Integrating Data Centers at the Edge: Addressing Growth and Interconnection Challenges

Aditie Garg, Nicholas Long, Roberto Vercellino, and Otto Van Geet, National Laboratory of the Rockies

ABSTRACT

The rapid growth of artificial intelligence (AI) and machine learning is driving unprecedented electric demand from data centers that is projected to increasingly strain and saturate the electric grid. By 2030, 90% of AI workloads are expected to shift toward inference-based workloads, requiring interconnection of multiple data centers (less than 20 MW) to be sited near end users. Although these data centers have relatively small individual loads, their aggregated demand can strain distribution systems, leading to multi-year interconnection delays and increased costs to customers.

This paper proposes an integrated planning approach to support the rise of distributed data centers while bolstering grid reliability, lowering infrastructure cost, and minimizing interconnection delays. Leveraging energy efficiency and demand flexibility from nearby loads can reduce the peak load of the grid, while low-grade heat from data centers can be repurposed for nearby building heating and domestic hot water needs, further reducing electric stress. This paper presents an integrated framework with an illustrative case study in the San Francisco region that combines feeder-level hosting capacity analysis, targeted demand flexibility from nearby loads, and data center heat reuse to expand distribution headroom using existing infrastructure. The results demonstrate that the interconnection of two data centers (10 MW training and 1 MW inference) could be accomplished without the need for traditional grid upgrades if coordinated demand flexibility and data center heat reuse are employed to provide a peak reduction of average 17% in distribution feeder peak demand across varying scenarios under study.

Introduction

Artificial intelligence (AI) and machine learning are reshaping data center design, deployment, and electricity consumption patterns, prompting new interconnection requests to the electric grid. The digital infrastructure sector is in its fastest expansion phase yet, driven by generative and agentic AI. Studies estimate a total global data center capacity of about 205 GW (55 GW active, 15 GW under construction, and at least 135 GW in the pipeline), with analysts expecting active capacity to triple or even quadruple within five to seven years (Infrastructure Masons 2025). By 2030, AI consumption is expected to shift toward inference (~90% of workloads), accelerating deployment at the edge and in communities worldwide (Infrastructure Masons 2025).

Recent growth in AI inference type workloads has shifted attention toward smaller, decentralized “edge” and enterprise data centers sited close to end users to meet latency requirements. These data center facilities typically interconnect at the distribution level; while each site’s load may be modest, their rapid, clustered growth can cumulatively exceed feeder and substation capacity, triggering multi-year upgrade timelines and cost allocation disputes (Shehabi et al. 2016). With the limited capacity of hosting distribution systems, even modest new large loads can force reconductoring, transformer replacements, or substation expansions (Commonwealth Edison Company 2026). Traditional interconnection and planning processes are sequential and asset-centric, often requiring years for engineering, permitting, procurement, and construction timelines, amplified by ongoing supply chain constraints for distribution and large power transformers (DOE 2023). Also, utilities face challenges in planning infrastructure for large new loads when forecasted demand growth does not fully materialize or operates below expected utilization levels. Such uncertainty can lead to stranded asset risks and upward pressure on electricity rates, highlighting the importance of carefully designing tariffs and cost-recovery structures (Riu, I et al. 2025). Developers report grid availability and power provisioning as leading constraints for project schedules and site selection, with interconnection queues lengthening and contingency margins eroding (North American Electric Reliability Corporation [NERC] 2023). Without proactive feeder and demand centric strategies, the aggregated growth of data centers can raise local peak demand, degrade reliability, and increase costs for both operators and customers.

Prior work has addressed parts of this challenge, but largely in silos. On the demand side, sector-wide studies documented significant reductions in data center energy intensity via server virtualization, improved utilization, and advanced air/liquid chip cooling, yet showed that total electricity use remains highly sensitive to workload growth and deployment choices (ASHRAE TC 9.9 2021). At the grid interface, distribution hosting capacity analysis methods have matured for distributed energy resource interconnections and are increasingly being adapted to assess "hosting capacity" for large new customers, offering feeder-level headroom estimates under grid constraints (Commonwealth Edison Company 2026). Existing frameworks demonstrate that targeted efficiency and load flexibility from energy consumers on the same distribution network can reduce peaks, load requirements, and enable participation in demand response programs (DOE and NREL 2020-2021, ACEEE 2020). Data centers could also provide direct load flexibility through workload scheduling, IT power capping, controllable cooling, and coordination of on-site uninterruptible power supplies (UPS) and backup generation (LBNL 2012). Complementary thermal strategies reclaim low-grade data center heat via heat pumps for district heating of nearby buildings, displacing electric or fossil heating and reducing local winter peaks, validated by European deployments in Stockholm and Odense and supported by ASHRAE's broader thermal guidelines (Yuan et al. 2023, Huang et al. 2020). Also, implementing heat reuse can reduce data center cooling energy requirements while simultaneously displacing building heating demand, delivering efficiency benefits on both sides of the energy system (D. Gardiner et al. 2025). Recent policy developments further highlight growing interest in this approach. Legislation passed in Virginia directs the state's energy agency to evaluate opportunities for reusing waste heat from data centers and to develop strategies to connect these facilities with potential heat users (Virginia General Assembly 2026).

What remains largely unaddressed is an integrated feeder and demand-centric planning approach that simultaneously quantifies local headroom and constraints, applies targeted energy efficiency in adjacent buildings, operationalizes local demand flexibility, and capitalizes on heat reuse to alleviate local peak demand (Garg et al. 2025). Emerging state and utility planning resources emphasize aligning interconnection processes, energy generation alternatives, and flexible load programs for large new customers, including data centers. This paper proposes and illustrates a sample study that evaluates an integrated framework combining demand flexibility from nearby buildings and data center heat reuse to expand distribution capacity, reduce interconnection delays, enhance grid reliability, and reduce infrastructure requirements (reducing overall costs for upgrades) for the rise of data centers. In addition to addressing technical interconnection challenges, these strategies align with broader policy goals of improving grid reliability while providing affordable solutions for customers. By leveraging existing infrastructure, flexible demand, and data center heat recovery, the proposed framework reduces the need for immediate capital-intensive grid upgrades. This approach supports ongoing federal and state priorities to maintain grid reliability and minimize costs to ratepayers while enabling the rapid growth of AI-driven digital infrastructure.

Data Centers and AI Workload Types

The rapid expansion of AI workloads has driven diversification in data center infrastructure. Broadly, AI-related data centers can be grouped into three categories based on their scale, purpose, and deployment characteristics: (1) hyperscale training centers, (2) enterprise/co-location centers, and (3) inference centers. Each of these categories has distinct implications for size, location strategy, latency requirements, and their potential to leverage flexible loads from nearby buildings for grid support.

