Measuring What Matters: A New Framework for Data Center Performance

September 2026

Prepared by:

Nora Wang Esram, CEO
Reilly Loveland, Associate Director
Mariel Wolfson, Senior Technical Writer
Anna Johnson, Senior Project Manager
Sayali Lamne, Project Manager

Contents

List of Figures

List of Tables

Acknowledgements

This report was made possible through the generous support of Blue Horizons. The authors are grateful to external and internal reviewers, colleagues, and sponsors for their support. We particularly thank the following external reviewers for invaluable feedback that improved the quality of the report:

Melissa Baker, Elizabeth Beardsley, Gail Hampsmire, Ian LaHiff, and Lucas Tuffoli, U.S. Green Building Council

Christopher Binder and Sara Rademacher, Green Building Initiative (GBI)

Danielle Byrnette and Jeff Loiter, National Association of Regulatory Utility Commissioners (NARUC)

Alex Dews, Institute for Market Transformation (IMT)

Jay Dietrich and Jay Paidipati, Uptime Institute

Holly Elwood, Global Electronics Council (GEC)

Miranda Gardiner, iMasons Climate Accord (iCA)

Patrick Hughes, National Electrical Manufacturers Association (NEMA)

Abe Silverman, SilverGreen Consulting

Adam Sledd, Dominion Energy Innovation Center

Rodney Sobin, National Association of State Energy Officials (NASEO)

Forest Tanier-Gesner, PAE Consulting Engineers

Brian J. Walker, Ph.D.

Thank you to the following NBI internal reviewers:

Tristan Grant, Director of Codes and Policy

Alexi Miller, Director of Building Innovation

Tegan Molloy, Director of Leadership and Market Development.

External review and support do not imply affiliation or endorsement. Lastly, the authors thank the NBI strategic communications team, Sarah Cadmus Burgher and Zack Loehle, as well as Cynthia Merse for copyediting and Marbry Walker for graphic design.

Executive Summary

A new era of data center growth requires new rules. Data centers that run artificial intelligence (AI) workloads are among the fastest-growing sources of electricity demand in the United States. In many regions, data centers are already straining grid infrastructure, raising costs for ratepayers, and inciting strong community opposition. As demand accelerates, policymakers, utilities, and communities are increasingly asked to evaluate projects without a consistent, credible way to assess the impacts of data centers on the electric grid, climate goals, economic development, or local communities. In response, more communities and states are setting moratoria on further data center buildout, which pushes data center development into less-regulated regions of the country.

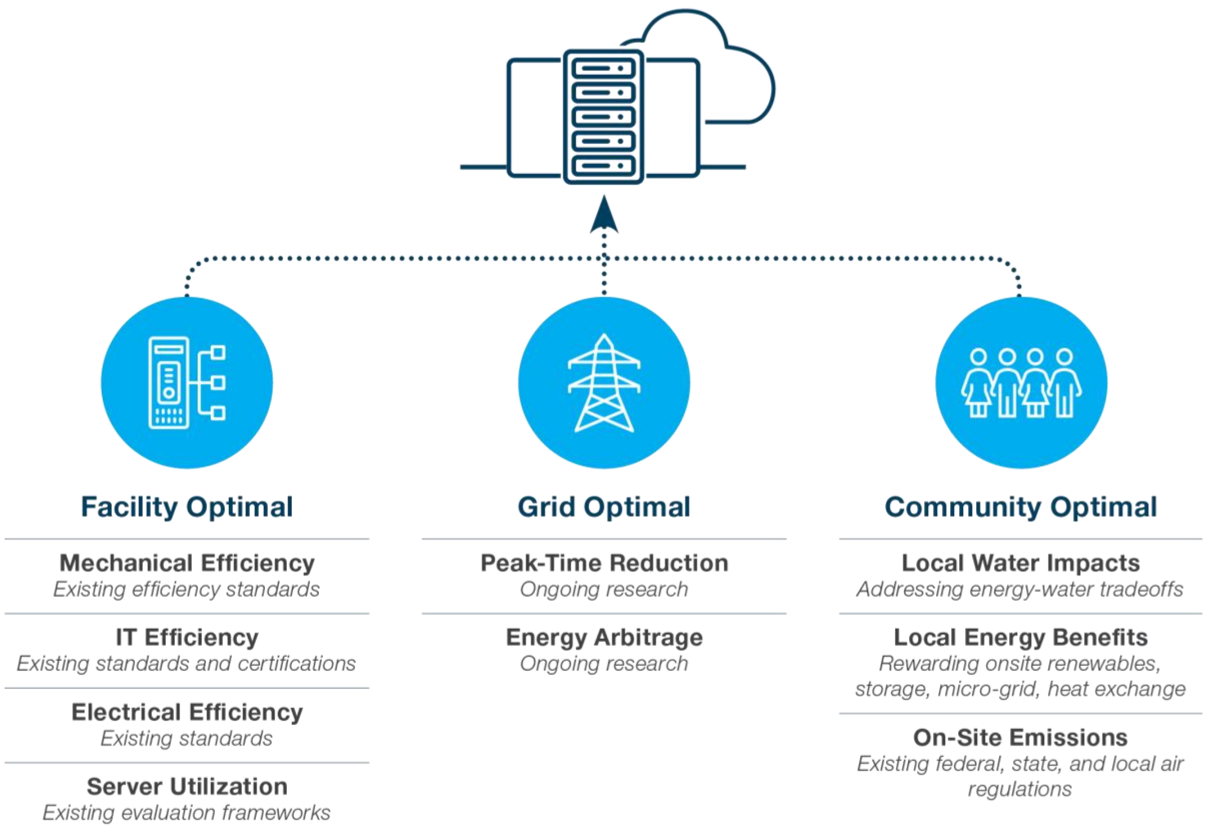
The lack of effective measurement of responsible data center design and operational practices has created a critical obstacle to data center development that would balance the interests of data center owners, utilities, and surrounding communities. Without shared metrics, standards, and performance targets, stakeholders often pursue competing priorities. This leads to fragmented decision-making and missed opportunities to align data center growth with energy goals, climate objectives, and risk mitigation for surrounding communities.

Measure what matters. NBI's technical report "The Data Center Energy Performance Evaluation Framework (DC-EPEF): Practical Standards for Facilities, the Grid, and Communities," summarizes current data center standards and metrics, identifies key gaps and opportunities, and outlines ways to leverage existing regulations and incentives to advance technologies that strengthen the grid and protect the environment. NBI proposes a **future-ready data center evaluation framework** as the best way to support state and local decision makers. The framework targets four dimensions of performance:

1. **Energy, water, and emissions metrics** that go beyond mechanical- and electrical-system efficiency to capture total resource consumption and environmental impact.
2. **Grid stress and demand flexibility targets** that measure a data center's ability to adjust workload and energy demand in response to grid conditions—which are increasingly critical as large loads multiply.
3. **Local flexibility and technological adaptability** to account for regional grid, water, climate, and community conditions while keeping standards current with rapid industry change.
4. **Tiered requirements** combining mandatory minimums tied to grid interconnection with more ambitious targets linked to financial incentives and certification—driving cost-effective adoption at scale without sacrificing ambition.

To address this need, NBI is initiating a **Data Center Energy Performance Evaluation Framework (DC-EPEF).** The DC-EPEF is not meant to replace existing codes and standards, or rating and certification systems for data centers. Instead, it builds upon the robust and ongoing efforts across industry, government, utilities, and standard-setting organizations. The DC-EPEF will allow policymakers and other key stakeholders to holistically benchmark data center performance and clarify how data center energy performance translates into benefits across energy, water, emissions, and grid reliability criteria, resulting in data centers that are facility-optimal, grid-optimal, and community-optimal.

DATA CENTER ENERGY PERFORMANCE EVALUATION FRAMEWORK (DC-EPEF)



View the detailed description for the DC-EPEF.

Data center growth is outpacing current codes, standards, and rating and certification systems. To reinforce and replicate best practices, there is a need for more effective and comprehensive metrics and evaluation criteria to establish a consistent baseline of social and environmental accountability and set expectations around data center impacts on energy, water, and emissions.

Today’s complex landscape of codes and standards faces three major challenges:

1. Energy codes focus primarily on the efficiency of mechanical and electrical systems while excluding other key factors related to data center system impacts.
2. The fragmented codes adoption landscape complicates comparisons of data center performance across jurisdictions and risks free ridership.
3. The pace of code development cannot keep up with the pace of innovation in the data center industry.

Building energy codes and standards originally intended to promote energy efficiency were not written to address the explosive electricity demand growth associated with data centers, or to mitigate impacts on grid reliability and resilience. They fall short in capturing measurable reductions in energy demand that can result from integrating hardware efficiency, operational controls, and load management strategies.

The challenges of large and dynamic data center loads make it necessary to move beyond efficiency. Modern data center standards must encourage operational flexibility that (1) helps make data center energy consumption responsive to grid conditions and (2) improves alignment of electricity demand with the availability of clean energy.

Furthermore, modern data center standards will need to recognize inherent performance tradeoffs. For example, advances in energy efficiency may increase water consumption, or flexibility measures may affect operational reliability. The most beneficial data center operational outcomes will likely vary by geography and local resource constraints. As a result, standards should support site-specific optimization.

Make performance easier to compare. Because sustainability rating and certification systems for data centers rely primarily on existing energy codes and standards, especially ASHRAE Standard 90.4, the strengths and limitations of the standards carry over to the rating and certification systems. Some certification systems establish prerequisites, while most allow projects to pursue sustainability improvements across multiple categories of their choosing. As a result, projects can achieve similar certification levels through varied combinations of sustainability measures—making it difficult, if not impossible, to conduct accurate comparisons across data center facilities.

Policymakers need to set not only consistent baselines—which in turn will establish necessary guardrails for all data center development—but also consistent performance targets that incentivize leaders to deliver the positive outcomes that matter most to their regions and communities. Using a consistent framework for both minimum requirements and optimal performance targets can create a clear pathway for continuous improvement over time.

Build on what already is working. Recent efforts, including the Green Grid's Data Center Resource Effectiveness metric, the iMasons Climate Accord Maturity Model, and the PNNL/ASHRAE/NEMA Data Center Energy Performance Framework all demonstrate movement away from single-purpose energy metrics toward integrated and systematic approaches to measuring data center performance. These approaches incorporate considerations such as the intersection of energy, water, and climate impacts; look at carbon emissions not just from facility energy demand but also from grid power consumed and the materials incorporated into data centers; and aim to assess efficient performance across the entire life cycle of data center development and operation. The next step is to integrate these promising new developments into a consistent framework that connects comprehensive metrics, comparable baselines, minimum requirements, high-performance targets, and measurement and verification methods that are readily accessible to policymakers, utilities, data center operators and investors, and communities.

Use existing policy mechanisms governing different aspects of data center development—they stand ready to carry stronger performance expectations forward. A data center evaluation framework must establish "do no harm" baseline standards while enabling states and jurisdictions to provide higher incentives for improved performance—and both tasks should be done through pre-existing regulatory pathways, such as building performance standards, interconnection agreements, energy procurement requirements, land-use permitting, and disclosure and reporting requirements. A coordinated framework will also be essential for overcoming key gaps in the policymaking landscape, including the difficulty of comparing across programs that use inconsistent performance metrics, limited operational visibility and fragmented oversight approaches, and limited opportunities to update performance expectations for existing facilities.

Recognize that rapid technological advancement enables data center designers and operators to meet strong performance expectations. Standards that do not support integrative technological approaches risk driving system performance to overemphasize a narrow range of dimensions; they could miss key environmental, grid, economic, or climate impacts and opportunities. Forward-looking, performance-based standards can help ensure that efficiency, reliability, and sustainability goals remain relevant even as the underlying technologies evolve. Server efficiency and utilization, cooling and electrical system efficiency, waste heat reuse, and load management strategies can all be integrated into high-performing data centers, with intelligent control systems supporting alignment between data center operational decisions and relevant grid and climatic conditions. Data center demand flexibility, demonstrated in test environments, promises pathways to mitigate grid stress while maintaining strong business models for data center operators.

Align the market around better outcomes. A framework leading to stronger and more comprehensive data center performance standards requires reinforcement across the market. Pressure from data center investors, customers, and regulators can push larger facilities to initially require stronger performance and to set precedents. Building these expectations into policy decisions and requirements shapes project outcomes broadly. Finally, accountability through verified reporting and public transparency builds community trust and supports the ongoing evolution of standards. Market and policy mechanisms should operate in a complementary fashion, resulting in the coordinated growth of data centers toward performance that is facility-optimal, grid-optimal, and community-optimal.

How to Use This Report

This report is divided into **four** sections.

Part I: Existing Data Center Energy Standards, Rating Systems and Evaluation Mechanisms

Part II: Existing and Emerging Policy Mechanisms and How They Can Integrate Data Center Performance Requirements

Part III: Data Center Technologies

Part IV: Key Stakeholders and Market Drivers

Readers interested equally in current approaches to data center evaluation, relevant policies and technologies, market drivers, and a proposed path forward will benefit from reading the entire report. Recognizing the length of the report, we also encourage readers to select the sections that will be most useful for their needs and decision-making. To that end:

Read **section 1** for an overview of standards, rating systems, and evaluation mechanisms currently applied to data centers. This section explains the applications and limitations of each option. For example, while ASHRAE 90.4 standards provide a solid baseline for data center energy design, their focus is efficiency in mechanical and electrical systems and excludes other factors related to data center energy use, including demand flexibility, water use, and on-site pollution that impacts communities and the environment. ENERGY STAR® focuses solely on benchmarking operational efficiency of mechanical systems. Rating and certification systems, such as LEED and Green Globes, are designed as holistic sustainability frameworks that allow projects to prioritize across multiple categories, but as a result are not direct indicators of data center energy performance.

Readers who prefer a summary of our metrics section can skip directly to **sections 1.5.2** and **1.5.3.** These sections explain that future data center performance evaluation approaches can and should build upon metrics that are (1) already familiar to the market and (2) supported by practical measurement and reporting methods while expanding the evaluation scope where important gaps exist (e.g., demand flexibility, carbon emissions, and computational output metrics).

Readers can use **section 2** to learn about existing policy pathways for regulating and incentivizing data center development. These regulatory pathways are already developed and ready to take on stronger and more specific performance requirements and targets, but they do not yet function in a coordinated way. For a summary of important policy gaps and opportunities, see **sections 2.5** and **2.6.** We recommend policymakers require baseline “do no harm” standards across the relevant dimensions of data center energy performance and environmental impact while also encouraging states and jurisdictions to further incentivize improved or market-leading performance.

Readers interested in data center technologies can refer to **section 3.** Rapid technological advances in computing and data center system design allow data centers to achieve resource-efficient and flexible performance that meets critical needs while limiting environmental impacts. A performance evaluation framework focused on outcomes can accelerate the adoption of best-in-class emerging technologies and reward market leaders who assume early adoption risks. Grid stress, reliability, and resilience are increasingly top-of-mind for regulators, and future-ready data centers must be ready to respond to this concern. For a discussion of those specific issues, see **sections 3.4** and **3.5.**

Readers most interested in key stakeholders and market drivers should refer to **section 4.**

The data center market is a complex ecosystem of organizations. Shared definitions, metrics, and verification methods would allow customers, utilities, governments, and communities to evaluate projects through a consistent framework rather than relying on claims from specific companies or developers, aiding market adoption of the most efficient and best performing systems and designs. For a summary of NBI's recommendations for how to achieve this, see **section 4.10** and the conclusion.

Finally, it is critical to accept the rapid pace of change in this industry. A static report cannot claim to be comprehensive or exhaustive. Given the high profile of data center development as a bipartisan political issue, new legislative and regulatory developments occur almost daily. The information in this report is current as of publication in September 2026.

Part I: Existing Data Center Energy Standards, Rating Systems, and Evaluation Mechanisms

Traditional energy standards and metrics are insufficient for AI data centers. They cannot keep up with the industry's rapid pace of change.

The data center industry, driven largely by the rapid surge in AI adoption, is putting massive and unprecedented demand on the electric grid. Advances in chip efficiency, the emergence of high-density AI-optimized server architectures, and highly efficient cooling and external heat exchange systems are reshaping facility design and operations. These technological shifts have introduced a new set of challenges and opportunities—straining the grid while offering pathways to strengthen it.

Energy codes and standards such as those developed by ASHRAE and the International Energy Conservation Code (IECC) have long been used to evaluate the energy performance of buildings, but the schedules on which they are updated preclude them from keeping pace with rapid changes in computing and data center technology. While ASHRAE data center standards are updated roughly every three years, data center growth rates have explosively increased over the last decade: from 2018 to 2023, data center annual energy consumption roughly doubled (from 76 TWh to 176 TWh), but from 2023 to 2028, projections for energy consumption growth range from 325 TWh to 580 TWh—doubling demand again even at the low end of this projected range [1]. Accelerated server energy demand, associated primarily with AI data center facilities, accounts for almost half of the net increase in global data center energy demand this decade, fundamentally changing not just the amount of energy consumed by data centers but the size, scale, and operational profile of the data centers driving the observed growth [2].

Rapid innovation is creating significant opportunities for data center developers and operators to reduce their own operational costs by improving the efficiency and grid-stabilizing performance of their facilities. Examples include advanced external heat exchange systems that use substantially less water than their predecessors, and innovative cooling systems with pumped refrigerants that improve thermal management in the information technology (IT) space. As power availability becomes an increasingly critical constraint, leading operators are highly motivated to maximize the performance and efficiency of their facilities.

However, strong business incentives for data centers to improve performance do not eliminate the need for effective codes and standards to establish a common baseline of accountability. In addition, traditional energy-efficiency standards were not designed to address the challenges associated with unprecedented growth in electricity demand from data centers. Society needs guidelines that can help manage data centers' impact on grid system reliability and resilience.

States and communities hosting the data centers being built today and those being planned need reliable ways to assess the implications of these projects—whether a proposed or operating facility represents minimum acceptable performance or industry-leading practice, or whether gains in one area are being achieved at the expense of the grid, the environment, and community benefits. To support informed decision-making, experts need to develop—and policymakers need to implement—effective codes, standards, and performance targets tailored to specific requirements and impacts of data centers.

This section surveys existing data center performance metrics, building energy codes and standards, green building certification systems, and relevant industrial efforts and commitments. We will discuss their connections, applications, limitations, and areas requiring improvement.

1.1 Data Center Performance Metrics: Energy, Emissions, Water

A growing number of data center metrics have been developed to assess different dimensions of performance, including facility energy efficiency, carbon emissions, renewable energy use, water consumption, energy reuse, and demand flexibility. To date, Power Usage Effectiveness (PUE) remains the most widely adopted and referenced metric, in part because it is simple to calculate and relies on data that many operators already track: total facility energy use and IT equipment energy use. By indicating the proportion of non-IT energy (lighting, power systems, cooling, etc.) relative to computing energy (IT equipment), it provides a straightforward indicator of infrastructure efficiency. Other metrics require more granular data that is not yet consistently collected or reported as standard industry practice. PUE has helped drive meaningful efficiency improvements over the past two decades, with average values declining substantially from roughly 2.0 in 2007 to around 1.5–1.6 today, with a theoretical minimum of 1 possible, and available technology that enables data centers to achieve PUEs well below 1.5, especially in newer AI and hyperscale data centers. However, many of the easiest infrastructure efficiency gains have already been captured and broader uptake of more efficient system designs across the industry will require more targeted strategies and more consistent measures of data center impacts.

No single metric is adequate on its own. PUE does not capture total load growth, emissions intensity, grid stress, or whether a data center can shift demand in response to grid conditions. It also does not capture the dramatic increase in power demand for information and communications technology (ICT) systems to manage and store data for AI training and inference. Another metric, Cooling Efficiency Ratio (CER) is a complement or alternative to PUE. Unlike PUE, CER indicates the efficiency just of the data center’s cooling system, including both IT space and external heat rejection. It is included as a metric under ISO 30134-7. Carbon Usage Effectiveness (CUE) is meant to address emissions but can be reduced through market tools like renewable energy certificates or offsets that do not improve local grid conditions or community emissions outcomes. Renewable Energy Factor (REF), developed by the International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC) [3], quantifies the renewable energy that a data center generates on-site, acquires through renewable energy certificates, or procures through a portion of utility electricity designated as renewable. Similar to CUE, REF does not account for efficiency and allows purchasing of renewable energy certificates that may not be location-specific and thus do not translate to local emission reduction. Two metrics—Energy Reuse Effectiveness (ERE) and Energy Reuse Factor (ERF)—were also introduced to account for beneficial reuse of waste heat, allowing facilities that export energy to district heating networks or on-site reuse applications to reflect those gains in reported efficiency [4].

Water Usage Effectiveness (WUE), introduced by The Green Grid in 2011, measures how efficiently a data center uses water. It typically ranges from 0 to about 2.5 L/kWh, with an industry average of roughly 1.8 L/kWh [5]. WUE helps quantify water consumption but does not distinguish between potable and non-potable water sources; it does not capture local water stress or fully reflect the tradeoffs between water use and energy efficiency. As of October 2025, The Green Grid has introduced a new water metric, the Water Usage Impact metric (WUI). WUI takes into account the water stress level in the data center's location. A data center in a drought-prone area would have a higher WUI than an equivalent data center in an area with more abundant water resources. There is a need for greater specificity in water usage metrics, to distinguish overall consumption from flow rates, and to break down water usage by sub-systems.

Even with these limitations, existing metrics provide an important foundation for developing codes, standards, rating systems, incentives, and procurement requirements. What is needed now is a more comprehensive framework that uses these metrics together—rather than in isolation—to evaluate facility performance, grid impacts, emissions, water use, and community outcomes. Just as importantly, the framework must address how the necessary data can be collected, verified, and reported in practical ways so that metrics can move from conceptual tools to actionable requirements for policymakers, utilities, communities, and data center operators.

1.2 Energy Codes & Standards Traditionally Applied to Data Centers

State and local jurisdictions typically rely on model energy codes to establish minimum efficiency requirements for buildings. For commercial buildings, the two primary reference documents are the IECC, developed by the International Code Council (ICC), and ANSI/ASHRAE/IES Standard 90.1 developed by ASHRAE. Jurisdictions may adopt the IECC commercial provisions directly, adopt ASHRAE 90.1 directly, or adopt a state-specific commercial energy code that references or modifies one or both. In practice, this means the IECC and ASHRAE 90.1 are not entirely separate systems: the IECC commercial chapter has long allowed ASHRAE 90.1 as a compliance path, while ASHRAE 90.1 serves as the technical backbone for much of commercial building energy regulation.

Data centers complicate this structure because they are commercial buildings, but they do not behave like typical offices, warehouses, or retail facilities. Their energy use is dominated by IT equipment, supporting electrical infrastructure, and continuous cooling loads. Even before the rise of graphics processing units (GPUs) and hyperscale AI campuses, a typical enterprise or colocation data center had an IT density of about 100 watts per square foot (W/ft^2) [6], far exceeding conventional building types (1–2 W/ft^2 for office buildings) [7], [8]. These unusually high and continuous energy demands make data centers a distinct building category that requires more specialized treatment than standard commercial energy code provisions can provide.

1.2.1 Data Centers are Difficult to Classify Under the Widely Used International Building Code

Data centers are not well defined in existing codes, which compounds the challenge of setting standards for their energy and environmental performance. The rapid pace of change in the data center industry outpaces slower code development cycles.

Currently, the International Building Code (IBC) (2024 edition) defines data centers in Section 202 but does not classify them under a specific occupancy group in Chapter 3. When a building type is not explicitly listed, the IBC requires it to be categorized under the occupancy it most closely resembles. As a result, code officials must classify data centers based on their use, hazards, and risk profile. In practice, data centers are often classified under multiple groups depending on their characteristics, including Group B (Business), Group F-1 (Moderate-Hazard Factory Industrial), or Group S-1 (Moderate-Hazard Storage). During the 2025 code development cycle for the IBC and the International Fire Code (IFC), several proposals were introduced to provide clearer direction for data center classification. For example, Proposal G38-25, which is currently pending future action at public comment hearings, aims to formally assign data centers to Group F-1.

