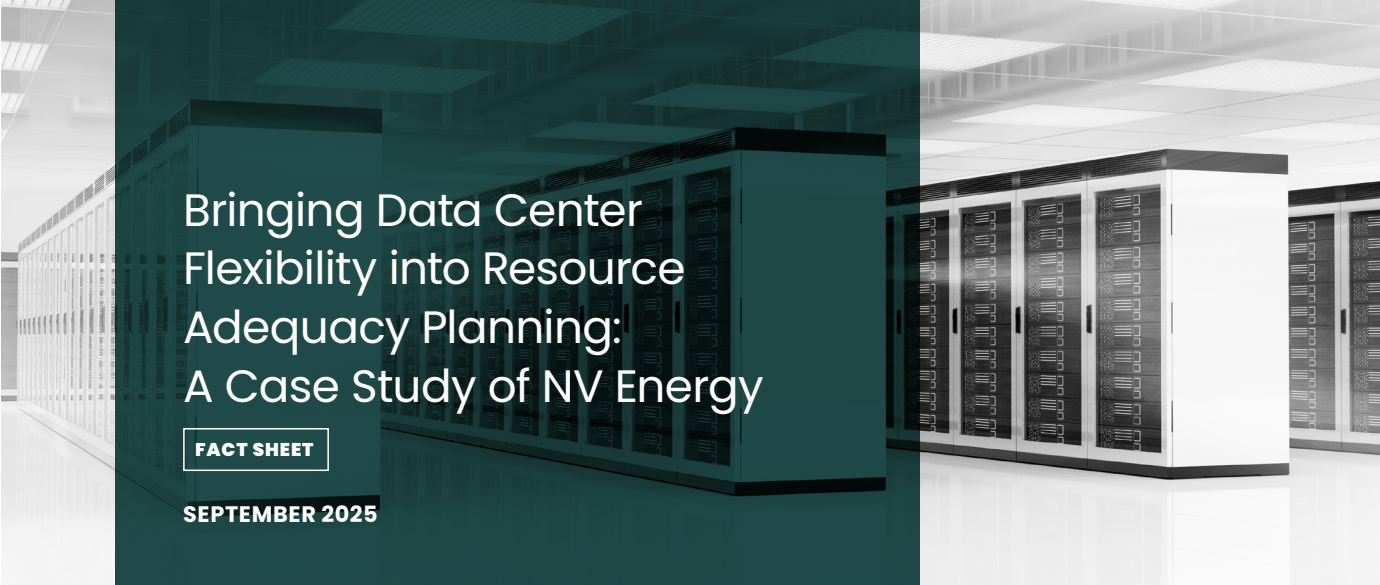




Bringing Data Center Flexibility into Resource Adequacy Planning: A Case Study of NV Energy

FACT SHEET

SEPTEMBER 2025

Background and context

Why was this study conducted, and what motivated the focus on data center flexibility in integrated resource plans (IRPs)?

This study was motivated by recent and anticipated data center load growth, projected to reach 7–12% of U.S. electricity consumption by 2028. While there is uncertainty associated with future load growth from data centers, if these projections come to fruition, data centers materially affect system reliability and resource adequacy. They are also driving major capital investment decisions for new generation and transmission, and delaying or cancelling thermal power plant retirements nationwide. In nearly all IRPs, data center demand is modeled as inflexible firm load. While this may reflect past preferences of data center customers, those assumptions could lead to significant delays in interconnecting data centers and lead to overbuilding generation resources. Flexible data center demand, either through backup generation, load curtailment, or load-shifting capabilities, can support resource adequacy, accelerate interconnection, and save billions of dollars. Recognizing, quantifying, and incentivizing this flexibility is, to-date, an unexplored tool that can help mitigate new generation capacity needs and to integrate data centers into resource adequacy frameworks and IRPs as potential reliability assets.

What are utility planners currently missing when evaluating large data center loads in resource adequacy studies?

Most resource adequacy studies treat data centers as static load additions, ignoring operating characteristics such as the ability to reduce or shift consumption during grid stress events. They often lack visibility into hourly demand characteristics, on both an aggregated and individual data center level, assuming—incorrectly—that demand is a flat block of load across all hours. In reality, the timing and quantity of data center load realization is uncertain, challenging resource adequacy planning and making solutions such as flexibility that can avoid capital-intensive investments especially valuable.

To date we are unaware of any IRPs that proactively consider potential data center flexibility to reduce or obviate the need for new capacity.