Hyperscalers are designed to support large-scale AI training workloads and global cloud services. These data centers are clustered to enable parallel processing for complex models such as generative AI and large language models. These centers are large, typically exceeding 100 MW in power capacity, and ranging into hundreds of MW up to multi-GW scale for the largest deployments (Coresite n.d.). These data centers are centralized in power-rich regions, often in rural or suburban areas with access to inexpensive and abundant energy supply, fiber connectivity for high bandwidth, and favorable land availability. Training jobs are generally tolerant to higher latency because they involve batch processing and do not require real-time responsiveness.

Enterprise data centers are facilities owned or operated by a single organization to support internal computing needs, including AI inference and training workloads. Co-location centers provide shared physical infrastructure where multiple customers lease space and power for their own computing equipment (Coresite n.d.). Enterprise or co-location centers are medium scale, typically in the range of 5–100 MW of IT load capacity. These data centers are often located near business campuses, or interconnection hubs, providing a balance between connectivity and operational proximity to users and organizational sites. Latency profiles for these centers are low to medium, require more responsive performance than hyperscale training but may not demand an extremely low latency.

Inference data centers are small, distributed computing facilities located close to end users or data sources to support latency-sensitive AI inference and real-time applications. These facilities bring processing power closer to the edge of the network, reducing dependence on distant centralized infrastructure. These data centers are individually small, typically under 5 MW of capacity, and in many cases may be modular or micro-scale deployments. These data centers are deployed across local distribution networks, often on distribution feeders near population centers (e.g., retail locations, corporate campuses) to minimize network latency to end users. Low latency is the defining characteristic of inference data centers, to meet the demands of real-time AI inference and control loops.

Across AI infrastructure, hyperscale, enterprise/co-location, and inference data centers represent a continuum of scale, siting, and grid interaction. According to Infrastructure Masons, it is expected that 90% of AI workloads will shift toward inference workloads that use AI trained models for inferencing (Infrastructure Masons 2025). As AI workloads shift from centralized training toward distributed inference, the grid integration challenge moves downstream, from transmission-level interconnections toward feeder- and substation-level constraints where coordination with nearby building loads can provide meaningful flexibility and capacity relief. Nationwide, utilities are projecting substantial load growth due to AI and increased data center interconnection requests (Wilson et al. 2024).

Approach

This paper proposes an approach to provide integration pathways for faster interconnection of data centers in the grid by coordinating building loads and thermal energy solutions among energy consumers on the same distribution network as data centers to reduce peak electrical demand while satisfying grid reliability and service level requirements. The approach leverages demand flexibility from nearby loads as well as data center heat reuse for electric- and thermal-demand reduction without compromising grid performance.

At a neighborhood scale, aggregated demand flexibility reductions can meaningfully expand feeder capacity, enabling siting and faster interconnection of data centers. Load-shifting strategies that exploit both electric and thermal flexibility can transfer demand from peak to off-peak periods in response to grid conditions in real time.

Data centers located on the distribution grid (edge of the grid) provide an additional opportunity due to their proximity to end-use thermal loads (e.g., multifamily domestic hot water loads), enabling heat recovery to reduce local electric heating demand. By capturing and repurposing low-grade thermal energy for neighboring heating load, data centers can offset electric heating demand and alleviate feeder stress.

The demand flexibility and data center heat reuse strategies are defined below.

Strategy #1: Demand flexibility. The net feeder electric demand is defined as the sum of electric residential ($P_{res_{electrical}}$) and commercial ($P_{com_{electrical}}$) building load, as defined in (1).

$$P_{base_{elect}}(t) = P_{res_{electrical}}(t) + P_{com_{electrical}}(t) \tag{1}$$

In this case demand flexibility is showcased by relaxing thermostat control setpoints (heating and cooling) in residential and commercial buildings along with dimming lighting systems on commercial building to reduce the building's daily electricity peak demand during demand flexibility events (Reyna et al. 2025, Xiong 2024). Thus, the reduced net electric demand at the feeder due to demand flexibility is defined as (2):

$$P_{net_{elect}}(t) = P_{base_{elect}}(t) - \Delta P_{flex}(t) \tag{2}$$

Where, $\Delta P_{flex}(t)$ is the dispatchable reduction from building demand flexibility lighting and HVAC thermostat control. In residential buildings HVAC settings are shifted with a 4°F offset during on-peak periods to reduce peak load of the building.

Strategy #2: Data center heat reuse. To reduce electric load needs for heating loads, residential and commercial heating loads are considered as an offtake for the data center heat. In this case heat reuse is modeled via principles of energy balance for the data centers. In steady-state operation, nearly all electrical energy consumed by the IT equipment and facility auxiliaries (e.g., power supplies,) is ultimately dissipated as heat. Neglecting transient storage and non-thermal exports, it is assumed the rate of heat rejection equals the electrical input power (Sun 2022). Accordingly, the instantaneous heat source (P_{heat}) is approximately equal to the data center electrical demand (P_{DC}) as defined in (3).

$$P_{heat} \approx P_{DC} \tag{3}$$

Treating the heat output as equal to the total data center electrical load provides a reasonable upper bound on available thermal energy for recovery. Practical delivery to neighboring loads is limited by heat recovery, distribution effectiveness, and available nearby heating demand; we represent this with a recovery factor η_{rec}. This recovery factor includes heat pump conversion that distributes energy within the thermal networks and assumes change in temperature of water sufficiently meeting data center supplied temperature requirement. Thus, the usable data center heat for nearby building heat loads is defined as (4).

$$P_{heat\,usable} \approx \eta_{rec} * P_{heat} \tag{4}$$

The reduced net thermal demand at the feeder due to data center heat recovery is defined as (5):

$$P_{net_{thermal}}(t) = P_{res_{thermal}} + P_{com_{thermal}}(t) - \eta_{rec} * P_{heat} \tag{5}$$

Under the assumed recovery effectiveness η_{rec} and given the observed mix of heat end uses with an effective coefficient of performance (COP) [1], the net feeder electric demand is calculated as:

$$\begin{aligned} P_{net}(t) &= P_{net_{elect}} + P_{DC} + \frac{P_{net_{thermal}}}{COP} \\ &= \left[P_{base_{electrical}}(t) + P_{DC}(t) - \Delta P_{flex\,comm}(t)\right] + \\ &\quad \left[\frac{P_{res_{thermal}}(t) + P_{com_{thermal}}(t) - \eta_{rec} * P_{heat}(t)}{COP}\right] \end{aligned} \tag{6}$$

Together, these measures are evaluated to reduce the net electric demand by providing demand flexibility during peak hours and supplying required heating loads during available data usable heat.