However, the ICC acknowledges that data centers differ significantly from conventional buildings in both function and user profile. To better address these differences, the ICC is developing a dedicated Data Center Guideline (G12), which aims to clarify existing compliance pathways, identify gaps in current codes, and recommend best practices for this rapidly evolving building type [9].

1.2.2 ASHRAE 90.4 is the Current Go-To Standard for Baseline Requirements in Cooling and Electrical Infrastructure

ASHRAE 90.4 is intended to provide a data center-specific compliance path for the mechanical and electrical infrastructure that serves IT equipment. ASHRAE 90.1 has referenced ASHRAE 90.4 as an alternative compliance path for large computer rooms and data centers. States that adopt ASHRAE 90.1 as their commercial building energy code must also comply with its companion standard—ASHRAE 90.4, the Energy Standard for Data Centers, first published in September 2016. Both standards are iterated over a three-year continuous maintenance cycle by the consensus committees. Note that OpenAI, maker of ChatGPT, was founded in December 2015, and a demo of ChatGPT debuted on November 30, 2022 [10]. Since then, agentic AI has proliferated (e.g., Anthropic's Claude, Microsoft's CoPilot, Google's Gemini, Meta AI, etc.). ASHRAE 90.4's three-year update cycle operates at a much slower pace than the rate of change in AI and data center technologies. However, the solution is not simply to bypass the consensus-based standards development process in pursuit of greater speed. Rather, standards must be designed to anticipate technology trends and incorporate a flexible framework that can accommodate innovation and evolving performance characteristics without requiring fundamental revisions every cycle.

The latest edition of ASHRAE 90.4, released in 2025, used a revised definition of a data center as a facility with an Information Technology Equipment (ITE) load greater than 10kw and a power density exceeding 20 W/ft² (215 W/m²) of conditioned floor area. For mechanical efficiency, the standard's annualized Mechanical Load Component (MLC) uses ITE design power split into two categories: ≤300 kW and >300 kW. Similarly, the standard's annualized Electrical Load Component (ELC) divides ITE load into two categories: <100kW and >100kW. While sufficient for enterprise or colocation facilities, this two-tier threshold may fall short of capturing the energy use dynamics of larger data centers, including hyperscale data centers, which typically range from 100 MW of demand to over 1,000 MW (1 GW) (as operated by Meta, AWS, and similar). Energy performance requirements are calculated based on modeled performance under historical weather data (Typical Meteorological Year, TMY3) and not measured based on actual conditions, which can vary substantially based on the climate of a data center's location. Energy performance requirements also do not consider other location and climate-dependent factors like water scarcity, and interactions among cooling energy systems, energy generation, energy use, and water intensity. Additionally, ASHRAE 90.4 has an alternative compliance pathway whereby a building can get credit for on-site renewable energy production (up to 5%), and trade-off performance between MLC and ELC [11].

Existing data center standards largely treat internal loads as separate components, overlooking the fact that performance is driven by how these systems operate together [12]. As load flexibility becomes increasingly critical in addressing unprecedented rates of load growth, traditional approaches to HVAC design may no longer be sufficient.

ASHRAE 90.4 only focuses on the infrastructure that serves the IT equipment and not the IT equipment itself, since energy codes do not regulate the performance of IT equipment or other plug and process loads. This means the most energy-intensive component of these facilities falls entirely outside the scope of the standard intended to govern them. Server efficiency has continued to improve over the last decade, but the growth of AI energy demand has overshadowed efficiency savings within facilities as well as the rate at which the grid can build additional carbon-free energy sources to accommodate increased energy demand [13]. This stands in contrast to data center growth in the 2010s, when server efficiency improvements kept pace with computing demand growth [14]. While metrics that measure server utilization and IT work per unit of energy remain useful for benchmarking performance within facilities over time, the wide variation across facilities in size and workload type makes simple comparisons across data centers difficult [15].

In highly energy-intensive facilities, there is typically a strong incentive to select efficient equipment, as energy costs make up a significant share of operating expenses, but additional measures are required to mitigate the grid impacts of growing demand for computational power. The inpouring of capital investment in AI and the race to deploy it may supersede traditional operational cost considerations that drive business decisions toward efficient equipment, ultimately creating a split incentive: developers and operators in a rush to deploy data centers may prioritize technologies and equipment that are readily available rather than those that optimize long-term operational performance and cost effectiveness. As a result, a strong business case for efficiency does not automatically translate into more efficient design decisions or equipment choices. This dynamic underscores the importance of energy codes and standards in establishing baseline performance requirements—but not prescribing specific technologies—and providing guardrails against inefficient outcomes.

1.2.3 California Title 24

California Title 24, Part 6 (Building Energy Efficiency Standards) is California's statewide energy code. It establishes mandatory and prescriptive energy-efficiency requirements for residential and nonresidential buildings and is updated on a three-year cycle by the California Energy Commission (CEC). For data centers and computer rooms, defined as spaces with an IT equipment power density greater than 20 W/ft², Title 24 contains dedicated requirements for the systems serving these spaces [16]. Unlike ASHRAE 90.4, which regulates data centers primarily through metrics such as MLC and ELC, Title 24 relies more heavily on prescriptive requirements. Key provisions include integrated air- or water-side economizers, variable-speed fan controls, restrictions on simultaneous heating and cooling, requirements for adiabatic humidification, fault detection and diagnostics, and energy metering and monitoring. Title 24 also provides a performance compliance pathway by evaluating whether a proposed building's modeled energy use is less than or equal to that of a code-compliant reference building. This allows tradeoffs among measures. California is currently considering additional data center requirements related to economizers, heat recovery, and fan controls as part of future Title 24 code cycles [17].

California's utilities recognized that conventional commercial building energy models were insufficient for evaluating data centers. In response, they developed data center-specific baseline methodologies for utility incentive programs, beginning with the 2013 Energy Efficiency Baselines for Data Centers [18] and later expanding the approach through the 2016 PG&E Data Center Baseline and Measurement and Verification Guidelines [19]. These documents established tailored methods for evaluating data center energy performance beyond what traditional building codes and compliance models could capture. The PG&E baseline methodology references Title 24 equipment requirements and code baselines as part of its approach for determining incentive eligibility and savings calculations. However, the utility baselines serve a different purpose than Title 24's compliance framework. Title 24 compares a proposed design against a code-defined reference building to determine compliance, while utility baselines are intended to estimate incremental savings and identify measures that merit financial incentives.

The PG&E guidelines explicitly acknowledge that standard practice in data centers is a moving target and that baseline assumptions must evolve as technologies mature and become commonplace. Utility baselines can, in principle, be updated more frequently to reflect evolving technologies and prevailing market practices. However, the PG&E guidelines have not been updated since the 2016 version was published.

1.2.4 Challenges and Opportunities in Data Center Energy Codes

Current data center energy codes present three major challenges.

Challenge 1: Energy codes focus primarily on the efficiency of mechanical and electrical systems while excluding other key factors related to data center energy use.

Energy codes leave out IT equipment efficiency such as servers, storage devices, and networking equipment; they do not address compute efficiency, including factors such as server utilization, workload management, and other operational practices that significantly influence overall energy performance. They generally do not account for water consumption, despite the important tradeoffs that often exist between energy and water use in cooling system design. This siloed approach constrains optimization and obscures true system-level efficiency. Ideally, standards would move toward an integrated framework that treats the data center as a unified, dynamic system—one that can be optimized across compute, cooling, power, and resource management simultaneously.

The limitations noted above do not mean that energy codes are irrelevant or unimportant. Codes are intended to establish minimum efficiency requirements; they set the floor, not the high targets, for building design. The key question, therefore, is not whether energy codes should drive fully optimized outcomes, but how far they should evolve to better reflect today's operational realities and load growth challenges associated with data centers.

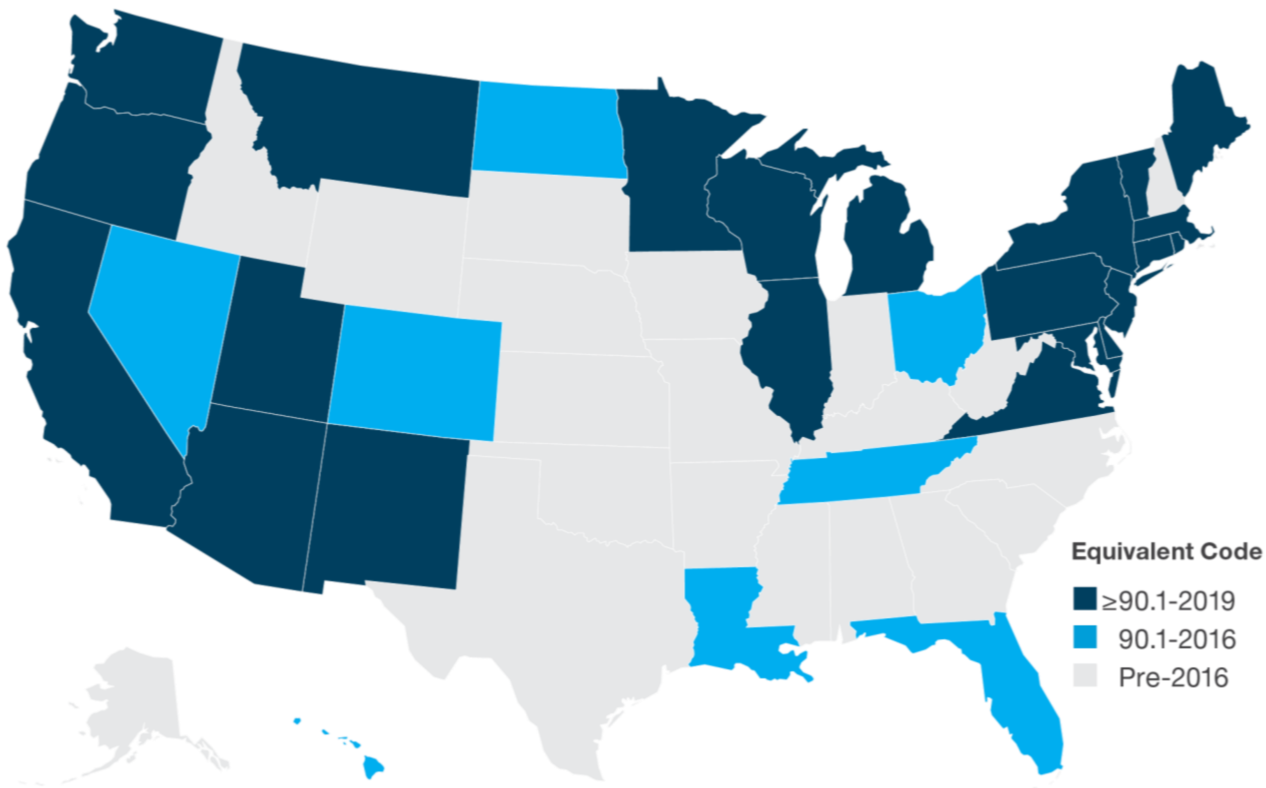
Data centers are increasingly significant loads on electric grids, raising concerns about grid reliability, peak demand, infrastructure investment, and community impacts. In this context, a narrow focus on annual energy reductions is likely insufficient. From a system perspective, *when* energy is used can be as important as *how much* energy is used. Energy codes should therefore consider pathways to encourage load flexibility, demand management, and other performance attributes that support grid reliability, while allowing appropriate tradeoffs among design strategies.

At the same time, energy codes may not be the most appropriate mechanism for regulating every aspect of data center performance, such as IT operations or water use. Rather than attempting to incorporate all requirements directly, codes could serve as the foundation of a broader performance framework by incorporating well-defined references to complementary standards that address these issues. Such an approach would help align requirements across related standards and reduce the risk of fragmented policy development. When policy is fragmented, designers and operators need to navigate multiple standards and compliance pathways, potentially increasing complexity, risks, and costs. **A coordinated framework that clearly defines the role of each standard is likely more practical than a single data center standard that covers all aspects.**

Challenge 2: The fragmented code adoption landscape complicates comparisons of data center performance across jurisdictions and risks free ridership.

The 2021 IECC incorporated ASHRAE 90.4 by requiring data center systems to comply with Sections 6 and 8, but it replaced ASHRAE 90.4's MLC values with IECC-specific design and annualized MLC tables (see Table 1. Maximum Annualized MLC Values Comparison for a Select Band of Climate Zones). By contrast, the 2024 IECC relies more directly on ASHRAE 90.4-2022 for data center mechanical and electrical compliance while reorganizing the treatment of data center systems within the broader commercial HVAC framework. As a result, the applicable efficiency baseline can vary not only by whether a jurisdiction adopts the IECC or ASHRAE 90.1, but also by which edition of the IECC and ASHRAE 90.4 is in force (Table 1). Figure 1 illustrates the current code adoption landscape across states.

FIGURE 1. STATUS OF STATE ENERGY CODE ADOPTION FOR COMMERCIAL BUILDINGS [20]



View detailed description of Figure 1.

TABLE 1. MAXIMUM ANNUALIZED MLC VALUES COMPARISON FOR A SELECT BAND OF CLIMATE ZONES

(Adapted from Standard 90.4-2016, Standard 90.4-2019, Standard 90.4-2022, and IECC 2021)

Climate Zone	Standard 90.4-2016	Standard 90.4-2019 and 2022	IECC 2021
4B	0.35	0.14	0.18
5B	0.33	0.14	0.16
6B	0.34	0.14	0.17

For most building types, variation in energy code adoption is not a major concern because real estate development is inherently local. Developers rarely choose to build offices, schools, or retail buildings in a different state solely because energy code requirements are less stringent. Data centers, however, are different. Developers and operators routinely evaluate locations nationwide based on access to low-cost electricity, transmission capacity, tax incentives, and fiber connectivity. At the same time, states and utilities actively compete to attract data center investment through economic development incentives, tax benefits, and favorable utility policies.

This dynamic creates the potential for free ridership and inconsistent program outcomes. Data centers are often built using highly standardized design templates, meaning that a facility constructed in a state with less stringent energy requirements may be nearly identical to one built in a more stringent jurisdiction. Yet if utility incentive programs calculate savings relative to local code baselines, operators in weaker-code jurisdictions may be able to claim substantial efficiency savings for measures that simply reflect common industry practice elsewhere. Given the scale of modern hyperscale facilities, the resulting incentive payments—and the associated costs borne by utility ratepayers—can be significant.

California's coordinated development of both Title 24 code baselines and utility program baselines offers a potential approach to addressing the challenges of (1) rapidly evolving technologies and design practices that outpace traditional code cycles (see Challenge 3 below) and (2) energy codes only covering some aspects of data center energy performance (see Challenge 1 above). This approach recognizes that building energy codes and utility efficiency programs serve different purposes and therefore may require different reference points for evaluating performance. Energy codes are designed to establish minimum acceptable performance, while utility programs seek to identify and reward performance that exceeds standard practice. At the same time, the two baseline frameworks must be developed in coordination. Misalignment between code baselines and utility program baselines can create unnecessary complexity for project teams, increase compliance burdens, and produce inconsistent signals regarding what constitutes high performance. California's experience demonstrates the value of maintaining separate but aligned frameworks that support both code compliance and incentive program evaluation while minimizing conflicts and duplication of effort.

More broadly, state and utility policymakers should consider developing data center-specific performance benchmarks that are comparable across jurisdictions and updated more frequently than energy codes. Such benchmarks could complement, rather than replace, code baselines by providing a more consistent basis for evaluating performance in a rapidly evolving industry. This approach could reduce free ridership, improve the consistency and effectiveness of incentive programs, and ensure that incentives reward meaningful advances in energy and system performance rather than differences in code adoption schedules.

Challenge 3: The pace of code development often struggles to keep up with the pace of innovation in the data center industry.

New computing architectures, cooling technologies, power systems, and operational practices can emerge and scale within a year, while energy code development, adoption, and implementation cycles often take considerably longer. As a result, many industry stakeholders view codes and standards as either irrelevant to current practice or, in some cases, as barriers to innovation.

While energy codes serve an important purpose and should continue to be developed through a rigorous consensus-based process, there may be value in exploring more agile mechanisms for addressing rapidly evolving technologies in the data center sector. Data centers vary in size, energy demands, power sources, climate, and local water stress. A one-size-fits-all (or "two-size" category in ASHRAE) approach is inadequate, particularly for the demands and impact of modern AI data centers. While it is unrealistic to expect codes to anticipate future technology, a more practical solution is to design performance-based code frameworks that are flexible and adaptable to accommodate innovation while maintaining accountability for outcomes. **This could include creating additional metrics to evaluate system-level performance, allowing tradeoffs between efficiency and flexibility**[1] **when appropriate, and creating clear connections to complementary standards or guidelines that can evolve more rapidly.**

1.3 Data Center Rating and Certification Systems

Many technology companies pursue voluntary green building certifications to demonstrate their environmental leadership, fulfil their corporate sustainability commitments, and respond to customers' environmental values. These certifications are intended to (1) reward best practices that go beyond minimum regulatory requirements, (2) in some cases qualify the company for state incentives by meeting voluntary certification standards,[2] and/or (3) facilitate community acceptance of data centers being built nearby. Although some of these standards have been adapted to data centers, they were not created specifically for data centers and their distinct requirements and impacts.

Rating and certification systems have historically served as market leadership tools to distinguish projects that were willing to practice sustainable design and operations; over time they built strong reputational value in commercial real estate markets. Their credibility in the data center sector is less established. This raises an important question: what should communities expect and require from data centers? Data centers are being built in a range of different states and communities across the United States. Some will care most about tax benefits and jobs, while others will care more about emissions. All likely care about impacts on quality of life. A certified data center may be more efficient or better documented than an uncertified one, but certification does not automatically address the full range of concerns that lead communities to oppose new facilities. While voluntary rating systems have value in some contexts, an important part of our framework is that it allows states and communities to set their own priorities once a "do no harm" baseline is met. Time will tell whether existing rating and certification systems can provide enough assurance in this rapidly changing sector. This section

---

[1] In this context, we use "flexibility" to refer to curtailing or shifting load in response to grid signals or scheduled events. Although there have been several encouraging pilots of load shifting, much work remains to be done to prove its viability as a consistent operational paradigm.

[2] For example, in Pennsylvania's Governor's Responsible Infrastructure Development Standards program, data center access to the permitting fast track is reliant in part on demonstrating adherence to voluntary standards.

examines what these systems currently include, what they do well, where they fall short, and how they could be improved to better reflect the real-world impacts of data center development.

The most recognized energy or green building rating and certification systems for data centers in the U.S. are ENERGY STAR, Leadership in Energy and Environmental Design (LEED), Green Globes, and Building Research Establishment Environmental Assessment Methodology (BREEAM). There are also several other rating, certification, and reporting systems in development or recently introduced. For example, the real estate- and infrastructure-focused GRESB Data Center Assessment was launched in 2026. Also in 2026, the U.S. Energy Information Administration data center energy use pilot was deployed to characterize energy usage in the industry and inform future data collection efforts.

This report focuses specifically on the energy-related criteria incorporated into these systems. In general, these systems do not establish new energy performance standards. Instead, they reference or build upon existing standards and codes such as ASHRAE and IECC and therefore inherit many of the limitations of current data center energy performance frameworks discussed above. Except for ENERGY STAR, these systems evaluate multiple dimensions of sustainability and allow tradeoffs across categories to some degree. As a result, a data center with relatively weak energy performance[3] may still achieve a high certification level by earning credits in other areas, such as sustainable materials. While this holistic approach can support broader sustainability goals, it can also make it difficult for policymakers, utilities, or other stakeholders to use certification status alone as a reliable indicator of superior energy performance.

1.3.1 ENERGY STAR®

ENERGY STAR certification provides a nationally recognized benchmark for energy performance, requiring facilities to achieve a score of 75 or higher on a 1–100 scale. ENERGY STAR relies on a statistical peer-comparison model. While this method is effective for broad benchmarking, its accuracy is constrained by the quality and relevance of the underlying reference dataset, which may not fully reflect rapidly evolving data center architectures and AI-driven workloads. Its data center statistical model, last updated in 2018, was based on PUE values of 60 data centers (ranging from 1.36 to 3.60) and has not fully kept pace with the rapid evolution of hyperscale and high-density data centers, which can limit its relevance for newer facility types [21].

The ENERGY STAR framework for data centers relies primarily on whole-building utility data, which limits visibility of system-level performance. Because on-site energy generation is incorporated into reported energy consumption, facilities may achieve strong ENERGY STAR performance scores even when underlying operational inefficiencies remain unchanged. In effect, energy generated on-site—whether from behind-the-meter renewable resources or combustion-based generation—can offset reported energy use and partially mask inefficiencies in facility operations.

[3] LEED requires projects to achieve minimum prerequisites in multiple areas, limiting excessive tradeoffs. According to the U.S. Green Building Council, data centers with LEED certification in the U.S. often achieve a LEED Gold rating, which requires them to achieve high scores on energy credits.

ENERGY STAR focuses exclusively on energy performance and does not account for water use, an increasingly important consideration in data center operations where tradeoffs often exist between water and energy consumption. The scoring approach also has limited sensitivity to geographic and climatic differences. Facilities located in regions with natural advantages, such as access to water-efficient cooling resources or favorable ambient conditions, may achieve better performance scores than otherwise comparable facilities operating under less favorable conditions. As a result, cross-regional comparisons may not always reflect differences in operational efficiency alone.

Finally, PUE—though simple and widely adopted—captures only infrastructure efficiency and does not capture IT equipment efficiency. This limitation is becoming increasingly significant in compute-intensive AI environments, where server utilization, accelerator efficiency, and workload characteristics can have a substantial impact on overall performance. Together, these limitations suggest a need to modernize benchmarking frameworks so they better reflect contemporary data center operations and more accurately distinguish among infrastructure efficiency, computational productivity, and broader sustainability outcomes.

1.3.2 Leadership in Energy and Environmental Design

The LEED rating system, developed by the U.S. Green Building Council (USGBC), is a voluntary, market-driven, consensus-based tool that serves as both a guideline and assessment mechanism for green buildings.[4] LEED evaluates sustainability across a broad range of categories, including energy, water, materials, site selection, indoor environmental quality, and innovation [22]. Many jurisdictions set local ordinances that require certain classes of commercial buildings to be LEED certified.

LEED evaluates data center energy performance primarily through whole-building energy modeling and compliance pathways based on ASHRAE 90.1 and ASHRAE 90.4. All projects must satisfy minimum energy performance prerequisites, with additional points available for projects that exceed those baseline requirements.[5] LEED further includes data center-specific modeling provisions and methods for quantifying electrical distribution losses through the LEED Data Center Calculator and the ASHRAE 90.4 ELC. LEED v5, released in 2026, references the most current provisions in ASHRAE 90.4-2022 for electrical and mechanical systems serving data centers.

---

[4] Certification is awarded at four levels—LEED Certified, LEED Silver, LEED Gold, and LEED Platinum. The certification evaluates measurable actions spanning site selection, water and energy efficiency, materials procurement, and indoor environmental quality, with additional points available through the Project Priorities credit for innovation, exemplary performance, regional priorities, and pilot credits. Under LEED Building Design and Construction (BD+C) for new construction, the process begins at the design phase, and the certificate is awarded after construction following third-party review by Green Business Certification, Inc.; the BD+C certificate does not expire.