How might a failure to evaluate potential data center flexibility affect capacity expansion plans and resource procurement?

If flexibility is ignored, planners may require additional firm generation to cover peak demand and meet their reserve margin requirements, raising costs and misallocating resources. Supply chain constraints and long lead times for new gas turbines mean that new infrastructure may face delays, slowing down data center interconnections. New fossil-fuel based generation that is mainly built to support these increased capacity needs threatens utility and state climate commitments, as do delayed coal plant retirements. And because data center flexibility may require specific designs and capital investments at the time of construction, ignoring opportunities for flexibility during planning and procurement processes risks new data centers being unable to operate flexibly once online.

Treating data center demand as a fixed, static block of load, potentially inflates the planning reserve margin and requires the utility to build new resources to cover rare periods of grid reliability risk, when limited load flexibility could lead to large reduction in capacity needs.

Historically, data centers have been seen as inflexible. Why does your analysis consider approaching this topic differently?

The scale and speed of data center load growth is causing both utility planners and data center developers to think about flexibility as a mutually-beneficial opportunity. To date, most discussions of data center flexibility have not properly considered the financial incentives of the tech industry. When asking an *existing* data center of their willingness to curtail load, they will invariably say they are unable and unwilling to do so, regardless of the payment, because there is a significant financial value of lost load. However, if the discussion is framed as an opportunity for faster energization—i.e., accelerated grid connection and significant reduction in up front capital in exchange for limited amounts of flexibility—the incentives for flexibility may be better aligned. The key is to plan for data center flexibility *up front* in the development process, when timelines can be accelerated and data centers can be designed to meet specific grid needs.

Data centers developers are interested in connecting to the grid as quickly as possible, and flexibility could offer a pathway to accelerated interconnection rather than waiting in queue for new generating capacity or transmission to come online. For existing ratepayers, potential flexibility is an opportunity to support reliability and avoid expensive investments in capital infrastructure. Both are recognizing the opportunity that on-site backup generation, which data centers already include at their facilities, can play to support flexibility. Innovations in the data center design and operations, including new data center designs that support flexible operations, and new software tools that can schedule workloads to minimize resource adequacy impacts, are creating new opportunities for flexibility that did not exist in the past.

Data centers developers are interested in connecting to the grid as quickly as possible, and flexibility could offer a pathway to accelerated interconnection rather than waiting in queue for new generating capacity or transmission to come online.

Approach and findings

How did you frame your analysis and what methods did you use?

We conducted an effective load carrying capability (ELCC) analysis, comparable to the existing capacity accreditation framework used by utilities nationwide for their IRPs. While data center flexibility may be novel, we wanted to evaluate it in a way that slots into the existing utility planning framework. Our idea was to quantify data center flexibility like we do any other generating resource, allowing utilities to determine what the potential capacity contributions of flexibility would be up front in their resource planning processes.

We also chose to be technology or flexibility agnostic in our assessment. Knowing that data center utilization and configurations will change—especially based on new incentives—we modeled flexibility without a specific technology, backup generation, or type of compute load in mind. Instead, we created technology agnostic options for data center flexibility based on their availability requirements. The flexibility benefits are then quantified in terms of ELCC, and translated into portfolio capacity requirements. In this way, data centers choosing to integrate at a utility could pick the flexibility option best suited to their needs.

How did you develop your assumptions around data center flexibility? How realistic are they?

Data centers were modeled in aggregate as energy-limited resources with defined availability (e.g., 99.5–99.9% uptime) and maximum depth of response (either 1GW or 2 GW), which were used to compute a total annual energy budget that ranged from 8.8 GWh to 87.6 GWh over the course of the year.

With these energy constraints, the model could choose when, and for how long, to call on data center flexibility. Depth of response was further constrained based on the hourly data center demand profile. These flexibility needs in any given weather year are manageable, especially when spread across a fleet of data centers. On net, just a few dozen hours of load flexibility in a given year could cut the total capacity needed by over two thirds.

Just a few dozen hours of load flexibility in a given year could cut the total capacity needed by over two thirds.

Why was NV Energy selected for this analysis?