Results

To evaluate the load-reduction strategies described above namely building demand flexibility and data center heat reuse a representative urban distribution feeder from the San Francisco region is analyzed using the feeder from the SMART-DS dataset (SMART-DS v1.0 n.d.). The feeder serves a mixed residential and commercial area with peak electrical demands of

[1] COP value may vary based on customer equipment and locations.

3.8 MW (residential) and 7.6 MW (commercial). The feeder head transformer is rated at 16 MVA and supplies 155 secondary transformers rated at 75 kVA each. The feeder topology and study region are shown in Figure 1.

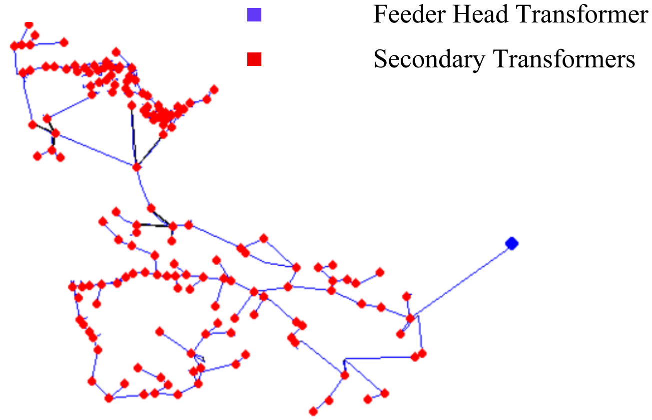

Figure 1. San Francisco area feeder under analysis (https://data.openei.org/s3_viewer?bucket=oedi-data-lake)

The baseline electrical and thermal load profiles were generated for buildings with electric only fuel types in the California region from ResStock™ and ComStock™ data (Commercial and Residential Load Profiles 2026). The thermal and electric load profiles are shown in Figure 2. In this analysis, the annual peak electrical load occurs on February 19, reaching approximately 11 MW. Thermal demand is dominated by commercial customers, which account for approximately 66% of total thermal load, with a peak thermal demand of 22 MBtu/h observed on February 24. Under baseline conditions, the feeder has approximately 5 MW of available headroom for new interconnections.

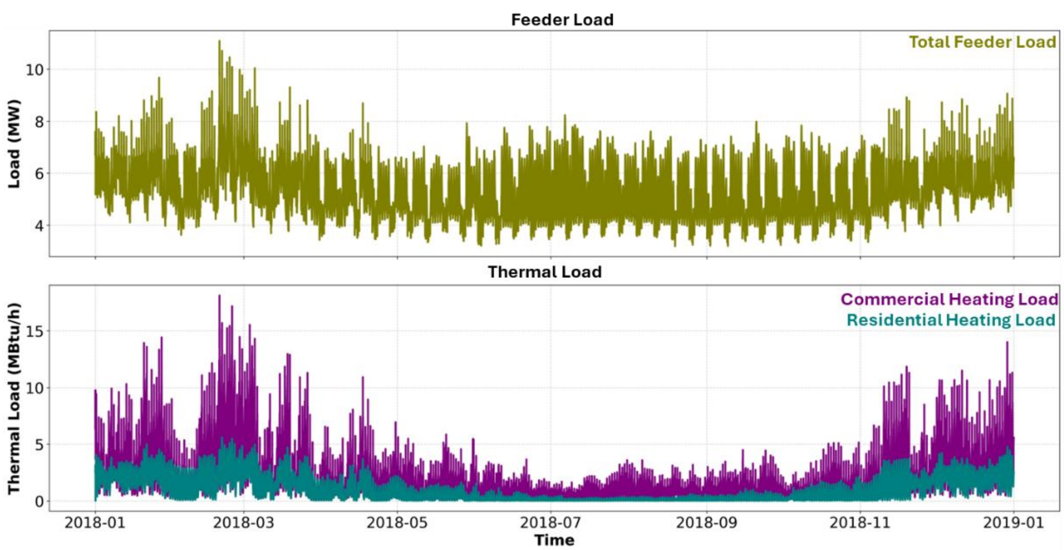

Figure 2. Base feeder electric and thermal load with COP = 1[2]

The feeder is assumed to interconnect two data centers: a 10-MW co-location AI facility and a 1-MW inference-oriented AI facility, operated at 60% and 80% of the rated capacity,

[2] Thermal loads are calculated based on electric heat consumption data from the ResStock and ComStock database considering a COP factor of 1 in this case.

respectively. The corresponding load profiles for colocation and inference data centers are shown in Figure 3 (Vercellino et al. 2025). These load profiles were generated using DIPLOEE (Datacenter Planning, Load Optimization for Energy Efficiency), the National Laboratory of the Rockies' bottom-up discrete simulation tool for data center modeling. DIPLOEE combines a library of power samples from typical AI workloads measured on the National Laboratory of the Rockies' cluster with facility-level utilization and user behavior distributions to simulate the submission, scheduling and execution of AI workloads at data centers, while tracking power consumption and throughput.