[5] Under LEED v4 for Data Centers, projects must demonstrate at least a 5% improvement in modeled energy performance relative to the ASHRAE 90.1 baseline, with at least 2% of the savings attributable to data center power and cooling infrastructure. LEED v4.1 and LEED v5 incorporate updated provisions from ASHRAE 90.1 and ASHRAE 90.4, including requirements for evaluating infrastructure performance at both full-load and part-load conditions.

LEED provides mechanisms to recognize the efficiency of IT equipment. Projects may earn credits for the use of ENERGY STAR-certified server equipment and for reductions in plug and process loads. These credits are weighted the same as for other “process-intensive” buildings, defined as those with at least 0.5W/ft^2 of process load intensity. In hyperscale and high-density data centers, IT equipment typically accounts for most of the total electricity use yet is minimally addressed in LEED’s energy performance scope. USGBC recognizes this limitation and is working on developing more suitable approaches.[6]

These provisions intend to recognize that IT equipment, electrical infrastructure, and cooling systems operate as an integrated system: improvements in server efficiency reduce cooling and power requirements, while improvements in infrastructure efficiency reduce the overhead energy required to support computing loads. However, ASHRAE 90.1 Appendix G, an energy simulation methodology underlying LEED’s primary performance pathway, was not originally developed to compare the performance of compute-intensive facilities. Unlike traditional commercial buildings, data centers are designed to optimize—rather than reduce—energy supply to IT equipment. This complicates the assessment of data centers through a green building lens, especially because the Appendix G methodology evaluates building energy performance by comparing a proposed design against a standardized baseline building. This approach captures the efficiency of regulated loads such as lighting, HVAC systems, and building envelope components, which are not the primary energy consumers in data centers. These challenges are compounded by the lack of understanding of energy modeling among the many stakeholders involved in data center development and operation.

LEED combines mandatory prerequisites with a flexible points-based certification structure. All projects must satisfy minimum requirements related to energy efficiency, commissioning, refrigerant management, metering and reporting, and water efficiency. Beyond these prerequisites, projects accumulate points across multiple sustainability categories. This flexibility is one of LEED’s strengths as a comprehensive sustainability framework because it allows project teams to pursue environmental improvements that are most relevant to their circumstances. However, it also means that certification level alone is not necessarily a direct indicator of relative energy performance. Projects can achieve similar certification levels through different combinations of credits, and strong performance in one sustainability category can partially compensate for weaker performance in another. Therefore, LEED certification should be interpreted as an indicator of overall sustainability achievement rather than a standalone measure of data center’s energy impact.

1.3.3 Green Globes

Green Globes originated in the early 2000s, derived from the U.K.-based BREEAM and later adapted for use in North America. It was designed as a general commercial building certification

[6] LEED v5 recognizes data centers as a distinct building type. Data center-specific strategies developed in earlier LEED-certified projects have been reviewed and validated by USGBC’s Technical Advisory Group (TAG) and incorporated into the online library under the Project Priorities credit category. In addition, project teams may use pilot credits to evaluate new strategies; once reviewed and approved by the USGBC TAG, these approaches can be incorporated into the Project Priorities category in future LEED updates.

system, rather than a framework tailored to energy-intensive digital infrastructure. As a result, data centers were historically treated as just another commercial building type within the standard 1,000-point scoring system,[7] where energy accounts for roughly one-third of total points [23]. Green Globes is derived from an American National Standards Institute (ANSI) consensus process, and each ANSI standard is updated on a two-year rolling basis.

Similar to LEED, Green Globes relies on external energy standards to define its energy performance criteria but offers a broader and more flexible set of compliance pathways. Rather than prescribing a single methodology, Green Globes allows projects to demonstrate energy performance through multiple parallel pathways. These include ENERGY STAR benchmarking (e.g., Target Finder), which compares building energy use against a national dataset; whole-building energy simulation using ASHRAE 90.1 Appendix G, with a compliance interpretation for the use of ASHRAE 90.4 for data center projects; performance relative to the IECC; and ASHRAE Building EQ, which evaluates operational or design-stage energy performance using standardized benchmarking metrics. More recent updates (2024) also recognize alternative pathways associated with net-zero energy or carbon certifications [23].

While this flexibility can accommodate a wide range of project types and regulatory contexts, it also introduces significant variability. Because each pathway relies on different baselines, modeling assumptions, and performance indicators, energy outcomes are not directly comparable across projects. This structure effectively allows project teams to select the compliance pathway that best aligns with their project constraints or maximizes scoring potential, rather than enforcing a consistent or data center-specific standard of energy performance.

A notable feature of Green Globes is the ability to designate certain criteria as “Not Applicable” (N/A) when they fall outside the project scope or are superseded by local code requirements. These determinations are made by a Green Globes Assessor as part of the third-party review process and are intended to ensure that projects are evaluated only against criteria relevant to their design and operation. For data centers, this flexibility can help tailor the assessment to the unique characteristics of the facility by removing measures that may not be applicable, such as criteria related to optional building features or uses not present within the project.

At the same time, criteria designated as N/A are removed from the total available points used to calculate the final certification score. This approach helps avoid penalizing projects for features outside their scope, but it can also make comparisons across projects more challenging because certification percentages may be based on different sets of applicable criteria. As a result, higher certification scores do not necessarily indicate better overall environmental performance when compared with projects evaluated against a broader set of requirements. For

---

[7] The Green Globes 1000-point scale allows for weighted criteria, wherein the assigned number of points for individual criteria reflects their relative impact and/or benefit on the sustainability of the building. Certification is awarded at four levels, based on a percentage score of total applicable points: One Green Globes (35-54%), Two Green Globes (55-69%), Three Green Globes (70-84%), and Four Green Globes (85-100%). For Green Globes for New Construction, two reviews occur during the assessment process. The first is the Preliminary Assessment (Design Review) and the second, the Final Assessment, occurs at substantial completion and includes a walkthrough of the project.

policymakers, this means that certification levels may provide useful evidence of sustainability achievement but may be less suitable as standalone metrics for comparing performance across projects or establishing uniform policy thresholds.

Only recently has the Green Building Initiative (GBI) begun repositioning Green Globes toward digital infrastructure. In 2025, it launched its first Data Center Campus Certification, designed specifically for hyperscale environments with shared infrastructure such as centralized cooling systems, power distribution, and redundant power architecture. This represents a clear shift toward a framework that is explicitly designed to account for the unique needs and sustainability challenges of data center campuses.

While this evolution may better align the certification with the operational realities of hyperscale facilities, it remains uncertain whether the updated energy credits will meaningfully raise performance expectations or primarily make certification more accessible. Will the new framework drive broader adoption of best practices in data center energy efficiency, or simply lower structural barriers to certification without materially changing underlying energy performance requirements?

1.3.4 Building Research Establishment Environmental Assessment Methodology

Similar to LEED, BREEAM is a globally recognized third-party certification program; it was originally developed by BRE Global and deployed in Europe.[8] Energy performance is evaluated against the applicable local building regulations as the baseline. Unlike LEED's point system, BREEAM uses its own proprietary metric—the Energy Performance Ratio (EPR) for New Construction, which measures performance improvement across three weighted parameters: operational energy demand, primary energy consumption, and CO_2 emissions.

The BREEAM Data Centers pilot scheme, introduced in 2020, is still in a pilot phase and has not been formally adopted in the United States [24]. In addition, its EPR is typically assessed against local energy codes, which in many U.S. applications ultimately align with ASHRAE 90.4. The evaluation of building performance against the applicable local building regulation means that the baseline could be lower or higher depending on where the data center is located. This variable baseline limits BREEAM's effectiveness for comparing performance across regions.

Overall, BREEAM is primarily developed within the U.K./European Union regulatory and planning context, and its assessment structure reflects European building practices and energy modeling conventions. In the United States, building energy codes are largely based on ASHRAE standards and state-level adaptations, which align more directly with LEED's compliance pathways and documentation structure. In addition, many U.S. incentive programs, procurement frameworks, and benchmarking systems are built around LEED and ENERGY STAR rather than BREEAM. As a result, despite its global applicability, BREEAM has seen

[8] BREEAM is applicable to both new construction and existing buildings. For new construction, assessment begins during the design phase, and the certificate is awarded post-construction following a licensed BREEAM Assessor's verification.

limited adoption in the United States due to the dominance of LEED, stronger institutional integration of U.S.-based standards, and lower market recognition within domestic real estate and regulatory systems. Current unfamiliarity with this system in the United States provides an additional barrier to its use for near-term data center performance benchmarking.

1.3.5 Challenges and Opportunities for Existing Data Center Rating and Certification Systems for Energy Performance and Sustainability

This report does not attempt to determine which rating or certification system is more rigorous. Rather, its purpose is to help policymakers understand the applicability, strengths, and limitations of existing frameworks when considering them for permitting, incentives, procurement, benchmarking, or other policy purposes.

Challenge 1: Certification systems are constrained by the underlying energy standards they reference.

Current sustainability certification systems applicable to data centers do not establish independent energy performance standards. Instead, they largely rely on existing industry standards, particularly ASHRAE 90.4 and related ASHRAE methodologies, to define energy performance requirements in the United States. As certification systems are increasingly being incorporated into permitting processes, incentive programs, and other policy mechanisms, the underlying standards they reference become highly influential. Consequently, the strengths and limitations of those standards are often inherited by the certification systems themselves.

Current data center energy metrics primarily focus on facility infrastructure efficiency. While these metrics provide important information, they do not fully capture issues that are becoming increasingly important to policymakers, including grid responsiveness, load flexibility, emissions impacts, and the timing of electricity consumption. Because renewable energy availability, grid stress, and marginal emissions vary significantly by location and time of day, metrics that treat all kilowatt-hours equally may not fully reflect the real-world impacts of data center operations.

Future energy frameworks could better account for regional and temporal differences in energy system conditions, allowing policymakers to evaluate data centers not only based on how efficiently they consume energy but also on how their operations affect broader grid and decarbonization goals. If the objective is to drive meaningful improvements in energy efficiency, emission reduction, and grid-responsive operations, stronger and more comprehensive underlying energy standards may be needed. Certification systems can then serve as effective implementation and market-transformation tools built upon those standards.

Challenge 2: Certification frameworks must continue evolving alongside data center operations.

Data centers differ from most commercial buildings because energy consumption is closely tied to revenue-generating activity. Additional computing capacity can increase service delivery and business value, creating a natural incentive for operators to improve efficiency. As a result, evaluating data centers solely through traditional building-efficiency metrics may overlook important aspects of performance.

A growing challenge for the industry is understanding how to assess the relationship between energy consumption and useful computational output. While measuring computational productivity remains difficult, it is becoming increasingly important as AI workloads grow and improvements in chip efficiency, server utilization, and facility infrastructure become more interconnected.

At the same time, data centers vary considerably in ownership models, development approaches, operational responsibilities, and customer arrangements. Future certification frameworks may benefit from better alignment with these differing business models and operational realities. Recognizing these differences could improve the ability of certification systems to assess impacts and identify meaningful opportunities for improvement.

Certification programs play an important role in the market by encouraging leadership, recognizing best practices, and accelerating adoption of emerging sustainability strategies. However, the data center market is evolving rapidly, and all major certification systems are actively updating their approaches to address high-density computing, AI workloads, changing energy infrastructure, and growing water-use concerns. For this reason, evaluating certification systems solely based on their published requirements may not provide a complete picture of their effectiveness. Additional experience and research are needed to assess how these frameworks perform in practice, including whether they consistently lead to measurable improvements in energy efficiency, water stewardship, emission reduction, and other sustainability outcomes. Real-world implementation, verification processes, and operational results are likely to be as important as the technical requirements themselves.

Challenge 3: Policymakers should understand what certification systems are designed to measure.

Most certification systems are designed as holistic sustainability frameworks rather than dedicated energy performance standards. They establish prerequisite requirements while also allowing flexibility for projects to pursue sustainability improvements across multiple categories, including energy, water, materials, transportation, site development, and innovation.

In some jurisdictions, certification may be voluntary, while in others it may be incorporated into permitting requirements, procurement policies, or development agreements. Regardless of how certification is used, policymakers should recognize that these frameworks are intended to balance multiple environmental objectives rather than optimize a single outcome.

This flexibility is one of the strengths of certification systems because it allows projects to address sustainability opportunities that are most relevant to their circumstances. However, it also means that certification levels should not necessarily be interpreted as direct indicators of energy performance. Projects with similar certification levels may achieve those results through different combinations of sustainability measures.

Overall, certification systems can provide valuable evidence of overall sustainability achievement, but they may be less effective as standalone tools for comparing data center energy performance or establishing uniform energy-efficiency thresholds across projects. Policymakers seeking to drive specific energy outcomes may therefore benefit from pairing certification requirements with additional performance metrics or more targeted energy standards.

Looking forward, the most significant opportunity may not be the creation of additional certification systems but rather the continued development of more comprehensive and relevant data center energy performance metrics. Certification frameworks have proven effective at encouraging market adoption and recognizing leadership. As underlying standards evolve to better address AI workloads, computational productivity, grid impacts, emissions, water use, and system-level efficiency, certification programs can help accelerate their implementation at scale. In this context, certification systems should be viewed not as substitutes for strong energy standards, but as mechanisms that help translate those standards into market practice.

1.4 ISO Standards and Other Evaluation Frameworks

1.4.1 ISO 14001 (Environmental Management) & ISO 50001:2018 (Energy Management)

Hyperscale data center operators are increasingly adopting ISO 14001 and ISO 50001, which focus on an entire organization's environmental and energy management rather than benchmarking individual buildings against others. Both use a Plan-Do-Check-Act continuous-improvement structure. ISO 14001 establishes a broad environmental management system [25] beyond energy, including criteria air pollutants and greenhouse gas emissions, water use, wastewater, disposal, recycling, toxics, fluids management, and other site-specific environmental risks. In practice, ISO 14001 provides the umbrella environmental management framework, while ISO 50001 provides the more targeted energy management system that should be integrated into that broader environmental approach. Organizations that already comply with ISO 14001 often find it easier to integrate ISO 50001 [26]. Participating sites have demonstrated sustained energy performance gains, often achieving approximately 4% annual energy savings over extended periods. Adoption of ISO standards remains voluntary and is generally driven by corporate sustainability commitments and internal design, construction, and operation standards.

Unlike prescriptive rating systems, ISO 50001 does not mandate specific metrics; instead, organizations are required to define their own energy performance indicators based on operational context. This flexibility allows data center operators to tailor performance tracking to their unique workloads, infrastructure, and operational priorities [27]. For data centers, the most effective energy performance indicator would be a metric like "work delivered per unit of energy consumed." In practice, most hyperscale data center operators continue to rely on metrics such as PUE and WUE—some of the first and most established metrics in the industry. Given that hyperscale companies manage large, globally distributed portfolios of facilities, PUE and WUE, although not ideal, offer a reasonable starting point to enable consistent benchmarking and performance tracking over time and internationally.

Compliance with ISO 50001 requires third-party certification and emphasizes ongoing performance improvement rather than one-time achievement. Supporting this approach, the 50001 Ready Navigator [27], developed by the U.S. Department of Energy, helps organizations implement ISO 50001-aligned practices [27].

ISO 50001 can support stronger internal energy management because it requires organizations to define energy performance indicators, establish baselines, and track improvement over time. However, because organizations have flexibility to define their own indicators, ISO 50001 does not by itself create a consistent basis for comparing performance across data center operators. Widely used metrics such as PUE and WUE can provide a common starting point, but they are not required under ISO 50001 and do not capture the full range of data center impacts, including computational efficiency, load growth, emissions, water stress, and grid effects. ISO 50001's value for data centers could be strengthened by incorporating additional, data center-specific indicators that go beyond facility efficiency, such as metrics for demand flexibility and the associated effects on energy costs, emissions, and grid impacts.

1.4.2 Carbon-Free Electricity Framework and 100% Renewable Energy

Hyperscale data centers are starting to go beyond baseline compliance standards and local regulatory requirements—many want to position themselves as sustainability leaders in a highly competitive industry and meet their corporate environmental, social, and governance requirements.

For example, to meet its goal of 24/7 carbon-free energy, Google tracks hourly electricity consumption and matches it with carbon-free sources in real time using time-based Energy Attribute Certificates, which require accurate hourly tracking of clean energy generation and trading [28].

Some hyperscale data center operators participate in global initiatives such as RE100, which commits companies to sourcing 100% renewable electricity by 2050 or earlier. However, RE100 is not tailored specifically to data centers and does not guarantee net-zero carbon emissions; participants span industries from apparel to manufacturing to biotech.

Emission reduction approaches based on frameworks such as 24/7 carbon-free energy accounting and RE100 commitments are not necessarily indicative of improved energy performance from a grid operator's perspective. While these approaches can reduce a company's reported emissions footprint, they do not always align with real-time grid conditions, location constraints, or system-level efficiency needs, and the addition of large power purchase agreements aligned with data center demand can also exacerbate grid supply and demand imbalances, unless there are behind-the-meter battery installations that smooth peak renewable power generation impacts. In practice, these approaches may also be difficult to scale across the broader data center industry, as they often require substantial financial resources, long-term procurement capability, and access to diverse clean energy markets—conditions more readily available to large hyperscale operators such as Google than to small or mid-sized developers.

Hyperscale operators' commitments to 100% renewable energy are noteworthy, but they are not a reliable way to meet policy objectives regarding operational energy, emissions performance, and grid impacts from the data center industry. Ambitious voluntary commitments sometimes overlap with local, state, and regional goals, but this patchwork of metrics allows data center operators to pick and choose among a variety of approaches, compromising the potential for a cohesive approach that could better reduce grid stress.

1.4.3. Data Center Resource Effectiveness

Data center performance metrics like PUE, WUE, and CUE were originally developed by The Green Grid (TGG) beginning in 2007 and are now managed through the Information Technology Industry Council and related international standardization efforts. Several of these metrics have subsequently been incorporated by international standard setting bodies, within ISO/IEC resource-efficient data center working groups.

TGG transferred management of these metrics to Working Group 1 (WG 1) of the ISO/IEC JTC 1/SC 39 for further improvements. ISO/IEC JTC 1/SC 39/WG 1 (Resource Efficient Data Centers) is actively developing a suite of international standards aimed at improving the measurement and management of data center energy and resource efficiency [29] [30]. Its ongoing work focuses on standardizing key operational metrics such as PUE and ERE, along with best practices for cooling optimization, power distribution, and server utilization. It also seeks to establish more consistent approaches to reporting and benchmarking energy performance, as well as frameworks for life cycle resource assessment. Unlike building certification systems, this effort emphasizes engineering-grade comparability and operational performance rather than point-based ratings, positioning it as a more technically grounded framework for understanding data center efficiency, which can be referenced by certification systems. This body of work remains a work in progress.

Recognizing the limitations of evaluating data centers through a single metric, TGG released Data Center Resource Effectiveness (DCRE) in 2025 as a more holistic framework for evaluating resource use. DCRE builds on earlier metrics by incorporating multiple dimensions of performance, including facility energy efficiency, water consumption, climate conditions, and local water stress. The framework also introduces WUI, which seeks to account not only for how much water is consumed, but also the relative water stress of the region in which a facility operates. Rather than serving as a single engineering metric, DCRE is structured as a multi-factor scoring and maturity framework intended to provide a broader view of data center resource performance.

DCRE remains a relatively new framework and has not yet achieved the level of industry adoption or recognition associated with PUE. Additional work is also underway within industry organizations to develop complementary metrics that better address IT work capacity, AI workloads, and other emerging aspects of data center performance. As these efforts mature, DCRE may provide a useful foundation for moving beyond single-attribute metrics, but additional testing and real-world implementation will be needed to determine its effectiveness as a benchmarking, certification, or policy tool. Given its early stage of adoption, it should be viewed as a promising emerging framework rather than a mature standard for regulatory or incentive programs.

1.4.4 Maturity Model

The iMasons Climate Accord (iCA) Maturity Model, released in 2025, is designed as a voluntary roadmap for measuring progress toward emission reduction in digital infrastructure. The model focuses on three areas that the iCA identifies as major contributors to life cycle carbon emissions: power, equipment, and materials. The maturity model was developed with the recognition that data center sustainability extends beyond operational energy efficiency. While metrics such as PUE have helped improve infrastructure efficiency, they provide limited visibility into embodied carbon impacts from equipment and construction materials or the carbon intensity of electricity used to power facilities. The iCA Maturity Model seeks to address these gaps by defining progressive maturity levels and best practices that organizations can use to evaluate and improve their decarbonization efforts over time.

A key characteristic of the maturity model is that it functions as a self-assessment and improvement tool rather than a certification program. Unlike standards or certification systems, there is no formal compliance determination or third-party approval requirement associated with maturity levels. Therefore, the maturity model is not a substitute for performance standards or regulatory requirements. Because it is voluntary and designed as a roadmap rather than a compliance framework, it does not establish minimum performance thresholds or provide a consistent basis for comparing facilities. Instead, its primary value lies in helping organizations understand their relative level of maturity and identify strategies for continuous improvement.

The maturity model places substantial emphasis on embodied carbon in construction materials and equipment supply chains, areas that are often outside the scope of traditional data center energy metrics and green building certification systems. This broader perspective aligns with growing industry efforts to improve transparency and reduce emissions across the entire data center value chain rather than focusing exclusively on operational energy use. This framework provides useful insights into emerging best practices and could help inform future policy discussions regarding life cycle emissions and data center decarbonization.

1.4.5. PNNL/ASHRAE/NEMA AI Data Center Energy Performance Framework

Another recent development is the AI Data Center Energy Performance Framework developed jointly by Pacific Northwest National Laboratory (PNNL), ASHRAE, and the National Electrical Manufacturers Association (NEMA). Unlike traditional data center energy standards, which primarily focus on minimum equipment or system efficiency requirements, this framework is intended as a comprehensive guide for the planning, design, commissioning, operation, retrofit, and modernization of AI data centers.

The framework addresses a broad range of topics, including planning and siting, integrated design, energy and thermal efficiency, grid-interactive operations, resiliency, commissioning, operations and maintenance, and retrofit strategies. It also explicitly considers both energy and water use and promotes demand flexibility and grid-responsive operation. The framework is structured as an online resource that can evolve as technologies and best practices change, recognizing that AI-related infrastructure is advancing more rapidly than traditional code development cycles.

Since the framework is not a code, standard, or certification system, it does not establish mandatory performance requirements, prescribe compliance thresholds, or provide a mechanism for benchmarking and comparing facilities. Instead, it serves as a best practice guide that directs users to relevant standards, technical resources, and design approaches. While it represents a significant step toward a more holistic view of data center performance—one that incorporates energy efficiency, water use, reliability, and grid impacts—it remains a voluntary guidance document rather than a formal performance standard.

1.5 Bridging the Floor and the Frontier: The Need for an Integrated Data Center Performance Evaluation Framework for System Outcomes

1.5.1 Effective Metrics Remain the Foundation

Since PUE was introduced more than 20 years ago, numerous additional metrics have been developed to evaluate different aspects of data center performance. These metrics vary in purpose, scope, and granularity, measuring topics such as infrastructure efficiency, water use, carbon emissions, and renewable energy procurement. Together, they have significantly improved the industry's ability to understand, benchmark, and manage energy performance.

Table 2 summarizes several of the most widely used metrics and their primary limitations.