NV Energy was chosen because its 2024 IRP included a focus on data center load growth, with substantial development expected in northern Nevada near Reno. Data centers served by NV Energy are expected to account for 18% of peak load and almost 30% of annual energy demand by 2030, up from 4.7% and 8.4% in 2025, respectively. Like many other parts of the country, this represents substantial near-term demand growth. These expectations have only increased since the release of NV Energy's IRP. As a result, NV Energy provides a representative case where data centers are a main driver of future system needs and where planning assumptions around data center flexibility could meaningfully affect portfolio outcomes.

While this study is an NV Energy case study, it is intended to be generally applicable to any utility performing an IRP. As a result, it can be used as a playbook for how capacity accreditation for data center flexibility could be considered in IRPs to reduce capacity needs and accelerate interconnection of new loads.

What were the key findings of the NV Energy case study regarding cost and capacity savings?

Our modeling shows that even modest levels of load flexibility can yield large capacity savings. For example, 1 GW of data center flexibility was able to defer 600–850 MW of new firm capacity needs. Across scenarios, data center flexibility resulted in ~\$300–400 million in net present value savings over the planning horizon. These savings were achieved because the data centers provided firm capacity, resulting in the capacity expansion model selecting fewer capacity resources like natural gas CTs, and instead selecting more lower cost ‘energy’ focused resources such as solar plus storage.

Operationally, these savings were achieved while only requiring, on average, 3–42 hours per year of load flexibility across the entire participating data center fleet. And because all data centers are not required to respond to each event, data center flexibility can be “dispatched” in such a way to minimize the number of data center hours by “staging” each data center that is dispatched to their maximum amount of flexible load. This is particularly true for smaller, but more frequent events. This reduces the average number of response hours for any single data center to 1–13 hours. In practice, the flexibility response could be achieved through the use of backup generation, storage, or load curtailments, or a combination of the three. Load curtailment could consider both reductions to the data center’s cooling system energy use (for example, through setpoint changes to a chilled water system), or workload shifting.

Like all resources, capacity contributions for data center flexibility were found to saturate as participation increases, with initial increments of flexibility providing the greatest value. At very high levels of data center demand growth, load curtailments would need to become more frequent and longer in duration to continue deferring new capacity.

How can the results be applied to other regions or utilities?

The framework introduced in this study, which allows data center flexibility to be quantified and contribute to capacity requirements via a demand-side ELCC, can be replicated in IRPs and capacity market studies. While the specific value of flexibility will depend on local reserve margins, resource mix, and policy constraints, this methodology provides a transferable approach for quantifying large load contributions to resource adequacy. Utilities and ISOs can apply it to evaluate avoided capacity needs, assess opportunities for accelerated interconnection, and develop large-load rates that consider avoided costs attributed to data center flexibility.

Policy and planning implications

What steps should utilities take to better plan for large load integration and flexibility?

Utilities should begin by collecting detailed load shape and operational data for large customers, segmented by end-use and flexibility potential. They should incorporate sensi-

tivity cases in IRPs to test a range of flexibility assumptions, and evaluate demand-side ELCC contributions alongside traditional supply-side resources. Conducting capacity expansion analysis with and without incremental load flexibility will help utilities with cost allocation. In particular, this process can help utilities quantify avoided costs, emissions, and new resource additions that are attributed solely to data center flexibility. This will ultimately help them quantify the value for load flexibility as they enter negotiations with new data center customers. Proactively designing interconnection agreements and rate structures that reward flexibility will help align new data center development with system reliability needs and cost effectively meet new load growth in a prudent manner for both existing and new ratepayers.

This study considered the vertically integrated utility IRP process, but what are the implications of this study for wholesale capacity markets and ISO planning processes?

ISOs can utilize the same framework to accredit demand-side contributions from large loads. This accreditation can be used to calculate capacity obligations for data centers and load serving entities, and to allow data centers to offer flexibility into the market to offset a portion of their capacity requirement. However, market rules may need to evolve to recognize and compensate for flexible demand participation from large loads, whose size and unique operational characteristics may require new participation models.

What are the next steps for improving data center modeling and flexibility characterization in future planning?

Future analytical work should develop weather-dependent data center load profiles, evaluate correlated resource adequacy risks, disaggregate demand by compute type, and consider the specific flexibility characteristics on-site generation, storage, and workload shifting. Additional studies using more detailed, multi-region network models, are needed to account for transmission considerations. Future analytical work should also consider how changes to load profiles, resource mix, and climate over the planning horizon will change ELCC calculations, and assess how capacity requirements and their associated costs would change against a range of environmental policies, such as emissions and renewables constraints.

■ See gridlab.org/datacenter_flex for more information.