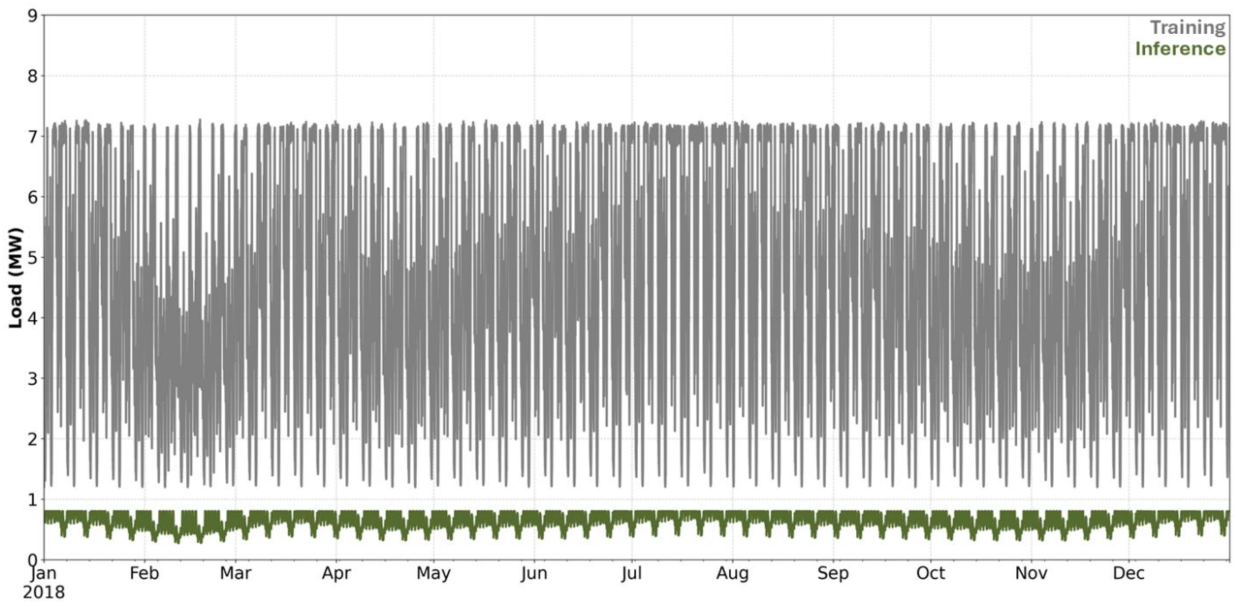

Figure 3. Data center load profiles for training and inference type workloads

With the interconnection of both data centers, the feeder's peak demand increases to 16.5 MW and the peak load day changes to March 15. The peak load exceeds the rated capacity of the feeder head transformer, as shown in Figure 4.

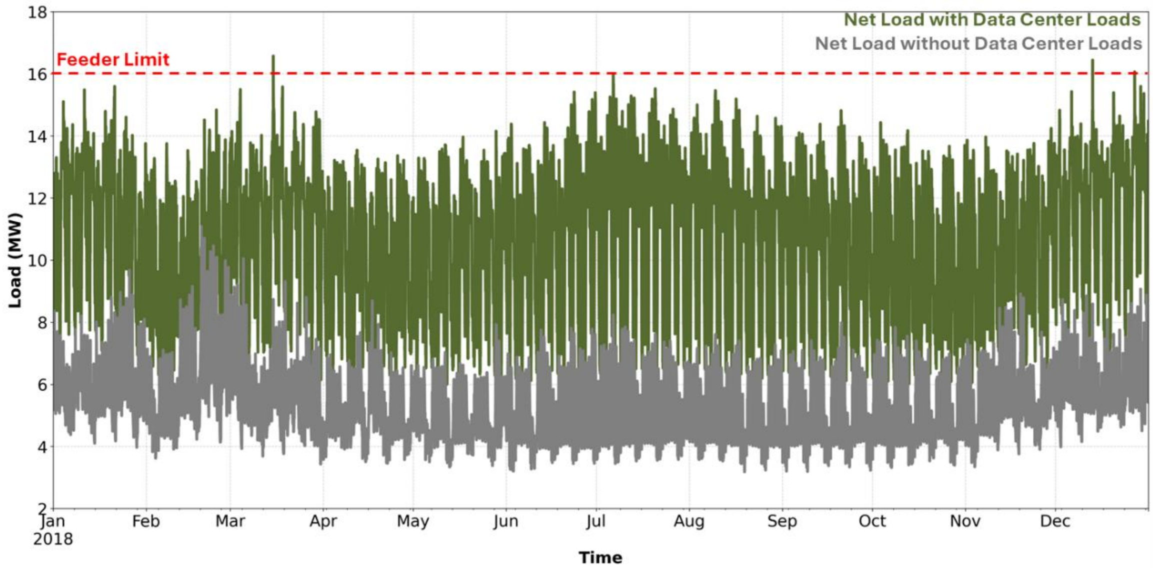

Figure 4. Feeder base load with and without data center interconnection

To enable interconnection without conventional capacity upgrades, two mitigation cases are evaluated:

1. Demand flexibility implemented through building lighting dimming and HVAC thermostat adjustments in participating residential and commercial buildings during peak events.
2. Demand flexibility combined with data center heat reuse to offset nearby building thermal loads, thus reducing electric (and gas) peak demand.

These cases are evaluated to quantify their ability to restore feeder headroom and maintain operation within the feeder's limits while accommodating new data center loads.

Strategy #1: Demand flexibility. Performance was assessed over a one-year period. Using this strategy, the peak electric load of the feeder is reduced by 0.2 MW (2.4% of peak base electric load reduction). The feeder base load (electric residential and commercial load) and managed load with reduced load due to demand flexibility as defined in equation 2 s is highlighted in Figure 5.[4] It is observed the peak load is reduced and shifted to off peak hours for 3 days continuously, based on the demand flexibility measures adopted during the analysis.

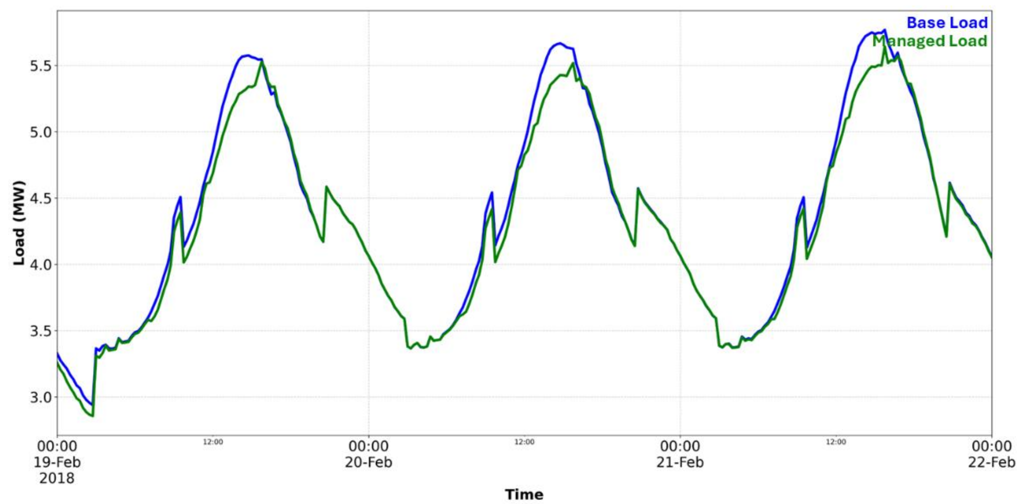

Figure 5. Feeder load on specific days in the grid comparing base load vs. building load with demand flexibility[3]

Strategy #2: Data center heat reuse. Waste heat from data centers can be recovered and used to serve thermal loads in nearby residential and commercial buildings. Based on prior studies, this analysis assumes a delivered thermal recovery factor of ($\eta_{rec} = 0.6$) (Pecchioli et al. 2025). Figure 6 shows the data center electrical load from the two data centers (10 MW and 1 MW) and the corresponding profile of recoverable heat with $\eta_{rec} = 0.6$ according to equation 4.