TABLE 2. DESCRIPTION OF METRICS AND LIMITATIONS

Metric	Description	Function and Limitations
Power Use Effectiveness (PUE)	Measures the efficiency of IT loads relative to total building load (inclusive of mechanical equipment, cooling, etc.) per kWh of IT power consumed over the course of a year. Measures historical performance, equivalent to metered annual Energy Use Intensity.	PUE indicates efficient building design and operation of the core mechanical and electrical components outside of IT systems. It does not by default indicate the degree to which a building negatively contributes to local grid stress or other energy impacts like emissions or water intensity. A building that reduces load based on demand signals to accommodate grid needs may demonstrate a higher (less efficient) PUE. This is a drawback of PUE. Additionally, PUE depends, in part, on cooling equipment, and may be correlated with higher WUE factors (as higher energy-efficiency cooling systems may use more water).
Mechanical Load Component (MLC)	Projected annualized factor that measures use of mechanical loads relative to total building loads (inclusive of IT components).	Relies on modeled estimates and historic climate data inputs, which can deviate from actual operational conditions of data center. Does not factor in water use or carbon intensity of system.
Electrical Loss Component (ELC)	Captures the electrical loss of uninterruptable power supply and IT equipment.	ELC pertains to the efficient design of electrical systems and components to limit electrical loss.
Cooling Efficiency Ratio (CER)	Compares the amount of heat removed to the energy consumed by cooling system operations.	Cooling consumes a large portion of data center energy. This metric captures the efficiency of cooling systems in the context of a specific data center's operation.
Carbon Use Effectiveness (CUE)	Measures a building's carbon intensity, per kWh consumed.	Does not necessarily indicate the degree to which actual energy drawn from the grid is served by renewable energy at any given point in time. CUE can be improved through the use of Renewable Energy Certificates, offsets, and non-time-based procurement. CUE does not consider efficiency of the energy used on site, and a building could be very resource inefficient in operation but still have a good CUE.
Renewable Energy Factor (REF)	Measures a building's renewable energy generation mix, per kWh consumed.	REF does not necessarily indicate the degree to which actual energy drawn from the grid is served by renewable energy at any given point in time. Can be driven down by RECs, offsets, and non-time-based procurement. REF does not consider efficiency of the energy used on site, and a building could be very resource inefficient in operation but still have a good REF.
Water Use Effectiveness (WUE)	Measures a building's water use intensity, per kWh of IT power consumed.	Can be inversely correlated with PUE, due to higher efficiency of liquid cooling. Also sometimes broken out into WUE_{site} and WUE_{source}, which respectively measure water use intensity within the data center itself, and water use intensity inclusive of data center off-site generation and transmission needs.

As data center technologies and operating practices continue to evolve—including the growth of AI workloads, liquid cooling systems, demand-flexible operations, and waste heat recovery—existing metrics alone may not provide a complete picture of facility performance. Most were developed to address a specific component or resource stream and therefore do not fully capture whole-building and system-level performance, time-dependent resource consumption, interactions with the electric grid, or tradeoffs among energy, water, carbon, and reliability objectives.

Effective metrics remain the foundation of any performance-based policy, standard, or certification program. Future approaches should build upon metrics that are already familiar to the market and supported by practical measurement and reporting methods, while expanding the scope of performance evaluation where important gaps exist. Potential additions may include demand-flexibility metrics, such as verified peak-load reduction, response time, duration, availability, and workload-shifting capability; time- and location-based emission metrics that reflect hourly grid carbon intensity and alignment with clean energy availability; water-impact metrics that distinguish total water consumption from local water stress; measures of energy reuse and waste-heat utilization; and computational-output metrics such as energy per AI task, energy per inference, energy per training run, useful compute delivered per kilowatt-hour, or other workload-adjusted productivity measures.

Many of these emerging metrics are not yet sufficiently mature, standardized, or practical for universal regulatory application. Nevertheless, they identify important dimensions of performance that are increasingly relevant to policymakers, utilities, and operators. Over time, incorporation of these metrics into standards, incentive programs, tariffs, interconnection agreements, and reporting requirements could help distinguish facilities that achieve broader societal objectives from those that perform well on only a limited set of metrics.

1.5.2 A Robust Energy Standard Must Extend Beyond Design Efficiency

A robust energy standard for data centers should be grounded in efficiency as the primary principle while recognizing that efficiency alone is not sufficient to address the full range of energy system impacts associated with rapidly growing compute demand. Existing standards such as ASHRAE 90.4 provide an important foundation by establishing minimum efficiency requirements for major electrical and mechanical systems. However, achieving broader public policy objectives may require an additional layer of performance accountability focused on how facilities actually operate over time.

At its core, energy standards should prioritize measurable reductions in energy demand through improvements in both hardware and software efficiency. This includes optimization of electrical and mechanical infrastructure, IT equipment, cooling systems, controls, AI models, and workload management. Demand-side efficiency remains the most universally applicable and scalable strategy for reducing resource consumption, lowering operating costs, and minimizing stress on energy infrastructure.

Yet efficiency alone may not adequately address the challenges posed by increasingly large and dynamic data center loads. Energy standards should therefore encourage operational flexibility that enables facilities to better align energy consumption with grid conditions. Potential strategies include demand response, dynamic load management, and workload shifting, where technically and operationally feasible. Such approaches can help reduce peak demand, improve utilization of existing grid infrastructure, and facilitate integration of variable renewable energy resources.

Energy standards should also encourage the coordinated use of renewable energy generation, energy storage, and flexible operations. However, renewable procurement should not be viewed solely as an accounting mechanism that offsets annual energy consumption. The timing of electricity production and consumption increasingly matters, particularly in regions with significant renewable penetration or localized grid constraints. Where practical measurement methods and data are available, future frameworks may place greater emphasis on the temporal alignment of electricity demand and clean energy supply.

Importantly, a comprehensive standard should recognize that optimization in one area can create tradeoffs in another. Improvements in energy efficiency may increase water consumption, flexibility measures may affect operational reliability, and the most beneficial outcomes can vary by geography and local resource constraints. For example, in water-stressed regions, reducing water consumption may be a higher priority than achieving marginal gains in energy efficiency. Explicitly acknowledging such tradeoffs can help ensure that standards encourage system-level optimization rather than narrow improvements in a single metric.

By making these tensions explicit rather than implicit, standards can better reflect real-world constraints in data center design and operation and be responsive to variation in resource constraints by data center geography. In doing so, they can both set enforceable performance boundaries and create space for continuous innovation and system-level optimization, ensuring that improvements in one dimension do not inadvertently undermine performance in another.

1.5.3 Toward a Data Center Energy Performance Evaluation Framework

Taken together, DCRE, the iCA Maturity Model, and the PNNL/ASHRAE/NEMA AI Data Center Energy Performance Framework demonstrate that the industry is beginning to move beyond single-purpose metrics toward more integrated approaches to data center performance. Each addresses a different gap. DCRE expands resource-effectiveness measurement beyond PUE by incorporating energy, water, climate, and regional water stress considerations. The iCA Maturity Model broadens the focus to life cycle carbon emissions associated with power, equipment, and materials. The PNNL/ASHRAE/NEMA framework provides guidance across the full life cycle of data center development and operation, including planning, design, resilience, operations, and grid interaction. While none of these initiatives function as a comprehensive regulatory standard, collectively they point toward the type of system-level approach that will be needed for the AI data center era.

The remaining challenge is translating these emerging concepts into a consistent set of metrics, baselines, and verification methods that can be applied by policymakers, utilities, operators, investors, and communities. The missing layer is a performance evaluation approach that both establishes clear guardrails for minimum performance and creates aspirational targets that encourage market leaders to go further.

To address this need, we propose a **Data Center Energy Performance Evaluation Framework (DC-EPEF).** The DC-EPEF is not intended to replace existing standards, certification systems, or performance metrics, nor does it seek to create an entirely new set of requirements. Instead, it is designed to build upon the substantial body of work already developed across industry, government, utilities, and standards organizations. Its objective is to provide a coherent structure that incorporates proven approaches, aligns ongoing initiatives, and identifies areas where important gaps remain.

The DC-EPEF prioritizes metrics that are already familiar to the market, supported by available data, and practical to verify. Rather than starting from scratch, it leverages existing standards, metrics, and reporting frameworks wherever possible. At the same time, it highlights emerging performance dimensions that may require additional metrics, reporting practices, or performance expectations as technologies and policy priorities evolve. In this way, the DC-EPEF balances continuity with innovation—building on established measures where they are effective while creating a pathway for incorporating new indicators as they mature.

Data center performance measurement is unlikely to be transformed by a single new metric or standard. Instead, progress will come through the integration of existing tools and the gradual development of additional measures where important gaps exist. By connecting established standards, emerging industry frameworks, and evolving policy objectives, the DC-EPEF provides a practical and adaptable structure for evaluating data center performance more holistically and consistently. It is intended to help policymakers and market actors distinguish facilities that achieve efficiency gains in a single dimension from those that deliver broader benefits across energy, emissions, water, grid reliability, and resource stewardship.

Part II: Strengthening Existing and Emerging Data Center Policy Mechanisms

Part I identifies the need for data center performance requirements that extend beyond traditional building efficiency. A stronger framework should account for energy and water use, emissions, grid impacts, load flexibility, and also drive performance improvements over time. Existing standards and metrics provide pieces of this framework but they are not consistently incorporated into the decisions that determine where, how, and under what conditions data centers are built and operated.

The data center policy challenge is not necessarily a lack of mechanisms. Building codes, interconnection agreements, utility tariffs, tax incentives, permitting processes, procurement requirements, and reporting rules already influence data center development. Although most were not originally designed to regulate data center performance at the scale it continues to grow, they provide established decision points where stronger metrics and operating requirements could be added more quickly than through the creation of an entirely new regulatory structure.

This section is not intended to provide an exhaustive inventory of data center policies. Instead, it identifies seven categories of existing and emerging policy mechanisms, provides selected examples of how these mechanisms are evolving, and considers where they can influence the data center development life cycle. Throughout this report, selected examples also illustrate how these structures are already beginning to incorporate stronger requirements.

Table 3 below defines the policy categories discussed in this section and highlights some potential areas of influence that these categories cover. Note that these policy categories can apply to data centers, although many were originally designed for commercial buildings or other types of loads. The examples listed under potential areas of influence illustrate possible applications and scope; they do not imply that all these approaches are currently used or fully implemented in existing policies.

TABLE 3. DEFINITIONS OF POLICY CATEGORIES

Policy Category	Purpose	Potential Areas of Influence
Building Energy Performance	Establish minimum or above code requirements for building systems and equipment	Energy efficiency, cooling, controls, commissioning, heat recovery
Interconnection and Energy Reliability	Determine how and when large loads connect to the grid	Load forecasting, grid capacity, queue management, backup resources, conditional service, reliability
Demand Flexibility and Grid Participation	Encourage or require loads to respond to grid conditions	Peak reduction, demand response, load shifting, grid services
Energy Procurement and Emissions	Influence the sources and emissions associated with energy use	Clean energy procurement/renewables, storage, carbon intensity, emissions
Utility and Tariff Regulation	(Re)allocate cost of energy and infrastructure and manage customer risk	Customer classes, contracts, minimum payments, cost recovery, incentives
Land Use, Permitting, and Environmental Issues	Address local development and community impacts	Water, site, noise, air quality, permitting, zoning, community benefits
Reporting and Verification*	Confirm whether facilities meet projected or required performance	Benchmarking, disclosure, measurement, verification, ongoing reporting

**Reporting and verification are not yet widely adopted as a standalone policy pathway but is emerging as a potential mechanism.*

2.1 Existing Policies as Vehicles for Stronger Requirements

The above categories influence different aspects of data center development, but many rely on similar underlying structures: eligibility requirements, contracts, approval conditions, reporting obligations, and financial incentives. These structures can serve as vehicles for the performance metrics identified in Part I. Some potential examples:

- An interconnection agreement could include energy efficiency and load flexibility requirements.
- A tax incentive could require verified energy, water, or emissions performance.
- A local development agreement or community benefit agreement could establish standards for on-site emissions, battery storage, and waste-heat reuse that benefit local communities.

The value of these mechanisms is that they already affect project decisions and therefore may provide a faster pathway for stronger requirements for data center energy performance than creating entirely new legislation, ordinances, or incentives.

The examples in Table 4 are illustrative rather than comprehensive. Each shows how an existing policy mechanism could carry stronger performance expectations without requiring an entirely new regulatory or incentive structure.

TABLE 4. ILLUSTRATIVE POLICY AND REGULATORY MECHANISMS

Policy Category	Selected Example	What this Example Demonstrates
Building Energy Performance	New York A9136 (proposed): thermal energy network requirements	Construction permitting could be used to require waste-heat feasibility analysis, heat-recovery infrastructure, metering, and verified performance. The proposal remains in the Assembly Energy Committee.
Interconnection and Energy Reliability	Southwest Power Pool High Impact Large Load framework	Grid operators can coordinate load, transmission, and supporting-generation studies and offer conditional access subject to curtailment and reliability requirements.
Demand Flexibility and Grid Participation	Texas SB 6	Large-load policy can distinguish mandatory emergency curtailment from planned and incentivized flexibility and reliability.
Energy Procurement and Emissions	NV Energy Clean Transition Tariff	A utility tariff can allow a large customer to support new clean, firm resources through the regulated utility framework. The approved Google agreement supports 115 MW of enhanced geothermal capacity.
Utility and Tariff Regulation	Oregon POWER Act	A state can establish a separate large-load service classification, require contracts, and allocate the costs and risks of serving large loads to those customers.
Land Use, Permitting, and Environmental Issues	Lancaster, Pennsylvania Community Benefits Agreement	A legally enforceable local agreement can address clean energy, water, noise, air quality, site design, reporting, workforce, and community investment.
Reporting and Verification	New York S6394B (proposed): data center disclosure requirements	Preconstruction and annual reporting could compare projected and actual energy, water, emissions, and waste-heat performance and provide a basis for enforcement.

2.2 How Policy Mechanisms Can Carry Stronger Requirements

The examples above demonstrate that some existing policies already influence data center approval, infrastructure investment, utility service, and operations. A stronger, standardized framework is needed to establish consistent expectations for comprehensive data center energy performance and ensure that baseline “do no harm” standards across all relevant dimensions are in place for all data centers, while making room for states and jurisdictions to further incentivize improved performance. The metrics used within that framework should carry over into the policy mechanisms that already shape data center development, allowing codes, tariffs, interconnection agreements, permits, incentives, and reporting requirements to consistently apply the measures most relevant to their authority and point of influence.

2.2.1 Building Energy Performance

Building codes, equipment standards, commissioning requirements, and construction permits provide the most direct pathways for establishing minimum performance during design and construction. Existing requirements generally focus on mechanical and electrical infrastructure and may not fully address IT equipment, water impacts, grid interaction, or actual operating performance.

New York A9136 provides a potential approach. It would require applicable new data centers and major expansions to assess nearby thermal loads before receiving a construction permit and to design, construct, and operate facilities to capture and reuse waste heat unless a connection is found infeasible [31]. It would also establish heat recovery performance standards, metering, and annual reporting.

The proposal demonstrates how a familiar construction pathway could utilize metrics such as energy reuse, recovered thermal energy, metering, and verified energy performance. The primary limitation of building codes and permits is that they are strongest before occupancy and may not ensure that modeled performance is achieved throughout operation. Thus, this approach should be paired with a reporting and verification requirement.

2.2.2 Interconnection and Energy Reliability

Interconnection is one of the earliest and most consequential decision points in data center development. Utilities and grid operators can use interconnection studies, service agreements, phased energization, and conditional service to address projected load, infrastructure readiness, supporting available or future resources, and system reliability.

Southwest Power Pool's High Impact Large Load framework illustrates this approach. Its Conditional High Impact Large Load service can provide faster conditional access to the grid in exchange for potential curtailment during constrained periods. Its High Impact Large Load Generation Assessment coordinates the study of a large load with the generation intended to support it [32].

These mechanisms could carry requirements for forecasting accuracy, expected load ramp up schedules, backup generation disclosure, controllability, and ongoing monitoring. Their primary gap is the limited scope since interconnection processes typically focus on grid impacts and may not address water use, building efficiency, or broader community impacts.

2.2.3 Demand Flexibility and Grid Participation

Demand response programs, flexible load agreements, virtual power plants, and other incentivized grid programs can encourage data centers to reduce or shift demand when the electric system is constrained. These mechanisms create a pathway to define measurable expectations for peak reduction, response time, duration, availability, and load shifting capability. Load shifting—its potential and its considerable remaining challenges—is discussed in more detail in section 3.4.

Texas SB 6 demonstrates the importance of distinguishing between two types of load reduction. The enacted law establishes emergency authority to curtail covered large loads or direct the deployment of backup generation to protect system reliability. It separately directs the Electric Reliability Council of Texas (ERCOT) to competitively procure a reliability service from large-load customers of at least 75 MW, with advance notice for anticipated emergency conditions [33]. The first is a mandatory safeguard to help with overall grid reliability, while the second is a planned and incentivized form of grid participation.

Part III of this report dives deeper into the technologies and strategies for operation that have the potential to provide flexibility, including workload shifting, facility-system controls, and electrical storage. The policy role is to define the expected performance, establish participation terms, and verify whether the agreed flexibility is available when needed.

2.2.4 Energy Procurement and Emissions

Clean energy tariffs, renewable energy requirements, emission limits, and procurement conditions or availability can influence the energy resources used to serve large loads. These mechanisms can move beyond annual renewable energy use (and avoid confusion with renewable energy credits) by considering carbon intensity, timing, location, storage, and the addition of new clean on-site resources.

The NV Energy Clean Transition Tariff demonstrates how an established utility structure can support customer-specific clean energy procurement in a regulated electricity market. The approved agreement with Google supports 115 MW of enhanced geothermal capacity intended to provide full-time energy for the company’s Nevada operations [34].

Clean energy procurement requirements can be effective when paired with efficiency and flexibility expectations, to ensure that operational performance and grid reliability are also considered. Clean energy purchasing alone does not show whether a facility is efficiently designed or operating, effectively reducing demand during constrained periods, or aligning its electricity use with the availability of clean energy. As discussed in Part I, a stronger performance framework should evaluate energy efficiency, carbon intensity, and load flexibility together rather than allowing strong performance in one area to offset poor performance in another.

2.2.5 Utility and Tariff Regulation

Large-load tariffs, customer classifications, long-term contracts, minimum requirement charges, and financial guarantees are increasingly being used to reduce speculative service requests and protect other customers from the costs of building stranded infrastructure.

Oregon’s POWER Act directs the public utility commission to establish a separate service classification and tariff for qualifying large energy use facilities (using at least 20 MW) that primarily provide data processing, hosting, or related services. The tariffs must allocate the costs and risks of serving these facilities to the large-load customers and limit cost shifting to other ratepayers. Qualifying facilities must also enter into electricity service contracts lasting at least 10 years, identify when service will begin, and commit to minimum payments based on projected electricity use. These contracts may also include excess demand charges [35].

AEP Ohio provides another example. Its approved data center tariff includes a 12-year term, minimum demand charges, financial viability requirements, exit fees, and for certain customers, a financial commitment equal to 50% of total minimum charges over the contract term [36].

These mechanisms currently focus primarily on financial risk, but their contractual structures could also add a requirement to meet operational performance targets. Protecting ratepayers requires more than ensuring data centers pay for their fair share of costs. Performance requirements can help align data center growth with broader public interests by ensuring efficient operation, flexible load management, and cleaner energy use to create more value rather than potential long-term liability for both the grid and the communities in which they operate. Service terms could include higher efficiency thresholds, flexible load commitments, backup resource disclosure, and ongoing performance reporting.

2.2.6 Land Use, Permitting, and Environmental Issues

Local governments generally have limited authority over energy system operations, but they directly influence site selection, zoning, development approval, and community impacts. Conditional use permits, development agreements, and community benefits agreements can address issues that may fall outside building codes or utility proceedings, including water use, air quality, and micro-grids.

Lancaster, Pennsylvania and Chirisa Technology Parks created a legally enforceable community benefits agreement that includes clean energy commitments, limits on water use, closed loop cooling, noise limits, back up generation requirements, waste heat recovery, public reporting, local hiring, and financial contributions to community and sustainability funds.

Local mechanisms are strongest during site selection and development approval. Their main limitation is that local agencies may lack the technical information, expertise, or jurisdiction needed to evaluate actual operational performance, grid impacts, and utility integration.

2.2.7 Reporting and Verification

Reporting and verification connect projected performance with actual operating results. Requirements can cover energy and water use, emissions, peak demand, waste heat, flexibility performance, and compliance with conditions established through permits, tariffs, interconnection agreements, or incentives. Reporting is arguably one of the most important foundational policy pathways because other requirements cannot be evaluated without consistent data. It is an important component of a comprehensive approach because data can help better inform standards development, create more trust in communities, and contribute to a more collaborative and iterative approach when working with data centers to better manage their expansion. It is also one of the largest current gaps because reporting is frequently voluntary, inconsistent across companies and jurisdictions, or disconnected from enforceable requirements.

New York S6394 or the Sustainable Data Centers Act, would require disclosure before construction and annual reporting after operation. Proposed disclosures include projected and actual energy consumption (with required comparison), energy sources, emissions, water use, and waste heat production and reuse. The bill also includes public disclosure and potential penalties for noncompliance [37].

2.3 Incorporating Metrics into Policy Mechanisms

Part I finds that no single metric adequately describes data center performance. PUE remains useful for understanding facility infrastructure efficiency, but it does not capture absolute energy use, emissions, water impacts, grid stress, or the ability to shift demand. A more complete framework requires the incorporation of complementary metrics that address facility performance and broader energy system and environmental impacts.

The seven policy categories provide different vehicles for incorporating those energy or energy-related metrics. The same metric may apply across more than one policy category. For example, projected load may be reviewed during interconnection, incorporated into a utility contract, and later compared with actual demand through annual reporting. This overlap can create continuity across different stages of development. Not every mechanism needs to regulate every impact. Instead, each can apply the metrics most relevant to its authority and decision point.

TABLE 5. POLICY PATHWAYS AND POTENTIAL AVENUES FOR EXPANDED DATA CENTER REQUIREMENTS

Policy Category	Aligned Metrics and Requirements
Building Energy Performance	Mechanical and electrical system efficiency, energy recovery and reuse, IT equipment efficiency, controls, metering, commissioning, verification
Interconnection and Energy Reliability	Projected peak load, load ramp up, forecast verification, non-fossil fuel backup generation, controllability, infrastructure readiness
Demand Flexibility and Grid Participation	Temporal flexibility, spatial flexibility, facility-system flexibility, energy and thermal storage, computational optimization, workload shifting capability, demand response participation
Energy Procurement and Emissions	Carbon intensity, renewable energy, hourly or time-based clean energy matching, battery storage, emissions
Utility and Tariff Regulation	Efficiency thresholds/requirements, load flexibility, forecasting accuracy/requirements, financial commitments, load flexibility, performance monitoring, reporting
Land Use, Permitting, and Environmental Issues	Water use, air quality, heat island effects, micro-grids, water heat reuse
Reporting and Verification	Actual energy and water use, emissions, peak demand and load profiles, projected load growth, environmental impact reporting, third-party verification

2.4 Regulatory Touchpoints Across the Data Center Life Cycle

The policy mechanisms described influence data center development at multiple points. They operate through separate regulatory, utility, and approval processes when a project moves from initial planning through ongoing operation and future expansion. Each touchpoint can apply different elements of a stronger, standardized performance framework, but the requirements will be most effective when they use consistent metrics, definitions, and reporting methods. These touchpoints help reveal what gaps might exist in utilizing existing policy mechanisms and how a larger framework can best align across separate regulatory processes.

2.4.1 Planning and Siting

Mechanisms such as zoning, tax incentives, environmental review, and community pushback shape projects before major development decisions are finalized. These processes may require early disclosure of projected electricity demand, water use, emissions, cooling and other major mechanical systems, backup generation, and potential local impacts.

This is the earliest opportunity to determine whether a proposed site and energy strategy are compatible with available infrastructure, environmental conditions, and community priorities. However, utilities, regulators, and communities may not receive complete project information until after significant siting decisions have already been made.

2.4.2 Interconnection, Design, and Approval

Interconnection studies and requests, utility tariffs, service agreements, building codes, permits, and development agreements determine how a project will receive power and what conditions it must meet before construction or energization.

