 | CA |
| N_BajaCA | CAMX |
| Alberta | CAN |
| BritishColumbia | CAN |
| AZPublicService | DSW |
| ElPasoElectric | DSW |
| PublicServiceNM | DSW |
| SaltRiverProject | DSW |
| TucsonElectric | DSW |
| WAPA_LwrCO | DSW |
| Nevada | NEV |
| VEA | NEV |
| NWMT | NWPP |
| PacificorpEast | NWPP |
| WAPA_UprMO | NWPP |
| Avista | PNW |
| BPA | PNW |
| ChelanCountyPUD | PNW |
| DouglasCountyPUD | PNW |
| GrantCountyPUD | PNW |
| IdahoPower | PNW |
| PugetSound | PNW |
| PACWSouth | PNW |
| PortlandGeneral | PNW |
| SeattleCL | PNW |
| TacomaPower | PNW |
| PublicServiceCO | RMPP |
| WAPA_ColMo | RMPP |

**ERCOT**

| PLEXOS Node | Market |
|---|---|
| LZ_Houston | ERCOT |
| LZ_NORTH | ERCOT |
| LZ_SOUTH | ERCOT |
| LZ_WEST | ERCOT |

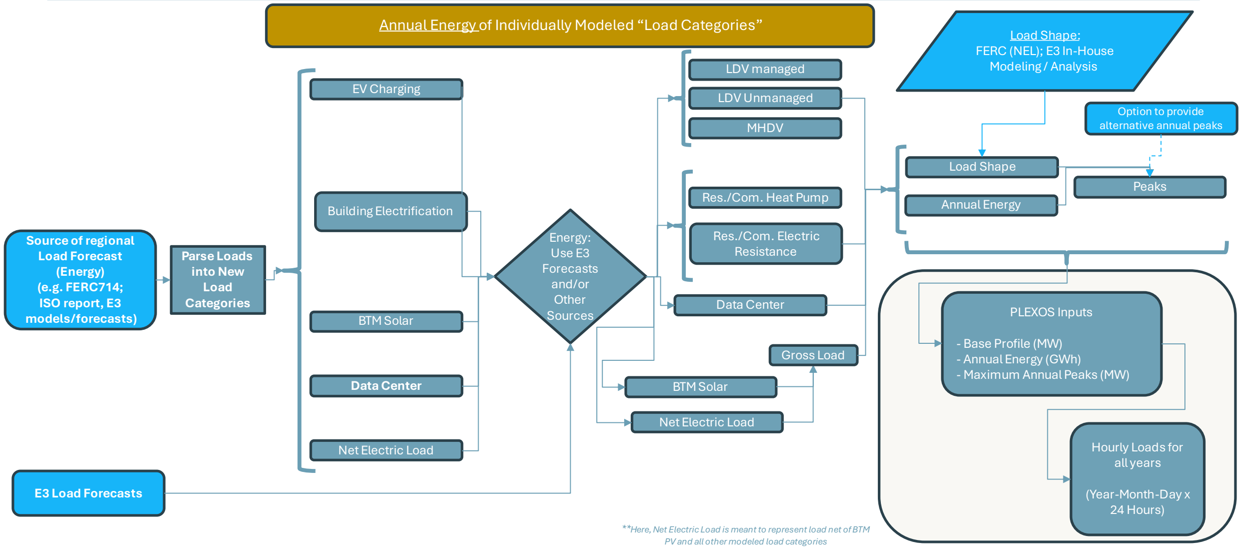
**SPP | MISO | ISONE**

| PLEXOS Node | Market |
|---|---|
| isoNE-Boston | ISONE |
| isoNE-CT | ISONE |
| isoNE-Maine | ISONE |
| isoNE-WCMA | ISONE |
| isoNE-SEMA | ISONE |
| isoNE-NH | ISONE |
| isoNE-RI | ISONE |
| isoNE-VT | ISONE |
| MISO_LRZ_1 | MISO |
| MISO_LRZ_10 | MISO |
| MISO_LRZ_2 | MISO |
| MISO_LRZ_3 | MISO |
| MISO_LRZ_4 | MISO |
| MISO_LRZ_5 | MISO |
| MISO_LRZ_6 | MISO |
| MISO_LRZ_7 | MISO |
| MISO_LRZ_8 | MISO |
| MISO_LRZ_9 | MISO |
| SPP_South | SPP |
| SPP_North | SPP |
| SPP_Panhandle | SPP |

**NY | PJM**

| PLEXOS Node | Market |
|---|---|
| NY-A-West | NYISO |
| NY-B-Genesee | NYISO |
| NY-C-Central | NYISO |
| NY-D-North | NYISO |
| NY-E-MohawkValley | NYISO |
| NY-F-Capital | NYISO |
| NY-G-HudsonValley | NYISO |
| NY-H-Millwood | NYISO |
| NY-I-Dunwoodie | NYISO |
| NY-J-NYC | NYISO |
| NY-K-LongIsland | NYISO |
| SERC | SERC |
| PJM_AEP_East | PJM |
| PJM_AlleghenyPower | PJM |
| PJM_East | PJM |
| PJM_Northwest | PJM |
| PJM_ComEd(NI) | PJM |
| PJM_DAY+DEOK | PJM |
| PJM_Dominion_VP | PJM |
| PJM_DuqLCo | PJM |
| PJM_FirstEnergy | PJM |
| PJM_Central | PJM |

## General E3 Process Flow for Modeling Major New Future Loads including Data Centers



***Here, Net Electric Load is meant to represent load net of BTM PV and all other modeled load categories*

---

## Example Total Load Buildup for a Data Center Heavy Utility

- **PLEXOS generates hourly loads for each load category, in each zone**
  - E3 models baseload as gross load net of electrical resistance heating load to account for fuel switching to heat pumps in the future
- **Each load type is summed for each hour to assess total energy, peaks, system shape**

Average Hourly Load Stack, 2050