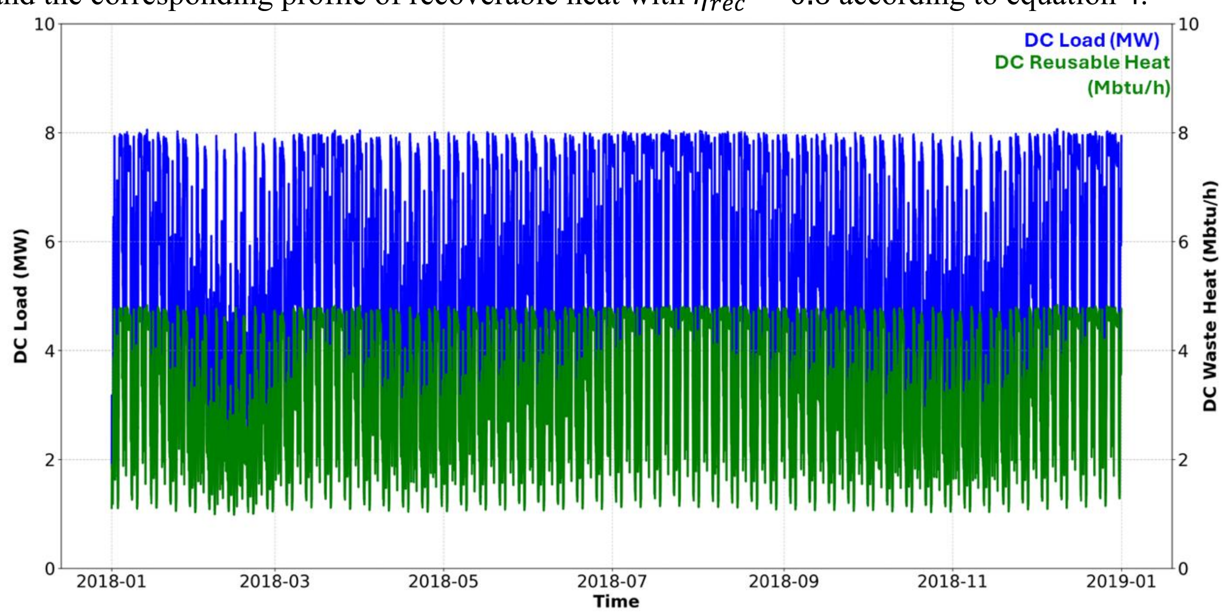

Figure 6. Data center load and usable data center (DC) heat profiles

The results indicate substantial recoverable thermal energy when proximate loads with heating demand are available. In aggregate, the data centers provide approximately 113.5 GBtu of recoverable heat over the study period, with an average of 311 MBtu/h per day of usable

[3] Note this study assumed a preliminary flexibility solution that included flexible loads from residential and commercial buildings with the building lighting dimming and HVAC thermostat adjustments in participating residential and commercial buildings during peak events. In reality various other solutions such as feeder peak load reduction management solution can be deployed along with energy efficiency to achieve larger peak load reduction.

thermal output. However, during several hours, excess unused data center heat (shown in Figure 7) remains unutilized due to insufficient thermal demand in the residential and commercial buildings thermal loading. This unused heat will still require conventional heat rejection methods such as cooling towers or dry coolers in order to protect the data center's IT equipment. That may require additional energy or other solutions such as thermal storage to store the excess heat, but these solutions are ignored in this analysis. Moreover, the utilization rate of DC heat is low due to differences in DC heat profiles and thermal load shapes with peak of DC heat reuse not aligning with thermal load shapes as highlighted in Figure 8. On March 15 and March 14, the data center reusable heat is seen to have a lower or rather flat profile during the peak hours thus reducing thermal load marginally, while leaving a significant data center heat unused during off peak load hours.

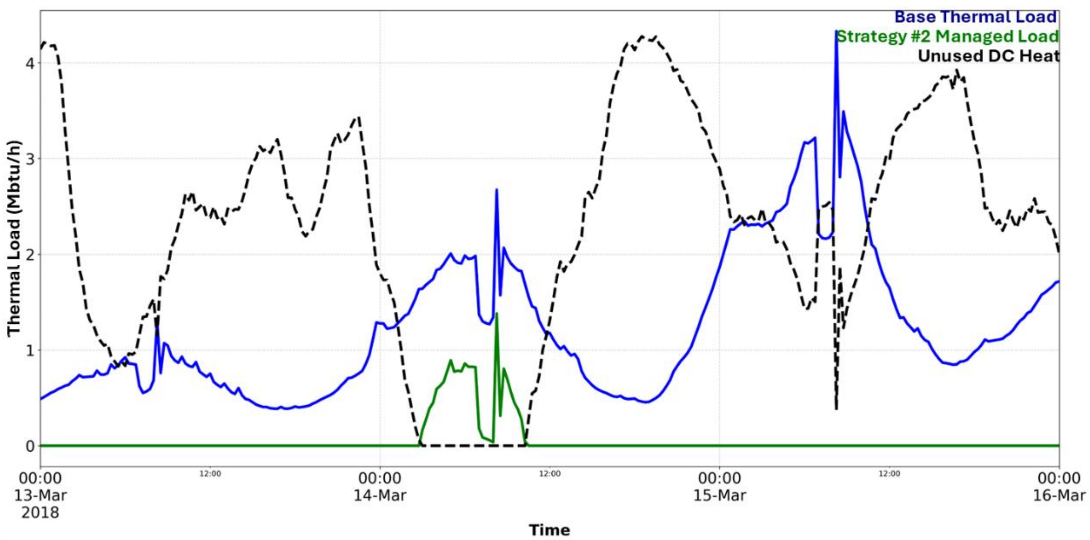

Figure 7. Thermal load profiles: base residential and commercial thermal load vs managed residential and commercial thermal load vs unused data center heat[4]. Base thermal load is the residential and commercial heat load, managed load is the reduced feeder heating load by utilizing data heat reuse as defined in equation 5 and unused DC heat is the excess DC heat that remains unused due to insufficient offtakes (heating load) for DC heat.