These mechanisms can establish requirements for:

- Load forecasts and ramp up schedules
- Infrastructure costs and financial commitments
- Building and system efficiency
- Demand flexibility and controllability
- Clean energy procurement and emissions
- Backup resource disclosure
- Water and environmental performance
- Reporting and verification

This is a critical stage because early performance goals can be translated into enforceable project requirements. The primary gap is that utility, building, environmental, and local approvals are often conducted separately and may not apply the same performance expectations. This is one key reason an overarching framework could help create structure for multiple regulatory agencies.

2.4.3 Construction and Commissioning

Inspections, commissioning, incentives, and utility energization provide opportunities to confirm that required equipment, controls, meters, and demand response technologies were installed and commissioned appropriately.

This stage connects design requirements to actual facility capabilities. Commissioning and verification can confirm that systems operate together as intended before the facility begins full operation. However, many approval processes focus primarily on whether equipment was installed rather than whether the completed facility can achieve its projected performance.

2.4.4 Operations

Once a facility is operating, utility tariffs, demand response programs, benchmarking requirements, and reporting can track and compare against actual performance. These mechanisms can measure energy and water use, emissions, peak demand, load profiles, forecast accuracy, and participation in flexibility programs. This remains a significant gap in the current landscape. Reporting is often voluntary or inconsistent, and responsibility may be divided among the data center developer, building owner, operator, and tenants.

2.4.5 Future Expansion or Retrofit

Data center campuses frequently expand in phases, add computing capacity, or replace major cooling, electrical, and IT systems. New utility service requests, permit modifications, tariff changes, and major equipment replacements provide opportunities to update load forecasts, benchmark performance, and apply updated requirements.

Without these triggers, existing facilities and later project phases may continue operating under older standards even when their energy demand or environmental impacts increase substantially.

2.4.6 Enforcement Across the Building Life Cycle

Different entities have authority or influence at each stage of data center development. Figure 2 summarizes the primary points where governments, regulators, utilities, federal agencies, and standards organizations can shape project requirements.

FIGURE 2. WHO SHAPES DATA CENTER REQUIREMENTS ACROSS THE BUILDING LIFE CYCLE

Who Can Regulate	Planning & Siting	Design & Approval	Construction	Operations & Commissioning	Future Expansion or Retrofit
Local Government & Constituents	✓	✓	✓		✓
State Government & PUCs (dependent on state)	✓			Opportunity!	✓
Utility	✓			✓	✓
Federal (EPA and FERC)	✓	Technical Support	Technical Support	Technical Support	Technical Support
Standards & Certification Bodies		✓		✓	✓

View Accessible Table of Figure 2.

2.5 Key Gaps Across the Policy Landscape

Existing policy mechanisms affect nearly every stage of data center development, but they do not yet function as a coordinated approach. Different entities oversee siting, interconnection, design, environmental impacts, utility service, and ongoing operations, and the requirements established at one stage may not carry forward to the next. Addressing this fragmentation will require a more consistent framework that carries common metrics, reporting expectations, and performance requirements across these separate decision points.

The most important gaps to address and consider in evaluation framework development include:

- **Inconsistent metrics:** Codes, utility agreements, permits, and reporting programs may use different definitions and measurement methods.

- **Projected versus actual performance:** Modeled performance and goals set during planning are not always compared with operating results.
- **Fragmented oversight:** Utility, environmental, building, and local reviews are frequently conducted separately.
- **Limited operational visibility:** Behind-the-meter generation, storage, and flexible operations may not be fully visible to utilities, regulators, or communities.
- **Split responsibility:** Developers, owners, operators, and tenants may control different parts of a facility's performance.
- **Limited triggers for existing facilities:** Expansions, retrofits, and ownership changes do not always require compliance with updated performance expectations.

A standardized framework could help address these gaps by establishing common metrics, definitions, and verification methods that can be used across different policy mechanisms. Each authority would apply the requirements relevant to its jurisdiction, while the underlying performance expectations would remain consistent throughout the project life cycle.

2.6 Using Existing Policy Pathways to Advance a Stronger Framework

Existing policy mechanisms provide a starting point for stronger data center requirements. Building codes can establish design minimums. Interconnection agreements and tariffs can address forecasting, flexibility, reliability, and cost responsibility. Procurement requirements can influence energy sources and emissions. Local approvals can address environmental and community impacts. Reporting and verification can determine whether facilities deliver the performance promised during development.

Not every policy mechanism needs to address every aspect of data center performance. Instead, a stronger standardized framework can establish a common set of metrics and expectations, while codes, contracts, tariffs, permits, incentives, and reporting requirements apply the elements most relevant to their authority and point of influence. Such a framework provides a shared language for understanding data center energy performance across multiple dimensions and, importantly, how those dimensions relate to one another. It can help identify where metrics should be complementary and where tradeoffs may be appropriate, recognizing that optimal outcomes will vary based on local environmental conditions, grid characteristics, and resource priorities. Rather than promoting a single objective, the framework supports a more balanced approach to evaluating and managing data center performance.

Some new policies will still be necessary where existing authority is incomplete or there is a gap in oversight, such as in operations in Figure 2. However, incorporating stronger requirements into the mechanisms that already shape data center investment, approval, utility service, and operation may allow meaningful improvements to be implemented more quickly and consistently.

The following section, Part III, examines the technologies and operational strategies already available to help data centers meet these stronger performance expectations. Part IV discusses market adoption, identifying the stakeholders and market dynamics that can accelerate uptake of the framework. Policy can establish the requirements, but rapid change will depend on aligning the incentives, investment decisions, procurement practices, and business priorities that shape how data centers are sited, designed, built, and operated.

Part III: Data Center Technologies

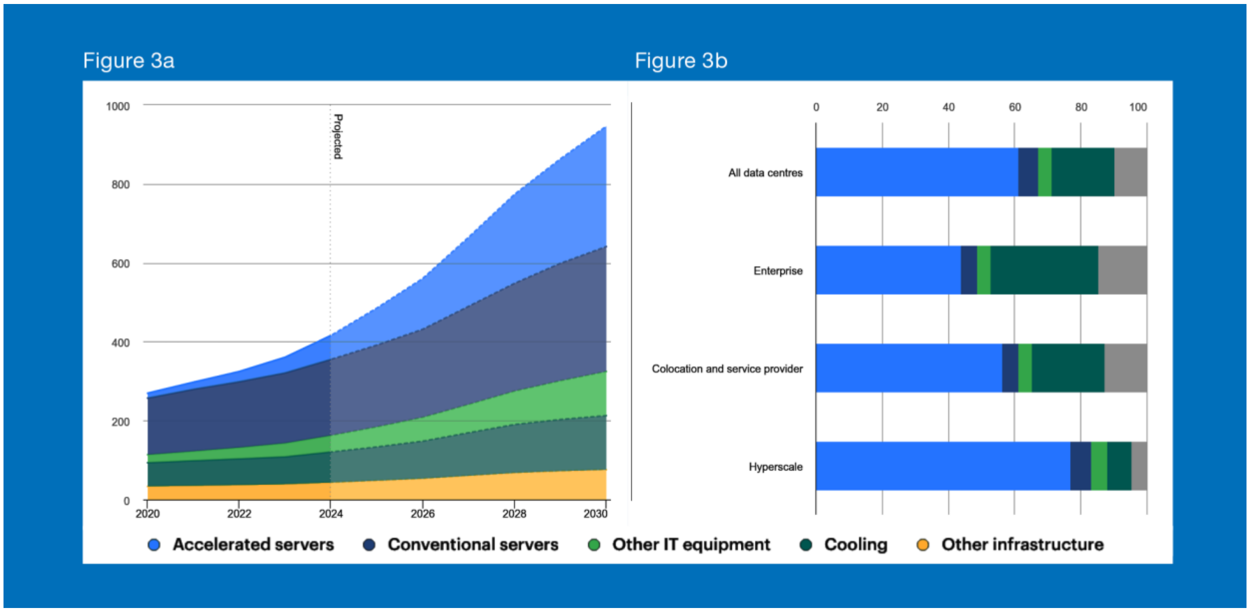
Understanding the current state of technology and anticipating future trends is essential for developing effective standards and codes, particularly in rapidly evolving sectors such as data centers and digital infrastructure. Data centers can achieve resource-efficient and flexible performance that produces valuable work while limiting environmental impacts. This requires integrated, holistic design approaches that incorporate a mix of mechanical, electrical, and technological system efficiency and performance strategies. Policies—including mandatory requirements, voluntary standards, and incentives—which do not support these integrative technological approaches hinder the ability of policymakers and impacted communities to advance responsible data center growth and risk driving system performance to optimize around a narrowed range of dimensions, missing key environmental, grid, economic, or climate impacts.

Additionally, standards that are grounded only in today's technologies risk becoming obsolete as system architectures, workloads, and energy supply models continue to change. Conversely, forward-looking policies can help ensure that efficiency, reliability, and sustainability goals remain relevant even as the underlying technologies evolve. As a result, an effective performance evaluation framework requires not only a clear understanding of existing technical performance but also a structured view of where the industry is heading. This section provides an introductory overview of data center energy use trends and patterns and discusses opportunities to improve data center energy performance.

3.1 Energy Use Trends and Patterns in Different Types of Data Centers

Data center energy usage is driven primarily by cooling and IT system demands (e.g., servers, storage systems, and networking equipment). The proportion of energy use by different data center processes varies by region. A global average of 62% of energy use goes to IT equipment, 20% to cooling, and the remaining energy to other infrastructure and IT uses (e.g., storage and network uses). In the United States, however, a greater proportion of energy use, on average, goes to IT system components: an estimated 57% of energy goes to IT equipment, 13% to storage and network uses, and 30% to infrastructure demands, including cooling [38].

FIGURE 3A. GLOBAL DATA CENTER ELECTRICITY CONSUMPTION IN TWH, BY EQUIPMENT, BASE CASE, 2020-2030, LAST UPDATED 10 APRIL 2025, [39], FIGURE 3B. SHARE OF ELECTRICITY CONSUMPTION (TWH) BY DATA CENTER AND EQUIPMENT TYPE, 2024. LAST UPDATED 10 APRIL 2025, [39]



View detailed description of Figures 3A and 3B.

Servers account for roughly 60% of the total electricity use of data centers [39], although this share varies in individual data centers depending on server configuration, workload type, and data center classification (Figure 3a). Cooling systems represent the second-largest source of electricity consumption in most data centers. Figure 3b illustrates market trends in electricity consumption by data center type. It shows that although hyperscale facilities operate with much higher server counts and server densities, they have succeeded in designing cooling systems that can use as little as approximately 7% of the data center's total electricity, whereas cooling systems for less energy-efficient enterprise data centers may consume up to 30%.

Enterprise (on-premise) data centers are facilities owned and operated entirely by a single organization to host its internal IT infrastructure. The diversity of IT infrastructure needs can reduce utilization rates and overall PUE as a result. These facilities range in size from small deployments of roughly 5,000–20,000 square feet supporting 500–2,000 servers (around 1–5 MW of power demand) to medium-scale installations that can reach 5–20 MW [40]. While they offer high levels of control over security, compliance, and data governance, enterprise data centers tend to be less energy efficient than hyperscale data centers, with PUE typically in the 1.5–1.8 range, although newer facilities are improving toward approximately 1.4 [41].

Enterprise data centers' load profiles remain fundamentally different from hyperscale or colocation facilities. Demand is often tied to business activity, peaking during working hours and falling sharply overnight and on weekends. Utilization of computing capacity by enterprise data centers has historically been low—sometimes averaging only 5–15% at the server level—which has been a key driver of migration away from traditional physical server deployments toward cloud computing, virtualization, and software-defined infrastructure, allowing resources to be provisioned and managed more flexibly through software systems [42]. However, from a grid perspective, enterprise loads offer only moderate flexibility: some demand can be shifted or deferred during off-hours, but core business-hour computing remains closely tied to user activity, limiting the potential for large-scale demand response participation.

Hyperscale data centers are large, centralized computing facilities owned and operated by cloud service providers such as AWS, Microsoft Azure, Google Cloud, Meta, Apple, and Oracle. They are designed primarily to support cloud computing, AI training, and large-scale data processing rather than the needs of a single enterprise. These facilities are typically much larger and more power intensive than other data center types, ranging from over 100 MW to multiple gigawatts for emerging AI “mega-campus” designs. Operationally, hyperscale facilities are characterized by higher overall utilization and near continuous sustained demand. When intermittent changes in demand do occur, they can cause large fluctuations in load with substantial grid impacts, in contrast to the more variable patterns seen in enterprise environments. Overall utilization can reach roughly 40–50% for AI inference workloads and up to 80% for AI training, reflecting much more intensive use of installed compute capacity, based on self-reported industry data to Lawrence Berkeley National Laboratory (LBNL) [43]. These facilities also achieve high efficiency, with companies like Google reporting fleet-wide PUE as low as 1.09, driven by advanced cooling and infrastructure optimization.

While some workloads are temporally flexible (e.g., AI training), real-time inference and streaming workloads require constant availability. This creates a mixed demand profile, where some electricity use can be optimized for time-of-use pricing or renewable availability, while other portions behave as firm, always-on load. There are increasing calls for hyperscalers to begin to actively manage this flexibility, such as scheduling training during periods of high wind generation in Texas or shifting compute across regions to align with cleaner grids, while also participating in demand response programs and testing infrastructure-level flexibility such as HVAC pre-cooling during summer temperature peaks and uninterruptible power supply (UPS)-based grid support [41].

Colocation data centers sit between enterprise-owned facilities and hyperscale cloud platforms, offering a shared infrastructure model where clients lease physical space (power, cooling, and physical security) while owning and managing their own IT equipment (servers, storage, and networking). Essentially, they are not responsible for the building or its infrastructure. These facilities are often located within large multi-tenant campuses and are frequently connected to high-capacity grid infrastructure, such as distribution substations that support dense computing loads. Utilization levels are roughly similar to enterprise data centers, as they represent an aggregation of individual enterprise owners in a single space, ranging typically from 21–35% [43]. Energy efficiency also varies by facility design, with PUE values commonly between 1.3 and 1.6 depending on age and cooling technology [41].

What distinguishes colocation facilities most is the diversity of their tenant base and resulting load behavior. A single site may host hundreds of organizations with very different operational needs, from financial services firms requiring high uptime to media, cloud, or research workloads with more variable demand. This diversity creates a “peak smoothing” effect, where aggregated facility load is more stable than any individual tenant’s profile.

TABLE 6. OPERATIONAL EFFICIENCY AND DEMAND FLEXIBILITY ACROSS DATA CENTER TYPES

Data Center Type	PUE	Utilization	Operational Efficiency
Enterprise (On-Premises)	1.5–1.8 (newer ~1.4)	~11–20% (low; high idle capacity)	Low-moderate (fragmented, underutilized infrastructure)
Colocation	1.3–1.6	~21–35%	Moderate-high (shared infrastructure improves efficiency vs. enterprise)
Hyperscale	~1.09 (fleet best-in-class)	~40–50% (inference), up to ~80% (training)	Very high (optimized at scale, high capital efficiency)

3.2 Efficiency Improvement Opportunities within Data Center Facilities

One of the leading hyperscale operators, Google, reported achieving an annual average PUE of 1.09 in 2024, compared to an industry average of approximately 1.56. Much of this improvement reflects rapid technology advancement in recent years—for example, facilities that use integrated liquid cooling can achieve PUE values near 1.10, compared to more traditional cooling design [44]. The PUE average of 1.56 reflects the full data center standing inventory, including centers that are decades old. Lowering PUE to 1.09 means that 84% less overhead energy is used for non-computing tasks like cooling and power distribution than the average data center. According to the Google, this performance was achieved through a layered or “full stack” approach that optimizes everything from physical infrastructure to chip design and software rather than focusing on a single system or technology.

AI-driven data center growth is fundamentally changing the underlying efficiency profile of computing infrastructure. Traditional servers rely primarily on central processing units (CPUs) for general-purpose workloads and typically consume around 750–1,000 W per unit, with relatively low idle power draw. In contrast, AI-accelerated servers combine CPUs with GPUs, tensor processing units (TPUs), and application-specific integrated circuits (ASICs)[9] to handle large-scale, compute-intensive tasks, but at much higher power levels—often 10 kW or more per server. They also exhibit significantly higher idle power consumption, roughly nine times that of conventional servers, meaning inefficiencies persist even when workloads are not fully active [45].

[9] ASICs are specially designed circuit chips optimized to efficiently complete specific tasks, but less able to be adapted to other, more general uses. TPUs are a type of ASIC developed by Google that are designed for neural network machine learning tasks.

This shift increases both the absolute energy demand and the importance of improving utilization and hardware efficiency in hyperscale environments.

In response, AI-driven data centers are employing various strategies to improve overall system efficiency. Below we highlight four: server efficiency, cooling efficiency, flexible operating temperatures, and software algorithm efficiency.

3.2.1 Server Efficiency

Server efficiency has become a major focus as workloads shift from traditional CPU-based systems to AI-accelerated infrastructure. Companies are increasingly adopting low-power, highly efficient processors like ARM-based chips, which can reduce energy use by up to 60% compared to traditional designs. Companies are also deploying highly specialized AI chips, such as Google's TPUs, which deliver dramatically better performance per watt than earlier generations. In addition, new architectures like chiplet-based processors, faster memory systems (e.g., high bandwidth memory), low-power double-data rate, and "near-memory" designs reduce the energy lost when moving data between components [46] [47]. Overall, IT efficiency gains, especially for AI chips, are leading to much more efficient performance per user query. These IT efficiency gains, however, are currently overwhelmed by growing AI workload demand [48].

A common argument is that server efficiency does little to curb electricity demand because data center operators can simply add more servers and continue growing their load. This may be true at the facility level when speed to market remains the primary driver. From a long-term perspective, however, data centers with more efficient IT equipment, higher server utilization, and more efficient AI algorithms are likely more competitive and financially viable because they can produce more useful computing output with less energy. For states, utilities, and local communities, these are often the types of facilities worth attracting. A data center that delivers greater economic and computing value per unit of energy is more likely to be an asset to the grid and community than one that merely consumes large amounts of power without optimizing its performance.

3.2.2 Cooling System Efficiency

Cooling efficiency is another major area of improvement. Cooling accounts for a substantial portion of the non-IT energy load in data centers, but across a range of system designs, technology advancements now allow for meaningful reductions in overall PUE while also balancing other local impacts, including water use, outdoor space requirements, and noise levels [49]. Efficient cooling systems, regardless of type, incorporate economizers which rely on ambient air temperatures to reduce the need for running mechanical chillers or compressors.

External heat rejection system design is the major driver of water use by data center cooling systems. Many data centers rely on water-cooled systems, which are more efficient than air-cooled systems. Open cooling towers used for heat rejection require continuous input of water as waste heat causes evaporation. System designs are available today to substantially reduce water consumption, in some cases avoiding water use entirely apart from the hottest days of the year. These designs include much more water-efficient evaporative and adiabatic cooling systems, and systems that use precision controls [50]. In water-stressed regions, closed-loop dry cooling technologies can also vastly improve water use efficiency while still delivering highly energy efficient systems [44]. Direct adiabatic cooling is particularly effective in hot, dry climates, where low humidity improves evaporation efficiency, or in centers operating at higher temperatures. Under optimal conditions, this approach can improve cooling efficiency by more than 85% and reduce air temperatures by up to 12°C [51].

Another key cooling efficiency strategy is improving airflow design within the data center itself. A widely used approach is hot aisle/cold aisle containment, where server racks are arranged so that cold air intake and hot air exhaust are physically separated. As AI workloads increase heat density, this kind of airflow control becomes even more important to prevent hotspots and inefficiencies. Studies using computational fluid dynamics modeling show that hot aisle containment combined with raised floor designs can improve air distribution efficiency by 28% and reduce air recirculation by 40% under typical cooling conditions [52] [53].

3.2.3 Operating Temperature Efficiency

Operating at higher temperatures is a third important efficiency lever. Traditionally, data centers were kept relatively cool, usually around 68–72°F (20–22°C), but modern equipment can safely operate at much higher temperatures, up to about 95°F (35°C). Allowing higher operating temperatures reduces the need for more energy-intensive cooling systems, with studies showing that each 1°F increase in temperature can reduce cooling energy use by roughly 4% [54]. For example, Amazon has reduced its water usage substantially in part by raising the temperature threshold for when it runs chillers, holding off until the ambient temperature is above 85°F; raising the temperature threshold reduced water consumption by half, compared to similar facilities running at lower temperatures [55].

3.2.4 Software and Algorithm Design Efficiency

Software and algorithm design efficiency allows data centers to increasingly improve efficiency, for example by using automated cooling control systems. One major innovation is the use of machine learning to control cooling systems in real time. Google, working with DeepMind, developed an AI system that replaces manual or rule-based cooling control with predictive algorithms. The system continuously analyzes data from sensors across the facility, forecasts heat generation, and automatically adjusts equipment such as pump speeds and cooling tower fans. This approach has reportedly reduced cooling energy consumption by up to 40% in Google's data centers [56]. It highlights a broader shift toward treating cooling as a dynamic optimization problem rather than a fixed mechanical process.

Energy-efficient algorithms also reduce the amount of computation required to complete tasks. For example, Google's "AlphaEvolve," an AI coding agent built on large language models, is used to discover and optimize algorithms across Google's infrastructure. The company reports that one AlphaEvolve-driven optimization has recovered roughly 0.7% of global computing resources, effectively allowing more work to be completed without increasing total compute capacity [57]. This highlights how small algorithmic improvements can scale into meaningful systemwide efficiency gains.

Another important trend is the shift toward task-specific AI models. Instead of relying exclusively on large, general-purpose models, companies are increasingly deploying smaller, specialized models designed for narrow applications. These models require significantly less computational power, which reduces both energy use and infrastructure demand while improving efficiency for specific tasks [46]. Alongside this, firms are investing in model and code optimization, refining AI architectures and software systems to improve hardware utilization and reduce unnecessary computation.

3.3 Increasing System Efficiency through Waste Heat Recovery

Data centers are among the most energy-intensive building types, and a large share of that energy is ultimately lost as heat. Rather than treating this heat as a byproduct to be discarded, it can be treated as a usable energy resource. Waste heat can be captured through liquid loops or heat exchangers that transfer thermal energy from server systems into heated water or refrigerants. This recovered heat can then be transported through pipelines to nearby buildings that require heating. This is the basis for thermal energy networks, emerging district-scale systems that connect multiple buildings to shared heating and cooling infrastructure. Importantly, this approach can work across different cooling technologies, including air cooling, water cooling, and advanced liquid cooling, since all systems ultimately generate recoverable heat that can be captured and redirected.

The benefits of waste heat recovery extend to both data center operators and surrounding communities. For operators, closed-loop cooling systems that enable heat reuse can significantly reduce water consumption—by as much as 80% compared to traditional open-loop systems that rely on evaporation. These systems also improve overall resource efficiency by shifting cooling from a waste-disposal process to an energy-recovery process. For communities and utilities, the systemwide benefits can be even larger. If widely deployed, waste heat recovery could reduce the need for up to 54 new power plants nationwide, avoiding an estimated $22.1 billion in infrastructure costs [58]. It can also help ease grid constraints by lowering net electricity demand growth. In some district-scale applications, integrating thermal energy systems can reduce data center power consumption by up to 30% by replacing conventional electric cooling with shared thermal infrastructure.

International experience shows that these systems are already technically feasible at scale. European countries have moved more aggressively than the United States in integrating data center heat into district energy systems, including incorporating this practice into data center standards with energy reuse targets [59]. In Stockholm, Sweden, waste heat from data centers is fed into the city’s district heating network, supplying enough energy to heat approximately 30,000 apartments each year while cutting overall heating system emissions by about 50 grams of CO_2 per kilowatt-hour [60]. Across Europe, dozens of similar projects are in operation or development, including initiatives by Google in Finland, Microsoft in Denmark, and Equinix in France, where recovered heat is used for applications ranging from residential heating to facilities like public swimming pools

Policy momentum to advance thermal energy networks in the US is beginning to emerge, though unevenly across states. In Virginia, recent legislative efforts have attempted to explore and accelerate waste heat reuse, including HB2578, which sought to identify reuse opportunities but did not pass, and HB323, which remains under consideration and directs the state to study pathways for expanding data center heat recovery. By contrast, New York has taken a more structured approach. The state's Utility Thermal Energy Network and Jobs Act, enacted in 2022, authorizes utilities to pilot district-scale thermal systems under regulatory oversight. One example is Con Edison's pilot in Manhattan's Chelsea neighborhood, which plans to transfer waste heat from a nearby data center into a centralized heat pump system serving a New York City Housing Authority residential community.

3.4 Reducing Grid and Climate Impacts through Data Center Energy System Design and Operation

3.4.1 Power Quality and Reliability Requirements for Data Center Operation

Ensuring power quality and reliability is essential to data center operations, as interruptions in power are associated with substantial financial consequences. According to Uptime Institute's *2024 Annual Outage Analysis,* outages are both frequent and increasingly costly. More than half (54%) of facility operators reported that their most recent significant outage exceeded $100,000 in total losses, while 20% experienced outages costing more than $1 million—a share that has risen for four consecutive years. For large enterprise facilities, the average cost of downtime can reach approximately $540,000 per hour, underscoring the extreme sensitivity of digital infrastructure to even short disruptions [61].

Power-related issues remain the leading cause of these incidents, cited by 54% of affected operators. This reflects not only grid instability, but also the growing complexity of integrating diverse energy sources into data center operations. As compute loads increase and uptime expectations approach "five nines" reliability, even minor deviations in voltage, frequency, or supply continuity can cascade into systemwide failures. In this context, power quality is not simply about avoiding blackouts—it is about ensuring consistent, high-fidelity electricity delivery at all times.

These reliability requirements are driving a data center operator preference for firm, dispatchable power sources—generation that can be ramped up or down on demand, independent of weather conditions or time of day. Among available options, natural gas combined-cycle plants remain the dominant near-term solution due to their dispatchability, relatively fast development timelines, and established supply chains, despite their carbon intensity. Nuclear power offers a carbon-free alternative with very high capacity factors (often above 90%), and interest in nuclear-backed power purchase agreements is growing, though high capital costs and long development timelines remain barriers. Geothermal and hydropower also provide firm, low-carbon electricity, but are geographically constrained and, in the case of hydropower, increasingly vulnerable to climate variability.

By contrast, solar and wind generation are inherently intermittent. These characteristics make them difficult to rely on as standalone power sources for facilities that require uninterrupted operation. Battery energy storage systems are often positioned as the bridge between intermittent renewables and firm demand, but their capabilities remain structurally limited. Current battery systems (2–4 hours of discharge duration) cannot economically cover overnight demand, multi-day wind lulls, or prolonged weather events. While long-duration storage technologies (8–24+ hours) are under development, they are not yet commercially viable at the scale, cost, or reliability required by hyperscale data centers.

These dynamics explain why data centers, despite strong clean energy commitments, continue to rely on fossil-based or other firm generation sources. Renewable energy alone cannot meet the stringent power quality and reliability requirements of modern data infrastructure without significant advances in storage or complementary firm capacity. Therefore, the pathway to decarbonized data centers is not simply about adding more renewables—it requires a coordinated system of clean firm generation, long-duration storage, and grid integration strategies that can deliver both sustainability and reliability at scale.

3.4.2 Increasing Data Center Demand Flexibility to Reduce Grid Impacts

A 2025 study from Duke University's Nicholas Center [62] brought attention to data center load flexibility as an opportunity to mitigate strain on the electrical grid, finding that, by leveraging flexibility, short, modest reductions in usage could allow 100 GW of large new loads to be accommodated with limited grid impacts. Additional research and demonstration projects are now exploring how precisely AI load flexibility can be harnessed for maximum grid benefit with minimal disruption to productivity. Interest in this capability is growing, and several notable demonstration projects have been completed, but more research, testing, and real-world usage are required to prove that load shifting is compatible with the reliability and performance demands of data centers.

Electric Power Research Institute's (EPRI's) DCFlex initiative indicates strong interest in using flexibility to support future data center infrastructure [63]. Their workstream on grid-informed data center designs outlines a framework for assessing data center flexibility potential. Planned next steps will continue to build the evidence base through use cases and real-world field trials. EPRI identifies three main categories of factors affecting an AI data center's ability to flex its load: technical, economic, and operational. By far the most important factor is the opportunity cost—loss of income—to companies when computational capability is unavailable.

Ability and willingness to flex are also affected by Service Level Agreements (SLAs), particularly in colocation data centers (where the facility owner controls building infrastructure and individual tenants control IT), as well as competition to provide high levels of uptime to customers and customer preferences. That said, given the imperative to accelerate time to power/speed to market, EPRI cites the possibility of some data centers connecting to the grid with only a partial load at an earlier date than the full load can be accommodated. Some types of workloads are more suitable for flexibility than others: batch workloads that can be scheduled (e.g., backups) or triggered (e.g., queued tasks), can tolerate delays better than interactive workloads (e.g., e-commerce), which are driven by users and where delay has a higher opportunity cost. A data center's ownership model also influences its flexibility. A colocation data center that serves multiple businesses/clients with different priorities is currently thought to be less amenable to flexing than a data center owned by a single entity.

3.4.3 Demonstrations of Data Center Demand Flexibility Capabilities

Demonstration projects show that data center load can be adjusted under certain conditions, but they do not establish that all data centers can provide reliable, repeatable, grid-scale flexibility. For policymakers and utilities, the central question is not simply whether a data center can reduce load in a pilot setting, but whether that flexibility can be measured, verified, compensated, and incorporated into grid planning in a dependable way. Below we highlight recent and ongoing field demonstration projects that indicate several distinct forms of data center flexibility when certain technological and workload criteria are met:

- **Emergency curtailment:** A data center's ability to rapidly cut power usage in response to unexpected grid-dispatched events while maintaining service-level agreements for critical workloads.
- **Behind-the-meter storage and islanding:** A data center's ability to consistently shift power demand away from evening peak demand times, integrating load flexibility strategies across computing workload, cooling, and backup systems.
- **Planned peak shifting:** Shifting to islanded operation and providing power back to the grid via utility-scale, behind-the-meter batteries in response to grid requests.
- **Geographic workload shifting:** Locational shifts of workload between geographically separated facilities in response to grid constraints.

These categories matter because they have different grid values and policy strategy implications. Emergency curtailment can help protect reliability during stressed conditions, but it may be too infrequent or uncertain to serve as the primary flexibility model. Planned peak shifting may be easier to incorporate into tariffs, demand response programs, and forecasting because it can be scheduled in advance. Behind-the-meter storage and islanding can provide valuable grid relief, but only if utilities and grid operators have sufficient visibility into the assets and confidence in their dispatchability. Geographic workload shifting could be powerful for hyperscale operators with large distributed portfolios, but it is less available to enterprise or colocation data centers.

Emergency curtailment: In the U.K., a collaboration of EPRI, EmeraldAI, National Grid, and Nebius demonstrated a "power flexible operational paradigm" for five days in December 2025 [64]. A cluster of 96 Nvidia Blackwell Ultra GPUs autonomously responded to 22 "live dispatch events," including surprise events. The cluster was able to cut power use by 30% in under 40 seconds, a performance similar to what industrial battery storage could provide, while still meeting the terms of its SLAs for critical workloads.

Behind-the-meter storage and islanding: Another demonstration was completed at the National Lab of the Rockies (NLR; formerly National Renewable Energy Laboratory) in early 2026. The experiment used NLR's controlled simulation environment, the Advanced Research on Integrated Energy Systems (ARIES) platform. ARIES powers Vulcan, NLR's 70 MW utility-scale test platform [65]. The experiment used utility-scale batteries in battery energy storage systems to shift load and return energy back to the local grid and achieved fast responses to utility requests (under 10 seconds) as well as uninterrupted shifts to islanded operation; like the Nebius/EmeraldAI experiment in the U.K., this demonstration met the terms of all customer SLAs.

Planned peak shifting: EPRI's DCFlex initiative oversaw another experiment at a data center in Phoenix, involving Oracle, EmeraldAI, NVIDIA, the software company Databricks, and the utility Salt River Project. EmeraldAI's software enabled the data center to reduce power consumption by 25% during evening peak demand [66]. This experiment only manipulated computing workloads, but in future experiments at its nine other test sites, DCFlex will experiment with flexibility in both cooling and backup power. Flexibility in all three areas (workload, cooling, and backup) could enable larger reductions in power use. In April 2026, Silicon Valley Power (the municipal utility for Santa Clara, California) and EmeraldAI partnered for a pilot of data center load flexibility [67]. The project will use NVIDIA's GPUs and EmeraldAI's Conductor platform at a multi-megawatt scale to demonstrate that data centers can respond to utility-directed flexibility needs during peak hours.

Geographic workload shifting: EmeraldAI also self-reported a spatial flexibility experiment focused on shifting AI inference workloads between regions. This demonstration shifted load from Virginia to Illinois without compromising service quality [68].

These examples show that flexibility is not a single capability. It can come from IT workload scheduling, cooling-system controls, UPS and battery systems, backup power architecture, geographic distribution, or some combination of these strategies. As a result, policy should avoid treating "data center flexibility" as a generic claim. A more useful approach would define flexibility by measurable attributes: how much load can be reduced, how quickly, for how long, how often, under whose control, and with what effect on service quality, emissions, water use, backup generation, and total system costs.

The demonstration projects also highlight unresolved barriers to scaling data center flexibility. Grid operators need confidence that reductions will be available when called upon. Regulators need standardized measurement and verification protocols. Utilities need visibility into behind-the-meter assets, load-management actions, and backup generation. Data center operators need compensation structures that reflect opportunity costs, service-level obligations, and reliability risk. Customers and tenants need assurance that participation in flexibility programs will not compromise service quality. Without these elements, flexibility may remain a promising technical capability rather than a dependable system resource.

The broader lesson is that data center flexibility should not be treated as a universal capability or a voluntary public-relations claim. It should be treated as a defined, verified, and differentiated performance attribute. When technically feasible and appropriately compensated, flexibility can help data centers reduce grid impacts, accelerate interconnection, improve utilization of existing infrastructure, and support clean energy integration. But its value will depend on whether it is reliable enough for grid operators, transparent enough for regulators, and practical enough for operators and customers to sustain over time. None of this has yet been sufficiently demonstrated, and considerably more work is needed.

3.5 Accelerating the Adoption of Emerging Technology through a Performance Evaluation Framework

A performance evaluation framework can help to accelerate the adoption of emerging technologies by creating a clear set of expectations and verification mechanisms. While data center researchers, operators, and developers have demonstrated the viability of high efficiency technologies that can substantially reduce data center energy and water demand, as well as grid and community impacts, these technologies are not consistently incorporated into data center design. Clear and consistent metrics and performance requirements can provide confidence that these new approaches deliver measurable benefits while maintaining credibility and accountability. In addition to setting clear benchmarks and performance baselines, policymakers and utilities can further reward market leaders that take on the risks and upfront costs associated with pioneering new technologies through incentives, expedited permitting, preferential rates, enhanced interconnection opportunities, or public recognition.

Thus, the performance evaluation framework not only establishes minimum performance requirements to protect public benefits but also creates a pathway for continuous improvement by recognizing and rewarding facilities that push the frontiers of efficiency, flexibility, and sustainability. The next part of this report further discusses these market mechanisms to adopt the framework and deliver results.

Part IV: Key Stakeholders and Market Drivers

Parts I through III establish the need for a stronger data center performance framework, identify the policy mechanisms that can carry its requirements, and describe the technologies available to meet them. However, a framework will not achieve broad impact through policy and technology alone. Adoption will depend on the wide range of stakeholders that shape how data centers are financed, sited, designed, built, and operated.

No single actor controls the full data center life cycle. Hyperscale technology companies create demand and establish procurement requirements, developers and operators secure sites and deliver projects, utilities control access to power, governments establish approval and incentive conditions, and designers, manufacturers, contractors, and service providers determine how requirements are implemented. Each operates with different incentives, timelines, information, and levels of authority.

This section identifies the stakeholders with the greatest ability to drive adoption of the framework and examines the market dynamics that can either accelerate or constrain implementation.

4.1 Key Stakeholders and Their Interests

The data center market includes a layered network of organizations that influence project demand, design, infrastructure, approval, construction, and operation. A snapshot of the key market actors and their key drivers, the concerns of either the developers that might limit growth or concerns of other stakeholders about data center growth around them, are listed in Table 7 below.

TABLE 7. KEY STAKEHOLDERS AND DRIVERS

Stakeholder Type	Examples of Organizations	Key Drivers / Primary Interest	Concerns Related to Development *(their own development or development around them)*
Data Center Developers and Operators	Compass Datacenters, QTS	Rapid deployment, land acquisition, build-to-suit growth	Permitting uncertainty, speculative load risk, incomplete tenant information, local opposition, infrastructure constraints
IT Manager	Any company which operates a data center day to day	Operational reliability, security, and productivity	Power availability and interconnection, grid reliability and resilience, tariff design, supply chain and workforce disruptions

Colocation / Real Estate Investment Trusts (REIT) Operators	Equinix, Digital Realty, Iron Mountain	Occupancy, lease competitiveness, stable rates, tenant attraction	Split incentives between operators and tenants, unclear tariff responsibility, capital constraints, inconsistent emissions accounting
Hyperscale Technology Companies	Google, Microsoft, Amazon, Meta, Oracle	AI growth, compute demand, power access, sustainability goals, rapid deployment	Interconnection queues, emissions scrutiny, water, community distrust, tension between sustainability and business priorities, rising capital costs
General Contractor	Turner Construction, JE Dunn, Hold Construction	Project growth, innovation, maintaining project pipelines	Labor shortages, supply chain issues, compressed schedules, local disruption concerns
Design / Engineering / Product and Service Providers	Architectural and engineering firms, HVAC manufacturers, IT equipment suppliers, electrical equipment suppliers	Performance optimization and technology integration	Limited operational data, rapidly changing AI workloads, uncertainty around future cooling and power needs
Standards and Certification Organizations	ASHRAE, ICC, GEC, NERC, USGBC, GBI	Industry standards and market transformation	Slow update and adoption cycles, industry resistance, incomplete operational datasets
Utilities and Operators, Regulators	Dominion Energy/ Virginia State Corporation Commission/ PJM; Oncor/ERCOT/ Public Utility Commission of Texas; AEP Ohio/ MISO/Public Utilities Commission of Ohio	Reliability, load growth, infrastructure planning, cost recovery	Speculative load forecasts, limited visibility into behind-the-meter resources, AI demand volatility, infrastructure timing mismatches, political pressure around rates, customer satisfaction
State and Local Governments	Loudoun County, Prince William County, Virginia; Dallas/Collin/Denton Counties, Texas; Franklin County, Licking County, Ohio	Economic development, affordability, reliability, community compatibility	Slow regulatory timelines, incomplete operational data, competing political pressures, limited authority
Research Organizations and Coalitions	EPRI, LBNL, PNNL, Uptime Institute, iCA	Understand efficiency and flexibility and their grid and water impacts	Limited operational datasets, uncertainty around scalability, translating pilots into policy
Advocacy and Community Groups	Environmental advocacy groups, residents/utility customers, community groups	Environmental justice, quality of life, transparency	Information mismatched or incomplete, trust, limited authority, uneven distribution of benefits

The stakeholder groups do not act independently but are instead part of an ecosystem of decision-making. Hyperscalers may establish performance expectations, but developers, designers, equipment providers, and contractors must deliver them. Utilities may establish service conditions, but regulators may need to approve the applicable tariff or agreement.

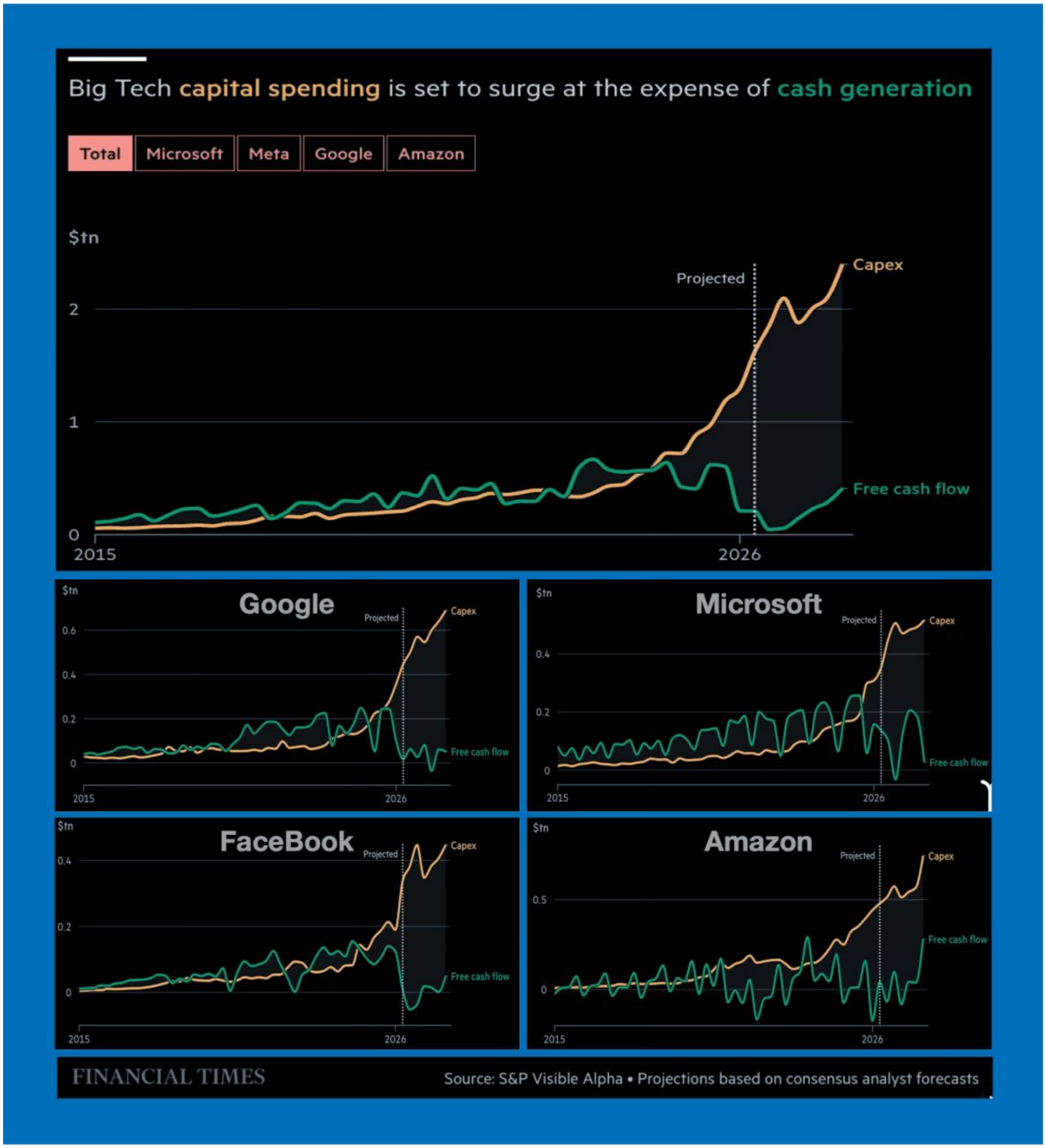
Governments can attach conditions to permits and incentives, while research organizations provide the metrics and technical methods needed to evaluate compliance.

These relationships are important because adoption is shaped not only by whether a technology or standard exists, but also by where the demand is, who pays for it, who is responsible for implementing it, and who is ultimately accountable for the result.

4.2 Driving Market Demand Through Procurement Pathways

Hyperscale technology companies and other large customers are among the strongest drivers of data center development. Their demand and finances drive much of the current expansion in data center capacity [69]. Even when they do not directly design or construct a facility, instead choosing to move into an existing space, they influence its power density, uptime requirements, cooling approach, energy sources, and environmental impacts through procurement specifications, leases, and operational expectations.

FIGURE 4. BIG TECH CAPITAL SPENDING IS TO SURGE AT THE EXPENSE OF CASH GENERATION. ADAPTED FROM DELONG 2026 [67]



View detailed description of Figure 4.

Procurement can accelerate change in the building sector because owner requirements are translated directly into project design, specifications, contracting, and material and equipment selection. These requirements affect actors throughout the building supply chain, from developers to manufacturers, designers, and energy providers. When large owners apply consistent design and performance expectations across multiple projects, they create a stronger and more predictable demand signal for higher-performing building systems, products, and services, making optimization of all three more commercially attractive [70].

Several data center companies, including Google, have ambitious sustainability goals [71], but corporate sustainability commitments are rarely the top driver for investment decisions. Sustainability decisions generally emphasize long-term environmental outcomes, while business and operations teams prioritize cost, reliability, available power, and speed to deployment. The most critical bottleneck for data center expansion is power. Interconnection queues now last four years or longer [72], significantly delaying new capacity [73] [74]. When these priorities conflict or slow progress, business needs are prioritized.

Stricter standards are therefore most likely to spread through procurement when they also support the bottom line and overall business outcomes. When requirements improve reliability, reduce costs, support faster interconnection, lower permitting risk, or provide greater certainty across markets, they are more likely to be adopted consistently rather than treated as optional sustainability measures.

4.3 Resolving Fragmented Ownership Through Clear Allocation of Responsibility

The data center market includes several ownership and operating models. Developers may acquire land and construct campuses before final tenants or operational characteristics are known. Colocation providers own and operate shared facilities, while tenants control much of the computing load. Hyperscalers may own and operate facilities directly or lease capacity from another company.

FIGURE 5. KEY CHARACTERISTICS OF DATA CENTER TYPES [74] [75]

	Enterprise / On-Premise	Colocation / REIT (Equinix, Digital Realty)	Hyperscale / Captive (AWS, Google, Meta)
Operate As	Owner or tenant	Third party real estate operator	Single tenant, usually owner occupied (sometimes leased)
Usage	Internal IT	Multiple competing tenants	Public cloud/AI
Scale	Small-medium (below 100 MW)	Medium-large (100–200 MW)	Huge (100 MW-1 GW+)

View accessible table of Figure 5.

This layered structure can make accountability confusing. Developers may provide projections during permitting before operating conditions are fully defined. Operators have access to and could report building- or portfolio-level performance but have limited control over tenant workloads. Tenants may drive energy use without being responsible for building improvements, utility tariffs, or site-level impacts.

The challenge is particularly significant in colocation facilities. Operators may control the central infrastructure, but tenants determine how much computing equipment is installed, when it operates, and whether workloads can be shifted. Responsibility for energy costs, flexibility participation, renewable energy use (bought or generated), and capital improvements may be divided among several parties.

These ownership structures influence adoption because investments are less likely to move forward when the party paying for an improvement does not receive its full benefit. Clearer allocation of responsibilities and requirements around use and financial investment over time through leases, service agreements, and operating contracts can reduce these split incentives. Stricter standards will also be easier to implement when they distinguish among building systems, tenant equipment, energy procurement, load flexibility, reporting, and verification rather than assigning more generic building-level requirements or assigning to just the building operator.

4.4 Catalyzing Service and Technology Markets Through Clear Target Setting

The data center delivery ecosystem, including the firms responsible for design, construction, equipment, and operations, determines how project expectations are translated into functioning facilities. Many of the efficiency and flexibility technologies discussed in Part III are already commercially available. Market adoption depends on whether they are consistently specified, properly integrated, successfully commissioned, and maintained during operation.