[4] These results are shown for the peak load day (March 15) with DC integration

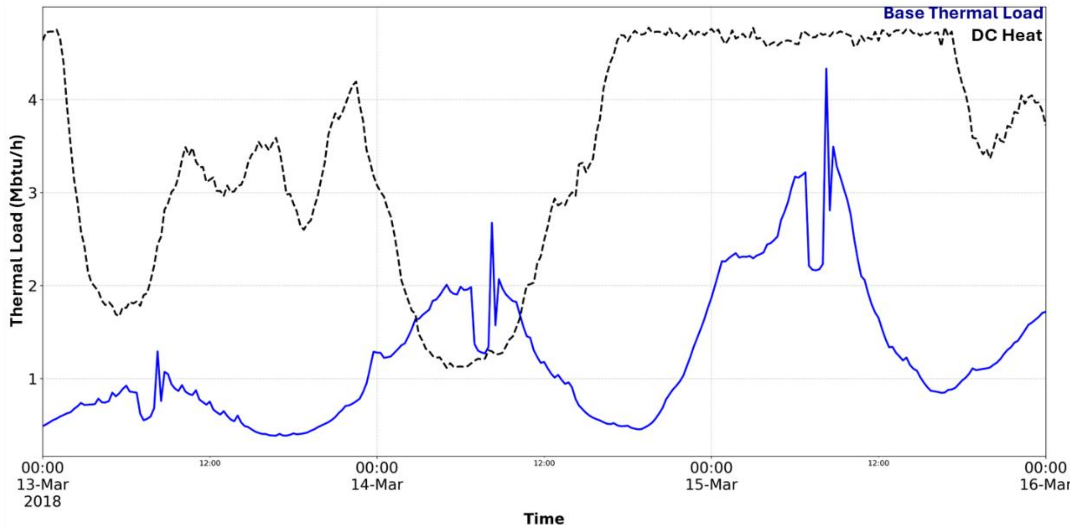

Figure 8 Base thermal load shape with DC heat reuse profiles for representative days. Base thermal load is the residential and commercial heat load; DC heat is the available reusable heat in the feeder for the peak load days.

As illustrated in Figure 9, nearly 77% of the available data center heat is unused, highlighting additional opportunities for thermal load displacement in locations with larger heating demand.

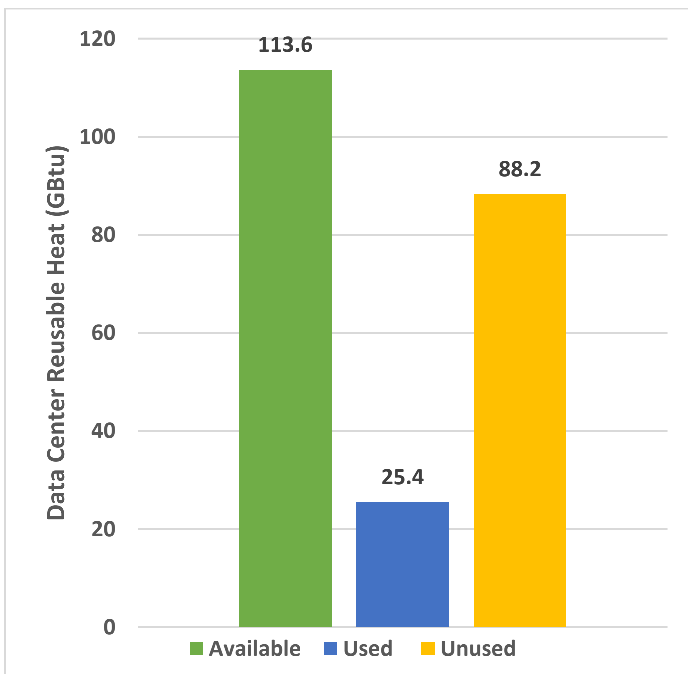

Figure 9. Data center reusable heat available vs. used vs. unused

As highlighted earlier, the feeder observes a peak load of 16.5 MW on March 15. Combining demand flexibility with data center heat reuse for residential and commercial heating loads, the peak feeder reduces by 4.4 MW, reducing the peak load to 12.1 MW. [5] Out of the 4.4 MW peak load reduction, 0.1 MW is reduced by residential and commercial demand flexibility and remainder of 4.3 MW is reduced from DC heat reuse highlighted in Figure 10[6]. Given the data center heat was only utilized at 23% in this case, additional peak relief from heat reuse is expected in locations with higher heating demand; for example, in colder regions with higher space heating demand, a larger fraction of data center heat could be utilized (assuming coincident demands).

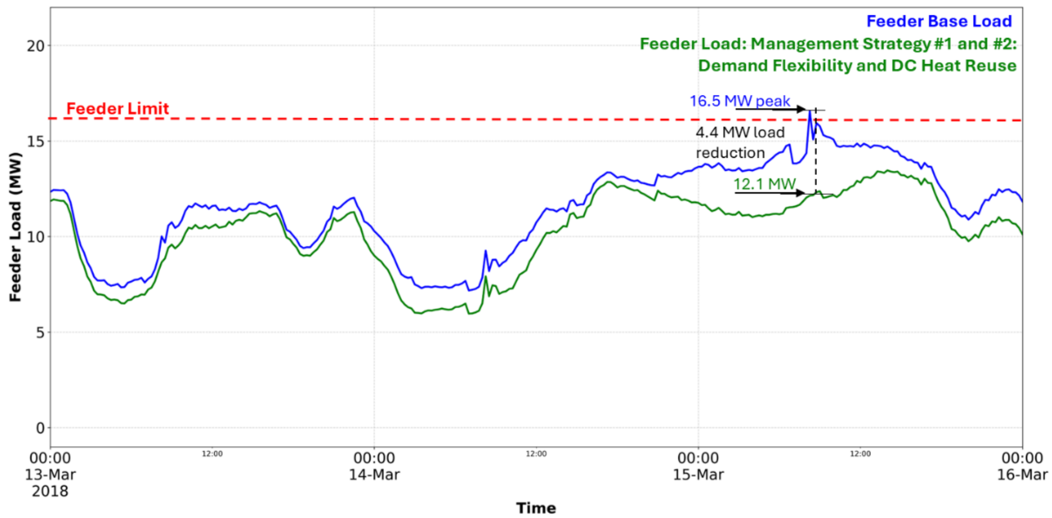

Figure 10. Feeder load with and without demand flexibility and data center heat reuse measures[6]

Given in a feeder varying COP factors may exist for the thermal loads, the results with varying COP factors is analyzed to include a COP factor of 2 and 3. Considering the same data center heat reduction base thermal loads, the peak load reductions are highlighted Figure 11. Combining demand flexibility with DC heat reuse the peak load reduction decreases as assumed COP increases, with peak load reduction falling from 4.4 MW to 1.6 MW and energy reduction from 27.6 GWh to 19.9 GWh across the COP factors from 1 to 3. Across all evaluated COP values, the combined peak-load mitigation strategies including demand flexibility and data center waste heat reuse reduced feeder peak demand and total energy consumption by an average of 17% and 6.3%, respectively, as summarized in the table below.

Table 1 Feeder peak load and energy reduction in percentages

	Peak Load Reduction (%)	**Energy Reduction (%)**
COP=1	26.7	7.4
COP=2	14.5	6.2
COP=3	9.7	5.3
Average	17	6.3

[5] Under the assumed recovery effectiveness η_{rec} and given the observed mix of heat end uses with an effective coefficient of performance (COP) of 1

[6] Feeder base load includes electric load, data center load along with heating loads converted to electric with COP =1. While the managed load (including strategy #1 and #2) includes data center load, managed electric load with demand flexibility and reduced thermal load converted to electric with COP of 1.

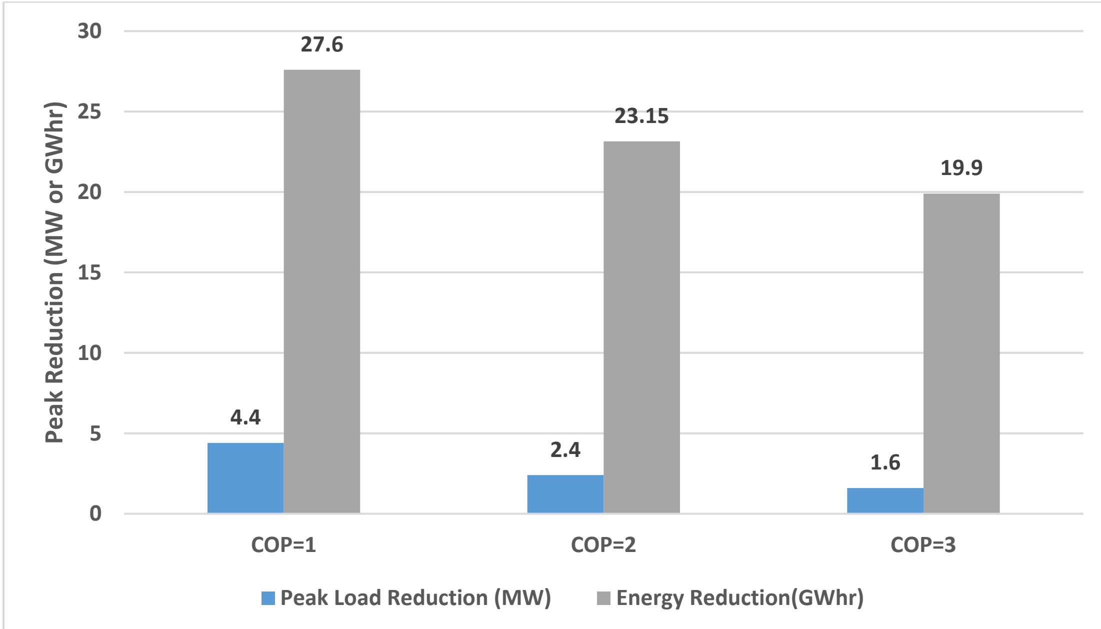

Figure 11 Load Reduction (Peak load and Energy) for varying COP values ranging from 1 to 3.

In summary, the demand flexibility-only case achieves a peak reduction of 0.2 MW, while the combined case incorporating data center heat reuse achieves a load reduction ranging from 4.4 MW to 1.6 MW for varying COP factors. In all cases the reduction is sufficient to maintain operation within the feeder's thermal limit. Importantly, these results are achieved without curtailing data center IT loads, demonstrating the potential of coordinated building and data center integration to unlock distribution hosting capacity.

Conclusion

This paper demonstrates that an integrated approach combining building demand flexibility and data center heat reuse can expand distribution hosting capacity and enable timely interconnection of emerging AI data centers without curtailing IT loads. Using an urban prototype feeder with peak load of 16.5 MW, coordinated lighting and HVAC flexibility reduced peak demand by 0.2 MW, while the combined strategy incorporating data center heat recovery achieved load reduction ranging from 4.4 to 1.6 MW for COPs from 1 to 3. In all scenarios, the total electric load reduction resulted in restoring feeder headroom and maintaining operation within transformer thermal limits. These operational measures deliver capacity relief comparable to traditional infrastructure upgrades, but on significantly faster timelines and with lower capital intensity.

Beyond enabling interconnections, the proposed approach directly improves grid reliability and affordability. By reducing peaks at the feeder head, these measures lower transformer loading, mitigate thermal stress on conductors, and reduce the likelihood of overload driven outages during extreme conditions. Leveraging existing building loads, controls, and recoverable data center heat shifts capacity creation from asset-heavy upgrades to operational optimization of resources already available to the grid. This reduces the scale and urgency of reconductoring, transformer replacements, and substation expansions lowering total system costs and limiting cost pass-through to customers.

The findings are especially relevant given the projected shift of AI workloads toward inference and the resulting growth of enterprise and edge data centers interconnecting at the distribution level. While individual facilities may be modest in size, their clustered deployment can quickly saturate feeder and substation capacity, triggering costly multi-year upgrade cycles. Aligning feeder capacity analysis with targeted building efficiency, grid-responsive loads, and thermal reuse unlocks near-term headroom, shortens interconnection timelines, reduces costs, and enables incremental capacity expansion using existing infrastructure.

More broadly, this framework provides a practical pathway to co-optimize digital infrastructure growth with community energy needs. Data center heat reuse displaces electric space and water heating during winter peaks, while flexible building loads provide fast-acting capacity relief during critical hours. It must be noted the demand flexibility measure may provide additional relief if feeder specific control strategies are considered. Also, the paper showcases underutilization of available data center heat, which suggests even greater benefits can be achieved in colder climates, and denser thermal-load corridors.

This study focuses on the technical potential of coordinated demand flexibility and heat reuse, practical considerations remain open. The capital cost and development timelines of thermal energy networks with heat-recovery capability are active areas of research, and their competitiveness relative to traditional electric distribution upgrades depends greatly on local utility programs, incentives, and customer adoption. These uncertainties underscore the need for future work assessing customer requirements, economic feasibility, and viable business models for cost recovery. Acknowledging these challenges is essential, even as emerging district-heating pilots demonstrate that data-center heat reuse can be successfully implemented under the right conditions.

Future work could extend this framework across diverse feeder types, and locations that evaluates incentives and market structures needed to scale coordinated approach towards utilizing building loads as a resource for data center integration. Collectively, these strategies offer a cost-effective and reliable grid solution to traditional grid upgrades for supporting rapid AI-driven load growth at the grid.