Data center projects remain highly customized, with designs shaped by site conditions, local regulations, operational requirements, and developer needs [75]. Design decisions vary according to workload, rack density, redundancy, cooling strategy, available energy resources, and future expansion plans. Designers and suppliers also have limited access to actual operating data from completed facilities, making it difficult to determine how projected performance compares with actual outcomes that could be improved in future projects. Uncertainty about future loads and operating conditions can encourage conservative or oversized designs as project teams prioritize reliability and capacity over energy efficiency [76].

Specialized mechanical and electrical labor remain limited, while equipment and material constraints continue to affect project delivery. Under compressed schedules, teams may rely more heavily on standardized designs, prefabricated systems, and familiar construction approaches [77]. At the same time, the growing scale, complexity, and speed of data center development have increased the use of modular construction, prefabrication, building information modeling, and other advanced design and delivery methods [78].

Consistent performance expectations can help the delivery market move from one-off solutions toward repeatable and scalable design practices. Designers can incorporate requirements earlier, manufacturers can develop products around common performance demands, contractors can build expertise in installation, and commissioning providers can use standardized testing methods appropriate for data centers. Clear outcome-based standards can therefore support innovation while allowing project teams to choose the technical solution best suited to each site and workload.

4.5 Supporting Speed to Market Through Well Defined Service Conditions

Access to adequate and reliable power is one of the strongest forces shaping data center site selection and development [79]. Available capacity, interconnection timelines, infrastructure costs, rate certainty, and reliability frequently determine whether a project proceeds and where it is located. Other factors including fiber connectivity, land, water, climate, and tax policy remain important, but power is often the binding constraint.

This gives utilities and grid operators considerable influence over market adoption. Interconnection applications, service agreements, tariffs, and energy procurement establish the terms under which a project receives power. When earlier energization or access to constrained capacity is linked to accurate forecasting, financial commitments, flexibility, or operational reporting, these requirements can become part of the standard approach as opposed to a secondary action.

Utilities must balance the revenue associated with large new customers against infrastructure risk, forecasting uncertainty, reliability, and the possibility that costs will be shifted to other ratepayers. Traditional utility planning processes were generally built around slower, more incremental load growth, rather than individual projects seeking hundreds of megawatts on development schedules that may be much shorter than utility planning and construction timelines [80].

Limited visibility into behind-the-meter generation, backup systems, storage, and workload-management strategies further complicate planning. These resources can materially change both projected and real-time demand, yet they may not be fully represented in utility models.

Because power access is so central to development, utility and interconnection processes may be among the fastest pathways for changing market behavior. Well defined and transparent service conditions can influence siting, design, operations, and investment well before a facility begins operating.

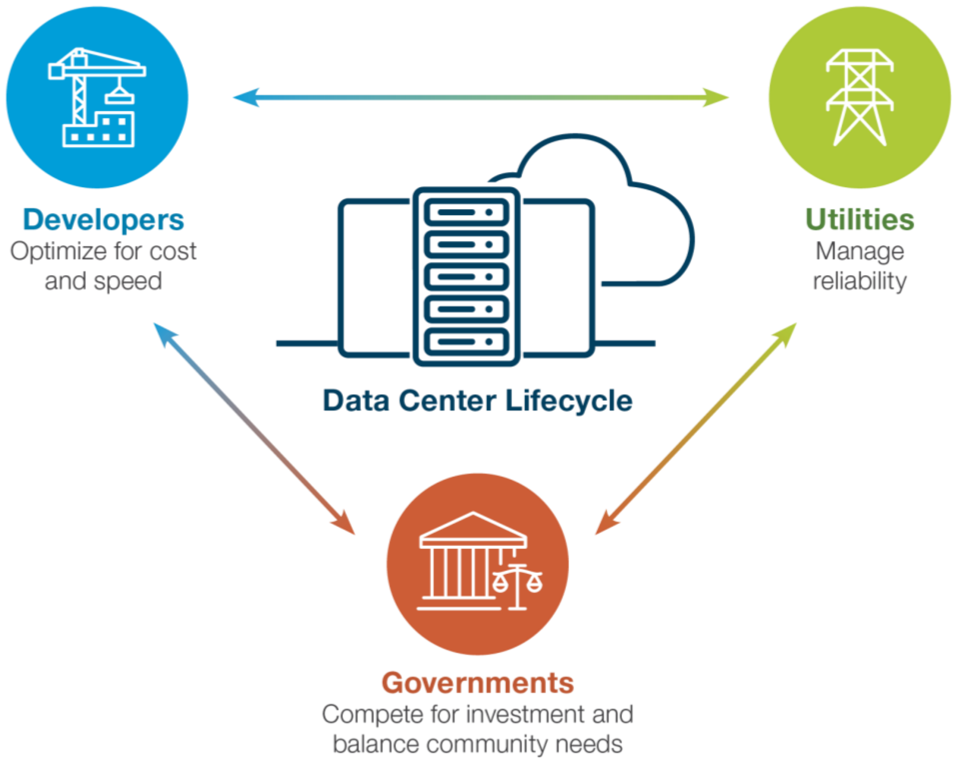
4.6 Establishing Consistent Expectations Across States to Avoid a Race to the Bottom

Developers and hyperscalers can compare jurisdictions and move projects toward locations offering available power, favorable rates, fast approvals, tax incentives, suitable land, and fewer regulatory constraints. States and local governments compete for these projects because of their investment, tax revenue, and potential economic development value.

This competition can discourage stricter requirements when jurisdictions fear that developers will move to a less regulated location. Conversely, stronger data center performance can also be rewarded with expedited approvals, incentives, infrastructure support, or quicker utility service for those projects that can meet stricter requirements.

The market is therefore shaped less by a single central decisionmaker than by negotiation among actors with different forms of leverage. Developers and hyperscalers can select among locations. Utilities control access to power but operate within regulated processes. Governments control incentives and approvals while balancing development with affordability and community impacts.

FIGURE 6. A SIMPLIFIED VIEW OF KEY STAKEHOLDERS IN THE DATA CENTER LIFE CYCLE



Greater consistency across jurisdictions can reduce concerns that stronger requirements will place individual communities at a competitive disadvantage or shift development toward locations with fewer protections [81]. A coordinated approach may also help prevent data center impacts from becoming concentrated in communities with less capacity to manage them. Common definitions and comparable requirements can also provide developers with greater certainty when assessing multiple markets. Standardization does not require every site to use the same technical solution, but it can establish a more consistent basis for evaluating project performance.

4.7 Earning Community Support Through Transparency and Accountability

Communities and advocacy groups can influence project approvals and local development, particularly when data center proposals raise concerns about utility affordability, environmental impacts, and how costs and benefits are distributed [82]. Data centers can generate substantial investment and local tax revenue, but many of their direct impacts are experienced near individual sites [83]. Residents living close to facilities may face persistent noise, construction activity, visual impacts, and pressure on local energy, water, and land resources, even when the broader jurisdiction receives financial benefits.

Opposition is often intensified when information is incomplete or arrives late. Communities may not know a project's expected power demand, cooling and water needs, backup-generation impacts (e.g., air quality), tenants, or expansion plans until after major decisions have been made. They may then be asked to rely on projections that are neither consistently reported nor independently verified, creating distrust that can be difficult to overcome. Recent research on community-centered data center development identifies late engagement and limited transparency as recurring features of stalled projects, while earlier engagement and measurable commitments are associated with greater community acceptance [84].

The growing use of data center moratoria demonstrates the practical consequences of this trust gap. By July 2026, lawmakers in 15 states were considering development bans or pauses, while local governments including Seattle, Baltimore, Dane County (Wisconsin), Chatham County (North Carolina), and Spokane, Washington had already adopted moratoria to study potential effects and develop more appropriate standards [85]. These actions show that community opposition can delay approvals, increase costs, create political risk, and make future development more difficult. Earlier disclosure, comparable reporting, and independent verification can make project impacts more understandable and help shift engagement from reactive opposition toward negotiation over enforceable conditions and community benefits. This can ultimately help shift the narrative about data centers by showing how they can contribute valuable assets to our built environment ecosystem if designed and managed appropriately.

4.8 Creating Strong Standards, Building Evidence, and Establishing Feedback Loops

Standards organizations, research institutions, technical coalitions, and certification bodies create the common technical language needed for broader market understanding and standardization. They define metrics, establish testing and calculation methods, evaluate emerging technologies, document best practices, and help translate technical findings into guidance for policymakers and industry.

Research and pilot projects are particularly important where technologies exist, but market actors remain uncertain about cost, reliability, operational risk, and scalability. Pilot projects and field studies can provide actual performance data, identify risks, and establish best practices before wider implementation [72]. Independent evaluation can further help utilities, regulators, developers, and customers assess technology performance and determine whether a technology or operating strategy is ready for broader deployment [86].

Standards and certification systems can also reward leadership and create market differentiation. However, they tend to evolve more slowly than data center technologies and workloads. Voluntary certification programs must balance technical rigor against broad market participation. Most importantly, they are ultimately not required, which can limit their ability to establish consistent and enforceable minimum performance expectations.

Their most important role in market adoption may be establishing comparability. Shared definitions, metrics, and verification methods allow customers, utilities, governments, and communities to evaluate projects on a consistent basis rather than relying on claims from specific companies or developers. Operating data can then be used to improve technical guidance, update standards, and reduce uncertainty and distrust in the market.

4.9 Driving Adoption Across the Data Center Life Cycle

Data center requirements must endure through a chain of decisions involving customers, developers, utilities, regulators, project teams, and operators. A performance or reliability expectation established early in development must be reflected in utility agreements and programs, project specifications, construction, commissioning, and ongoing operations to achieve a result that is ultimately more efficient and flexible. Table 8 shows where policy can influence each stage of the data center life cycle and which market actors are best positioned to turn those requirements into standard industry practice.

TABLE 8. POLICY AND MARKET INFLUENCE ACROSS THE DATA CENTER LIFE CYCLE

Development Stage	Where Policy Can Influence	Who Can Drive Adoption
Planning and Siting	• Set siting and permitting conditions • Require early disclosure and environmental review • Tie incentives to performance	• Data Center Developers and Operators • Hyperscale Technology Companies • Utilities and Regulators • State and Local Governments • Advocacy and Community Groups
Interconnection, Design, and Approval	• Establish interconnection and utility service requirements • Set building performance and flexibility expectations • Influence project size and system design • Tie utility incentives to performance	• Utilities and Regulators • Data Center Developers and Operators • Hyperscale Technology Companies • Design, Engineering, and Service Providers • Standards and Certification Organizations • Research Organizations and Coalitions
Construction and Commissioning	• Carry requirements into permits and specifications • Require testing and documentation • Verify performance before operation	• Data Center Developers and Operators • Colocation and REIT • Construction Industry • Design, Engineering, and Service Providers • Construction Industry • Standards and Certification Organizations
Operations	• Shape energy use through tariffs and flexibility programs • Set energy, emissions, and reporting requirements • Require ongoing performance verification	• Data Center Developers and Operators • Colocation and REIT • Hyperscale Technology Companies • Utilities and Regulators • Standards and Certification Organizations • IT managers
Expansion or Retrofit	• Trigger review when load or equipment changes • Apply updated service and performance requirements • Require new reporting or verification	• Data Center Developers and Operators • Colocation and REIT • Hyperscale Technology Companies • Utilities and Regulators • Design, Engineering, and Service Providers • Construction Industry

The table shows that policy adoption depends on more than establishing requirements. At each stage, market actors must incorporate those requirements into their decision-making. Aligning expectations across these life cycle points can help stricter standards become part of normal project delivery rather than separate or voluntary measures.

4.10 Market Adoption of a Consistent Performance Framework

The market dynamics discussed in this section show that stricter standards or high-performance targets will not be adopted through one policy, company, or stakeholder group. Adoption will occur when expectations are reinforced across the decisions that determine whether projects receive customers, capital, land, power, approvals, equipment, and public support.

Adoption of a more comprehensive framework to evaluate data center energy performance across multiple dimensions depends on three reinforcing market forces:

- **Market pressure:** Broader market expectations from investors, customers, and regulators can push large customers to require stronger performance through procurement, leasing, and site selection.
- **Project and policy delivery:** Stronger expectations gain traction when they are built into the decisions and requirements that shape how data centers are developed and operated.
- **Accountability and market learning:** Common metrics, independent research, verified reporting, and public transparency can demonstrate results and improve future standards and community trust.

When these forces reinforce one another, stricter performance requirements or higher performance targets are more likely to become part of normal data center development rather than remain isolated or voluntary practices. Market pressure creates the incentive to act; project and policy processes translate expectations into implementation; and performance data provides evidence of what is achievable.

Standardization supports this process by creating a common basis for action across jurisdictions, companies, and project types. A stronger approach should establish consistent performance outcomes, metrics, and reporting methods while preserving flexibility for facilities to select technologies suited to their location, workload, and operating needs. Verified operating data can then inform future revisions, creating a continuous cycle in which market expectations encourage stronger requirements, those requirements shape project delivery and operations, and actual performance supports the continued improvement of standards.

Conclusion: A Data Center Energy Performance Evaluation Framework for a Better Coordinated Path Forward

Today's data center landscape lacks a shared, credible way to assess whether a proposed or operating facility avoids harm to utilities and adjacent communities—or whether utilities and local governments are incentivizing truly best-in-class facilities. This gap points to the need for an evaluation framework that can serve as the connective layer between what gets built and how communities, utilities, and policymakers judge its performance.

We propose a future-ready Data Center Energy Performance Evaluation Framework (DC-EPEF) to help policymakers and communities understand what is being proposed and determine whether it represents the best available option—one that lowers long-term risks and liabilities while delivering broader benefits over the near and long term. An effective framework should cover four critical dimensions of data center energy performance.

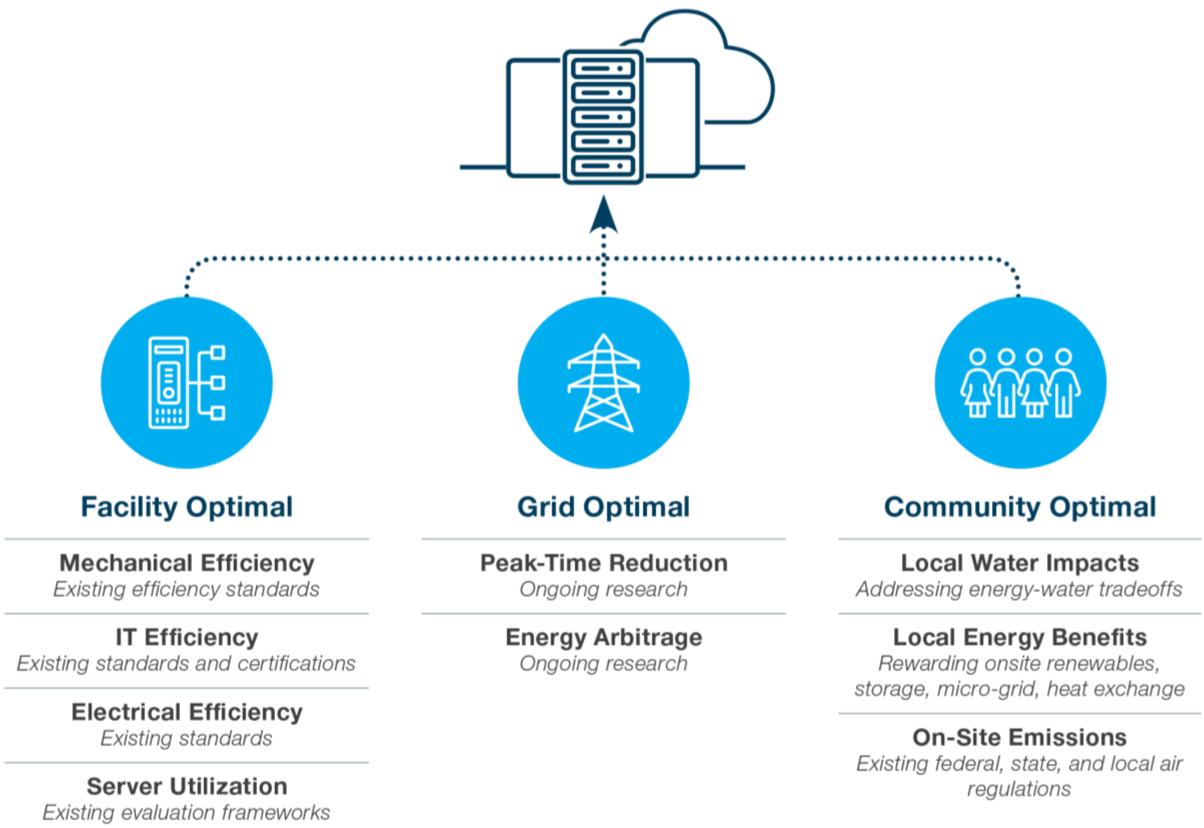
- **Energy, water, and emissions metrics** that go beyond mechanical and electrical system efficiency to capture total resource consumption and environmental impact.
- **Grid stress and demand flexibility targets** that measure a data center's ability to adjust workload and energy demand in response to grid conditions—which are increasingly critical as large loads multiply.
- **Local flexibility and technological adaptability** to account for regional grid, water, climate, and community conditions while keeping standards current with rapid industry change.
- **Tiered requirements** combining mandatory minimums tied to grid interconnection with more ambitious targets linked to financial incentives and certification—driving cost-effective adoption at scale without sacrificing ambition.

The proposed framework is organized around three pillars—Facility Optimal, Grid Optimal, and Community Optimal—to balance the interests of data center owners, utilities, and surrounding communities.

- **Facility Optimal** focuses on the internal performance of the facility, including mechanical and electrical efficiency informed by standards from ASHRAE, ENERGY STAR, and TGG; IT equipment efficiency informed by certification systems such as the Electronic Product Environmental Assessment Tool, (EPEAT) and ENERGY STAR; and server utilization practices informed by industry efforts such as those from the Uptime Institute.

- **Grid Optimal** addresses how the facility interacts with the power system through peak-time reduction, load flexibility, and energy arbitrage. NBI has developed GridOptimal metrics for commercial buildings, and those metrics have been integrated into green codes and green building rating systems. For data centers, grid metrics should be grounded in ongoing research and demonstrations on demand flexibility for large loads from EPRI and the National Labs, translating those findings into practical policy and program actions.
- **Community Optimal** captures local external impacts, including water use and energy-water tradeoffs; designs flexible energy management systems with tools such as storage, microgrids, and heat exchange; and complies with on-site air and environmental performance governed by existing federal, state, and local regulations, or community benefit agreements.

DATA CENTER ENERGY PERFORMANCE EVALUATION FRAMEWORK (DC-EPEF)



View the long description for the DC-EPEF.

Each component is defined by a metric, minimum requirements that may vary by location, and tiered targets that distinguish different levels of performance. Because the performance targets build on existing standards and research, they can be adopted flexibly—as mandatory requirements, voluntary ratings, or incentive programs—through current policy and market mechanisms.

Our framework builds on approaches that have already been successfully developed and implemented, while introducing new components informed by emerging research and pilot projects. New measures often take time to be incorporated into formal standards, and our framework can help bridge that gap by advancing these metrics sooner. The framework is focused on what can be measured, tracked, and verified. This is a significant gap with regard to community impacts, where there is often no reliable way to assess whether promised benefits—such as jobs, tax revenue, or support for schools, solar, and heat pumps—are fair and equitable.

We envision the framework as spanning a spectrum of instruments, from a baseline floor to aspirational leadership, and operating through two complementary channels. **Policy mechanisms** (e.g., building codes, interconnection agreements, permitting) establish the mandatory foundation: a “do no harm” floor for water, air, and environmental impacts, minimum thresholds for energy performance, legislation for consumer and community protection, and relevant state, local, and utility mandates that govern compliance. **Market mechanisms** drive performance above the floor through voluntary rating systems, state, local, and utility incentives, and corporate commitments that signal leadership. Existing best practice guides and tools can help designers and operators meet the defined targets and deliver expected outcomes. Through a common language for evaluating and rewarding responsible development, the framework will help move the industry from fragmented efforts toward coordinated action across multiple layers.

One of the biggest challenges facing the development of standards is that technology evolves faster than traditional regulatory cycles. Requirements tied to specific technologies can become outdated within a few years, potentially locking in today’s solutions while discouraging future innovation. Rather than prescribing particular technologies, data center performance requirements or targets should focus on defining desired outcomes and measuring performance against those outcomes. For example, instead of requiring liquid cooling or heat recovery, policymakers should focus on setting expectations for energy efficiency, on-site emissions, water use, or demand flexibility.

The data center industry—which is now largely inextricable from the AI industry—is highly technical and sophisticated. Compared with sectors such as commercial buildings or affordable housing, there are fewer market participants and substantially more capital available for building and using data centers. As a result, traditional deployment barriers—such as cost constraints, training needs, or limited technical capacity—are expected to be less significant. This creates an opportunity to be ambitious and push the boundaries of what can reasonably be expected and measured.

An effective performance framework should also be adaptive. Expectations should be linked to measurable, outcome-based metrics that can be updated over time as technologies, operating practices, and market conditions evolve, while maintaining a consistent overall framework. At the same time, performance targets should reflect local priorities and constraints. Water efficiency may be a higher priority in drought-prone regions, while grid reliability, affordability, or emission reduction may be more important elsewhere. A common framework with location-specific targets can provide consistency across jurisdictions while preserving the flexibility needed to address local needs and circumstances.

In summary, the goal is not to create a single idealized vision that every project must meet immediately. Rather, it is to establish an ambitious but practical pathway for continuous improvement. Policymakers should be clear about which outcomes are non-negotiable and where temporary tradeoffs or phased approaches may be acceptable as technologies, infrastructure, and markets mature. This is where requirements and incentives can work together. Requirements establish the minimum floor and define the guardrails for acceptable performance. Incentives reward projects that move beyond that floor by adopting emerging best practices, demonstrating leadership, and delivering broader public benefits. Using a common measurement framework for both creates a transparent pathway through which today's stretch goals can become tomorrow's baseline expectations, helping drive innovation while ensuring accountability and continuous progress.

Appendix A: Detailed Image Descriptions

Detailed Description of Data Center Energy Performance Evaluation Framework (DC-EPEF)

Facility Optimal:

- Mechanical Efficiency: existing efficiency standards
- IT Efficiency: existing standards and certifications
- Electrical Efficiency: existing standards
- Server Utilization: existing evaluation frameworks

Grid Optimal:

- Peak-Time Reduction: ongoing research
- Energy Arbitrage: ongoing research

Community Optimal:

- Local Water Impacts: addressing energy-water tradeoffs
- Local Energy Benefits: rewarding onsite renewables, storage, micro-grid, heat exchange
- On-Site Emissions: existing federal, state, and local air regulations

Return to image.

Detailed Description of Figure 1

Twenty-Two States Meet or Exceed Standard 90.1-2019

Arizona, California, Connecticut, Delaware, Illinois, Maine, Maryland, Massachusetts, Michigan, Minnesota, Montana, New Jersey, New Mexico, New York, Oregon, Pennsylvania, Rhode Island, Utah, Vermont, Virginia, Washington, and Wisconsin.

Eight States Meet Standard 90.1-2016

Colorado, Florida, Hawaii, Louisiana, Nevada, North Dakota, Ohio, and Tennessee.

The Remaining Twenty States Equivalent Code is pre-2016

Return to image.

Accessible Table of Figure 2

Who Can Regulate	Planning & Siting	Design & Approval	Construction	Operations & Commissioning	Future Expansion or Retrofit
Local Government & Constituents	Yes	Yes	Yes	No	Yes
State Government & PUCs *(dependent on state)*	Yes	No	No	Opportunity!	Yes
Utility	Yes	No	No	Yes	Yes
Federal *(EPA and FERC)*	Yes	Technical Support	Technical Support	Technical Support	Technical Support
Standards & Certification Bodies	No	Yes	No	Yes	Yes

Return to image.