Acknowledgments

This work was authored by the National Laboratory of the Rockies for the U.S. Department of Energy (DOE), operated under Contract No. DE-AC36-08GO28308. Funding provided internally by the National Laboratory of the Rockies. The views expressed in the article do not necessarily represent the views of the DOE or the U.S. Government. The U.S. Government retains and the publisher, by accepting the article for publication, acknowledges that the U.S. Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this work, or allow others to do so, for U.S. Government purposes.

References

ACEEE. 2020. *Grid-Interactive Efficient Building*. https://www.aceee.org/grid-interactive-efficient-buildings-gebs

ASHRAE TC 9.9, Thermal Guidelines for Data Processing Environments, 5th Edition, 2021. https://www.ashrae.org/file%20library/technical%20resources/bookstore/supplemental%20files/therm-gdlns-5th-r-e-refcard.pdf

Commercial and Residential Load Profiles. 2026. https://data.openei.org/s3_viewer?bucket=oedi-data-lake&prefix=nrel-pds-building-stock%2F

Commonwealth Edison Company. "Multi-Year Integrated Grid Plan and Direct Testimony of Commonwealth Edison Company." Illinois Commerce Commission, Docket No. P2026-0047 2026. https://www.icc.illinois.gov/docket/P2026-0047/documents/375126

Coresite, n.d. https://www.coresite.com/blog/five-types-of-data-centers

DOE and NREL. 2020–2021. *Grid-Interactive Efficient Buildings (GEB) Roadmap and Technical Reports*. https://www.energy.gov/eere/buildings/geb-technical-reports

DOE. 2023. *Transformer Supply Chain Assessments and Industry Updates* (distribution and large power transformers). U.S. Department of Energy. 2023. https://www.energy.gov/sites/default/files/2023-06/Implementing%20Final%20Guidance%20Distribution%20Transformers%20May%202023%20Update.pdf

Garg, A., C. Booten, O. VanGeet, S. Pless, K. Stenger, J. Vidal, R. Jackson, L. Lavin, and K. Prabakar. 2025. *Considerations for Distributed Edge Data Centers and Use of Building Loads to Support Large Interconnections*. Golden, CO: National Renewable Energy Laboratory. NREL/TP-5500-96700. https://docs.nlr.gov/docs/fy26osti/96700.pdf.

Huang, P., B. Copertaro, X. Zhang, J. Shen, I. Löfgren, M. Rönnelid, J. Fahlen, D. Andersson, M. Svanfeldt. 2020. "A review of data centers as prosumers in district energy systems: Renewable energy integration and waste heat reuse for district heating", Applied Energy, Volume 258, 2020, 114109, ISSN 0306-2619, https://doi.org/10.1016/j.apenergy.2019.114109.

Infrastructure Masons. 2025. "State of the Digital Infrastructure Industry: Annual Report 2025." https://imasons.org/wp-content/uploads/2025/05/State-of-the-Infrastructure-Industry2025.pdf.

LBNL. 2012. *Demand Response Opportunities and Enabling Technologies for Data Centers*. https://gridintegration.lbl.gov/publications/demand-response-opportunities-and

North American Electric Reliability Corporation (NERC). 2023 Long-Term Reliability Assessment. https://www.nerc.com/globalassets/programs/rapa/ra/nerc_ltra_2023.pdf

Pecchioli, C., A. Laval, J. Comella, D. Sickinger, O. VanGeet. 2025. "Data centers heat reuse 101". https://www.opencompute.org/documents/20230623-data-centers-heatreuse-101-3-2-docx-pdf

Shehabi, A. et al., 2016. *United States Data Center Energy Usage Report*, Lawrence Berkeley National Laboratory, 2016. https://eta-publications.lbl.gov/sites/default/files/lbnl-1005775_v2.pdf

SMART-DS Dataset, v1.0. n.d. Dataset., https://data.openei.org/s3_viewer?bucket=oedi-data-lake

Sun, J., S. Wang, J. Wang and L. M. Tolbert, "Dynamic Model and Converter-Based Emulator of a Data Center Power Distribution System," in *IEEE Transactions on Power Electronics*, vol. 37, no. 7, pp. 8420-8432, July 2022, doi: 10.1109/TPEL.2022.3146354. https://www.osti.gov/biblio/1982893

Vercellino, R., J. Willard, G. Campos, W. da Silva Pereira, O. Hull, and J. Mueller. 2025. "Measurement of Generative AI Workload Power Profiles for Whole-Facility Data Center Infrastructure Planning," *(submitted to) Nature Energy*.

Wilson, J., Z. Zimmerman, and R. Gramlich. 2024. "Strategic Industries Surging: Driving US Power Demand." Grid Strategies, LLC. https://gridstrategiesllc.com/wp-content/uploads/National-Load-Growth-Report-2024.pdf.

Yuan, X., Y. Liang, X. Hu, Y. Xu, Y. Chen, and R. Kosonen. 2023. "Waste heat recoveries in data centers: A review." *Renewable and Sustainable Energy Reviews*, Volume 188, 2023, 113777, ISSN 1364-0321, https://doi.org/10.1016/j.rser.2023.113777.

Riu, I., Patel, Kushal; Mettetal, Liz; Santoni-Colvin, Morgan; et al. *Tailored for Scale: Designing Electric Rates and Tariffs for Large Loads.* Energy and Environmental Economics, Inc. (E3). December 2025. Available at: https://www.ethree.com/ratepayer-study/

D. Gardiner, A. Dixon, E. Kent, and H. Schuster. 2025. "Data Center Heat Reuse: The Opportunity for States", https://www.dgardiner.com/wp-content/uploads/2025/12/Data-Center-Heat-Reuse-The-Opportunity-for-States.pdf

Virginia General Assembly, *Data centers; Department of Energy to identify opportunities for use of waste heat, report*, H.B. 323, 2026 Regular Session. https://lis.virginia.gov/bill-details/20261/HB323

Xiong, Jie, and Janghyun Kim. 2024. End-Use Savings Shapes Measure Documentation: Dispatch Schedule Generation for Demand Flexibility Measures. Golden, CO: National Renewable Energy Laboratory. NREL/TP-5500-89343. https://www.nrel.gov/docs/fy24osti/89343.pdf.

Reyna, Janet, et al. 2025. ResStock Technical Reference Documentation v3.3.0. Golden, CO: National Renewable Energy Laboratory. NREL/TP-5500-91621. https://www.nrel.gov/docs/fy25osti/91621.pdf.

Disclosures

AI-assisted tools used: ChatGPT
Used for: light editing
Human review: The authors reviewed and verified all content, calculations, and citations.