Detailed Description of Figures 3A and 3B

Figure 3A, on left side of panel

Figure 3A is an area chart that depicts the projected increase in global electricity consumption by data centers. The bulk of the expected increase, from roughly 400 TWH in 2024 to an expected 900+ TWH hours in 2030, is driven by a dramatic increase in accelerated servers, but there is expected growth in consumption from conventional servers, other IT equipment, cooling, and additional data center infrastructure as well.

Figure 3B, on right side of panel

Figure 3B is a bar chart that depicts the share of electricity consumption in TWH in 2024, split by data centers and equipment type. The top bar represents all data centers, for which roughly 2/3 of electricity consumption goes to accelerated servers. The majority of the remaining consumption goes to cooling, with much smaller allocations to conventional servers and other IT equipment. This pattern is qualitatively similar for the different types of data centers depicted in the remaining 3 bars: enterprise, colocation and service provider, and hyperscale data centers. A substantially greater proportion of energy demand goes to cooling for enterprise data centers, compared to the other types. Hyperscale data centers consume much more energy for accelerated servers, with relatively much smaller proportions of energy consumption for cooling. The distribution of energy consumption for colocation and service provider centers roughly mimics the distribution of energy consumption across all data center types.

Return to image.

Detailed Description of Figure 4

A composite image of five line charts, each depicting, in trillions of dollars, capex (capital expenditures) vs free cash flow for the major big tech companies (Microsoft, Meta, Google, and Amazon), as well as an aggregate of the four.

In aggregate, while capex and free cash flow stayed roughly similar for all companies until the early 2020s, a very large increase in capex kicked off and is projected to continue, greatly outpacing free cash flow growth; while in 2023 the total each for both capex and free cash flow across the four companies was slightly more than $0.5 trillion, past 2026, capex is predicted to surge beyond $2 trillion, while free cash flow remains at roughly the same level as 2023. This pattern is qualitatively similar, individually, for each of the major big tech companies, although Amazon expects greater increases in free cash flow than the other three.

Return to image.

Accessible Table of Figure 5

Dimension	**Enterprise / On-Premise**	**Colocation / REIT** *(Equinix, Digital Realty)*	**Hyperscale / Captive** *(AWS, Google, Meta)*
Operate as	Owner or tenant	Third-party real estate operator	Single tenant, usually owner occupied (sometimes leased)
Usage	Internal IT	Multiple competing tenants	Public cloud/AI
Scale of Demand	Small-medium (below 100 MW)	Medium-large (100–200 MW)	Huge (100 MW-1 GW+)

Return to image.

Detailed Description of Data Center Energy Performance Evaluation Framework (DC-EPEF)

Facility Optimal:

- Mechanical Efficiency: existing efficiency standards
- IT Efficiency: existing standards and certifications
- Electrical Efficiency: existing standards
- Server Utilization: existing evaluation frameworks

Grid Optimal:

- Peak-Time Reduction: ongoing research
- Energy Arbitrage: ongoing research

Community Optimal:

- Local Water Impacts: addressing energy-water tradeoffs
- Local Energy Benefits: rewarding onsite renewables, storage, micro-grid, heat exchange
- On-Site Emissions: existing federal, state, and local air regulations

Return to image.

References

[1] R. Mural *et al.,* “AI, Data Centers, and the U.S. Electric Grid: A Watershed Moment.” Feb. 2026. [Online]. Available: https://www.belfercenter.org/research-analysis/ai-data-centers-us-electric-grid

[2] IEA, “Share of electricity consumption by data centre and equipment type.” IEA, Paris, 2024. [Online]. Available: https://www.iea.org/data-and-statistics/charts/share-of-electricity-consumption-by-data-centre-and-equipment-type-2024.

[3] ISO, “ISO/IEC 30134-3:2016(en) Information technology — Data centres — Key performance indicators — Part 3: Renewable energy factor (REF.” 2016. [Online]. Available: https://www.iso.org/obp/ui/#iso:std:iso-iec:30134:-3:ed-1:v1:en.

[4] The Green Grid, “ERE: A Metric for Measuing the Benefit of Reuse Energy from a Data Center.” 2010. [Online]. Available: https://datacenters.lbl.gov/sites/default/files/EREmetric_GreenGrid.pdf.

[5] Data Center Knowledge, “A Guide to Data Center Water Usage Effectiveness (WUE) and Best Practices.” Jan. 17, 2025. [Online]. Available: https://www.datacenterknowledge.com/cooling/a-guide-to-data-center-water-usage-effectiveness-wue-and-best-practices.

[6] U. S. D. O. E. FEMP, “Best Practices Guide for Energy-Efficient Data Center Design.” Jul. 2024. [Online]. Available: https://www.energy.gov/sites/default/files/2024-07/best-practice-guide-data-center-design_0.pdf.

[7] CBECS, “CBECS: Principal Building Activities.” 2018. [Online]. Available: https://www.eia.gov/consumption/commercial/pba/office.php.

[8] NLR, “An Analysis of Plug Load Capacities and Power Requirements in Commercial Buildings.” 2014. [Online]. Available: https://docs.nrel.gov/docs/fy14osti/61528.pdf.

[9] ICC, “Data Centers, the I Codes and a New Data Center Guideline.” 2026. [Online]. Available: https://www.iccsafe.org/building-safety-journal/bsj-technical/data-centers-the-i%E2%80%91codes-and-a-new-data-center-guideline/.

[10] E. Strickland, “15 Graphs That Explain the State of AI in 2024 The AI Index tracks the generative AI boom, model costs, and responsible AI use.” Apr. 15, 2024. [Online]. Available: https://spectrum.ieee.org/ai-index-2024.

[11] ANSI/ASHRAE, “ANSI/ASHRAE 90.4-2025, Energy Standard for Data Centers.” 90.4. [Online]. Available: https://webstore.ansi.org/standards/ashrae/ansiashrae902025?srsltid=AfmBOooHHr-rsSTJrP85Gtvmn4r3D5FrDHNxJUajZnGtBU0e_dSAIpOI.

[12] K. L. T. R. S. McCord, “PNNL/ASHRAE/NEMA AI Data Center Energy Performance Framework.” Jun. 2026. [Online]. Available: https://www.ashrae.org/technical-resources/ai-data-center-framework.

[13] M. D. Henni and K. S. Mohammed, “Exponential AI growth and the physical limits of renewable energy systems,” *Energy Policy,* p. 115314, 2026.

[14] E. Masanet, A. Shehabi, N. Lei, S. Smith, and J. Koomey, “Recalibrating global data center energy-use estimates,” *Science,* pp. 984-986, 2020.

[15] J. Davis, “AI’s growth calls for useful IT efficiency metrics.” Jan. 15, 2026. [Online]. Available: https://journal.uptimeinstitute.com/ais-growth-calls-for-useful-it-efficiency-metrics/.

[16] California Energy Comission, “Computer Rooms & Data Centers Fact Sheet.” Aug. 2023. [Online]. Available: https://www.energy.ca.gov/sites/default/files/2023-09/2022_CEC-Computer_Room_and_Data_Centers_ADA.pdf.

[17] California Energy Codes and Standards, “Cycle Data Centers Efficiency Improvements.” 2028. [Online]. Available: https://title24stakeholders.com/measures/2028-cycle/data-centers-efficiency-improvements/.

[18] Integral Group, “Energy Efficiency Baselines for Data Centers Statewide Customized New Construction and Customized Retrofit Incentive Programs.” Mar. 01, 2013. [Online]. Available: https://www.pge.com/content/dam/pge/docs/save-energy-and-money/rebate-and-incentives/data_center_baseline.pdf.

[19] J. Stein and B. Gill, “PG&E Data Center Baseline and Measurement and Verification (M&V) Guidelines,” Pacific Gas and Electric Company, 2016. [Online]. Available: https://www.calmac.org/publications/2016_PG&E_Data_Center_Baseline_and_M&V_Guidelines.pdf.

[20] U.S. Department of Energy, “Building Energy Codes Program.” [Online]. Available: https://www.energycodes.gov/state-portal.

[21] Energy Star, “Technical Reference, "ENERGY STAR Score for Data Centers.” 2018. [Online]. Available: https://www.energystar.gov/sites/default/files/tools/Data_Center_August_2018_EN_508.pdf.

[22] USGBC, “LEED v5 certification is here.” [Online]. Available: https://www.usgbc.org/leed/v5.

[23] GBI, “Green Globes New Construction 2024 Technical Reference Manual.” 2024. [Online]. Available: https://app.thegbi.org/education/library/f835515b-ac48-4289-8750-60cd448d528a/.

[24] BREEAM, “BREEAM Data Centres Sustainability Guide.” Dec. 03, 2020. [Online]. Available: https://cdn.asp.events/CLIENT_CloserSt_D86EA381_5056_B739_5482D50A1A831DDD/sites/DCW-London-2020/media/data-centre-design-build--physical-security/Day2_1400_Weikai-Gong.pdf.

[25] ISO, “ISO 14001:2026.” [Online]. Available: https://www.iso.org/standard/14001.

[26] ISO, “ISO 50001.” [Online]. Available: https://www.iso.org/iso-50001-energy-management.html.

[27] "Achieve U.S. Department of Energy (DOE 50001 Ready U.S. DOE, “Energy Management System Recognition.” [Online]. Available: https://navigator.lbl.gov/guidance/recognition.

[28] Google Cloud, “A policy roadmap for 24/7 carbon-free energy.” Apr. 14, 2022. [Online]. Available: https://cloud.google.com/blog/topics/sustainability/a-policy-roadmap-for-achieving-247-carbon-free-energy.

[29] ISO, “ISO/IEC JTC 1/SC 39.” 2012. [Online]. Available: https://www.iso.org/committee/654019.html.

[30] ISO, “ISO/IEC JTC 1 N 16587.” Nov. 10, 2025. [Online]. Available: https://jtc1info.org/wp-content/uploads/2024/02/SC-39-Business-Plan-2023.pdf.

[31] NS Senate, “The New York State Senate.” Oct. 17, 2025. [Online]. Available: https://www.nysenate.gov/legislation/bills/2025/A9136.

[32] SPP, “SPP’s High Impact Large Load Interconnection Solutions: Powering Growth, Accelerating Opportunity.” [Online]. Available: https://www.spp.org/markets-operations/high-impact-large-load-hill-integration/.

[33] T. S. Legislature, “Texas Legislature Online.” Jun. 20, 2025. [Online]. Available: https://capitol.texas.gov/BillLookup/history.aspx?LegSess=89R&Bill=SB6.

[34] Nevada Public Utilities Commission, “Nevada Public Utilities Commission.” Mar. 11, 2025. [Online]. Available: https://puc-onbase.nv.gov/api/Document/Aaw7d8rSpNgSbvIAA

[35] HB 3546, “Oregon State Legislature.” 2025. [Online]. Available: https://olis.oregonlegislature.gov/liz/2025R1/Measures/Overview/HB3546.

[36] AEP Ohio, “AEP Ohio Updates PUCO on Data Center Load: Figures Show Tariff is Working.” Feb. 13, 2026. [Online]. Available: https://www.aepohio.com/company/news/view?releaseID=10753.

[37] New York State Senate, “Senate Bill S6394.” Mar. 13, 2025. [Online]. Available: https://www.nysenate.gov/legislation/bills/2025/S6394/amendment/original.

[38] ICEF, “Sustainable Data Centers Roadmap.” Innovation for Cool Earth Forum, Oct. 2025. [Online]. Available: https://www.icef.go.jp/wp-content/themes/icef_new/pdf/roadmap/icef2025_roadmap.pdf

[39] IEA, “Global data centre electricity consumption, by equipment, Base Case, 2020-2030.” IEA, Paris, 2025. [Online]. Available: https://www.iea.org/data-and-statistics/charts/global-data-centre-electricity-consumption-by-equipment-base-case-2020-2030.

[40] SolarTech, “How Much Electricity Does a Data Center Use? Complete 2025 Analysis,” IAEI Magazine. Accessed: Aug. 12, 2026. [Online]. Available: https://iaeimagazine.org/electrical-fundamentals/how-much-electricity-does-a-data-center-use-complete-2025-analysis/

[41] SOCOMEC, “Understanding the power consumption of data centers.” [Online]. Available: https://www.socomec.us/en-us/solutions/business/data-centers/understanding-power-consumption-data-centers

[42] P. Powell, “Hyperscale vs Colocation Data Centers,” IBM. Accessed: Aug. 28, 2026. [Online]. Available: https://www.ibm.com/think/topics/hyperscale-vs-colocation

[43] A. Shehabi *et al.,* “United States Data Center Energy Report.” 2024. [Online]. Available: https://eta-publications.lbl.gov/sites/default/files/2024-12/lbnl-2024-united-states-data-center-energy-usage-report_1.pdf.

[44] K. McCord, T. Loxley, and S. Rosenstock, "PNNL/ASHRAE/NEMA AI Data Center Energy Performance Framework." Jun. 2026. [Online]. Available: https://www.ashrae.org/technical-resources/ai-data-center-framework.

[45] A. Shehabi, J. S. Smith, A. Hubbard, and N. Lei, *United States Data Center Energy Report.* Berkeley, California: Lawrence Berkeley National Laboratory, 2024.

[46] IEA, *World Energy Outlook Special Report.* Paris: IEA, 2024.

[47] Micron, "Every watt matters: How low-power memory is transforming data centers." Mar. 14, 2025. [Online]. Available: https://www.micron.com/about/blog/applications/data-center/every-watt-matters-how-low-power-memory-is-transforming-data-centers.

[48] B. B. Tanner, K. Derek, C. F., A. Renda, E. Tabassi, and A. W. Wyckoff, "Global energy demands within the AI regulatory landscape." Apr. 10, 2026. [Online]. Available: https://www.brookings.edu/articles/global-energy-demands-within-the-ai-regulatory-landscape/.

[49] J. Dietrich, "Cooling systems: balancing cost, energy and water use," *Uptime Institute,* 2024.

[50] J. Dietrich, "Water is local: generalities do not apply," *Uptime Institute,* 2025.

[51] D. A. Kez, A. M. Foley, F. W. B. H. Wong, and A. Dolfi, "AI-driven cooling technologies for high-performance data centres: state-of-the-art review and future directions." Aug. 2025. [Online]. Available: https://pdf.sciencedirectassets.com/282793/1-s2.0-S2213138825X00102/1-s2.0-S221313882500342X/main.pdf?X-Amz-Security-Token=IQoJb3JpZ2luX2VjEEUaCXVzLWVhc3QtMSJIMEYCIQC6tv1shvpmrlt0cXkxd75AMhI5PgoZeN6HZ1wVnRA2egIhAPCPyfqFh5QROiCXSISH1YMin8FBKM9qWSevqZOy7020.

[52] D. A. Kez, A. Foley, F. Ahmed, and J. Morrow, "Data Center Potential Flexibilities and Challenges for Demand Response to Facilitate 100% Inverter-Based Resources: A Review." Nov. 17, 2022. [Online]. Available: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4269631.

[53] Google Data Centers, "Advancing responsible water use at our data centers." [Online]. Available: https://datacenters.google/water/.

[54] Energy Star, "Raise the Temperature." [Online]. Available: https://www.energystar.gov/products/data_center_equipment/5-simple-ways-avoid-energy-waste-your-data-center/raise-temperature.

[55] H. Clancy, "Behind Amazon's industry-leading water efficiency score." Jun. 11, 2026. [Online]. Available: https://trellis.net/article/behind-amazons-industry-leading-water-efficiency-score/.

[56] Google DeepMind, "DeepMind AI Reduces Google Data Centre Cooling Bill by 40%." Jul. 2016. [Online]. Available: https://deepmind.google/blog/deepmind-ai-reduces-google-data-centre-cooling-bill-by-40/.

[57] Google DeepMind, "AlphaEvolve: A Gemini-powered coding agent for designing advanced algorithms." May 14, 2025. [Online]. Available: https://deepmind.google/blog/alphaevolve-a-gemini-powered-coding-agent-for-designing-advanced-algorithms/.

[58] D. Gardiner, A. Dixon, E. Kent, and H. Schuster, “Data Center Heat Reuse: The Opportunity for States.” David Gardiner and Associates, Sep. 2025. [Online]. Available: https://www.dgardiner.com/wp-content/uploads/2025/12/Data-Center-Heat-Reuse-The-Opportunity-for-States.pdf#page=3.

[59] K. Weiß, “From EU Framework to National Action: How Germany Regulates Data Center Energy Use.” Climate Law A Sabin Center Blog, Oct. 24, 2025. [Online]. Available: https://blogs.law.columbia.edu/climatechange/2025/10/24/from-eu-framework-to-national-action-how-germany-regulates-data-center-energy-use/.

[60] EU Covenant of Mayors, “Stockholm, Sweden : Heat recovery from data centres,” EU Covenant of Mayors. [Online]. Available: https://eu-mayors.ec.europa.eu/en/news/stockholm-sweden-heat-recovery-data-centres

[61] Uptime Institute, “Annual outage analysis 2024.” Mar. 2024. [Online]. Available: https://datacenter.uptimeinstitute.com/rs/711-RIA-145/images/2024.Resiliency.Survey.ExecSum.pdf.

[62] T. H. Norris, T. Profeta, and D. Patino-Echeverri, “Rethinking Load Growth: Assessing the Potential for Integration of Large Flexible Loads in US Power Systems,” Nicholas Institute for Energy, Environment & Sustainability, Duke University, Text, Feb. 2025. Accessed: Aug. 12, 2026. [Online]. Available: http://nicholasinstitute.duke.edu/publications/rethinking-load-growth

[63] EPRI, “Grid Flexibility Needs and Data Center Characteristics,” EPRI, Jun. 2025. Accessed: Aug. 12, 2026. [Online]. Available: https://www.epri.com/research/products/3002031504

[64] C. Williams *et al.,* “Power Flexible AI Factories: A UK-First Demonstration of Grid Responsive AI Infrastructure (March 2026) | National Grid Partners,” Emerald AI, EPRI, National Grid, Nebius, Mar. 2026. Accessed: Aug. 12, 2026. [Online]. Available: https://www.ngpartners.com/stories/emerald-ai-whitepaper

[65] D. Vaidhynathan *et al.,* “Vulcan Test Platform: Demonstrating the Data Center as a Flexible Grid Asset,” NLR, 2025. doi: 10.2172/2583521.

[66] P. Colangelo *et al.,* “AI data centres as grid-interactive assets,” *Nat Energy,* vol. 11, no. 2, pp. 254–261, Feb. 2026, doi: 10.1038/s41560-025-01927-1.

[67] Emerald AI, “Silicon Valley Power and Emerald AI Launch Pilot to Demonstrate Flexible Data Centers in Santa Clara and Unlock Power Capacity for AI - Emerald AI,” Emerald AI. Accessed: Aug. 12, 2026. [Online]. Available: https://www.emeraldai.co/blog/silicon-valley-power-and-emerald-ai-launch-pilot-to-demonstrate-flexible-data-centers-in-santa-clara?_sp=07aea959-d44b-4358-8100-ed3d20fc793a.1784838113458

[68] H. K. Trabish, “Data centers are ready to negotiate flexibility for speed | Utility Dive,” Utility Dive, Jun. 26, 2026. Accessed: Aug. 12, 2026. [Online]. Available: https://www.utilitydive.com/news/data-centers-flexibility-utilities-speed-to-power/822588/

[69] B. Delong, “The Gonzo Size of the HyperScaler DataCenter Investment Boom: Chart of the Day.” May 12, 2026. [Online]. Available: https://braddelong.substack.com/p/the-gonzo-size-of-the-hyperscaler?utm_source=substack&utm_medium=email&utm_content=share.

[70] UKGBC, *Engaging the Supply Chain: A Practical Guide to Delivering Sustainability Outcomes.* UK Green Building Council, 2026.

[71] Google, "Driving sustainable innovation for society." [Online]. Available: https://datacenters.google/operating-sustainably/.

[72] LBNL, "DOE Data Center Load Flexibility Workshop Summary." Jan. 2025. [Online]. Available: https://eta-publications.lbl.gov/sites/default/files/2025-03/final_doe_data_center_load_flexibility_workshop_summary.v0307.pdf.

[73] FERC, "Commissioner Chang's Dissent on PJM's RRI, ER25-712." Feb. 11, 2025. [Online]. Available: https://www.ferc.gov/news-events/news/commissioner-changs-dissent-pjms-rri-er25-712.

[74] U.S. DOE, "Transmission Interconnection Roadmap." Apr. 2024. [Online]. Available: https://www.energy.gov/sites/default/files/2024-04/i2X%20Transmission%20Interconnection%20Roadmap.pdf.

[75] ARUP, "What is data center design?," November, 2025, [Online]. Available: https://www.arup.com/en-us/insights/what-is-data-center-design.

[76] U. S. D. O. E. FEMP, "Best Practices Guide for Energy-Efficient Data Center Design." Jul. 2024. [Online]. Available: https://www.energy.gov/sites/default/files/2024-07/best-practice-guide-data-center-design_0.pdf.

[77] Turner & Townsend, "Pressures on delivery due to rising demand." [Online]. Available: https://reports.turnerandtownsend.com/dcci-2023/industry-challenges.

[78] Schneider Electric, "Future-ready: How modular data centers and BIM meet growing market demands." Jun. 20, 2024. [Online]. Available: https://blog.se.com/datacenter/2024/06/20/how-modular-data-centers-bim-growing-market-demands/.

[79] CBRE, "Global Data Center Trends 2026." Jun. 17, 2026. [Online]. Available: https://www.cbre.com/insights/reports/global-data-center-trends-2026.

[80] NLR, "The Role of the U.S. Electric Distribution System in Serving Data Center and Other Large Loads." 2026. [Online]. Available: https://research-hub.nlr.gov/en/publications/the-role-of-the-us-electric-distribution-system-in-serving-data-c.

[81] G. Dunn, "When Data Center Developers Have Options, State Regulatory Treatment Is Key to Success." Mar. 17, 2025. [Online]. Available: https://www.gibsondunn.com/when-data-center-developers-have-options-state-regulatory-treatment-is-key-to-success.

[82] World Resource Institute, "From Energy Use to Air Quality, the Many Ways Data Centers Affect US Communities." Feb. 17, 2026. [Online]. Available: https://www.wri.org/insights/us-data-center-growth-impacts.

[83] JLARC, "Virginia Data Center Study." Dec. 2024. [Online]. Available: https://jlarc.virginia.gov/pdfs/presentations/JLARC%20Virginia%20Data%20Center%20Study_FINAL_12-09-2024.pdf.

[84] EPRI, "Partnering for Progress: Understanding Community-Centered Data Center Development." Jun. 05, 2026. [Online]. Available: https://community-centered.epri.com/executive-summary.html.

[85] N. C. State Legislatures, “Which States Are Banning Data Centers?” Jul. 01, 2026. [Online]. Available: https://www.ncsl.org/fiscal/which-states-are-banning-data-centers.

[86] U.S. DOE, “Industrial Technology Validation Program.” Jan. 08, 2026. [Online]. Available: https://betterbuildingssolutioncenter.energy.gov/industrial-technology-validation-program/about.

New Buildings Institute (NBI) provides trusted, independent expertise to create buildings that are better for people, communities, and the planet. Our work focuses on increasing energy efficiency and reducing carbon emissions and improving indoor air quality in the built environment. From driving leading-edge design, curating innovative technologies, creating community-driven public policy, and supporting cutting-edge market solutions, our work lowers energy costs, delivers improved health and safety, and increases resilience to improve lives. To learn more about NBI and our work, visit newbuildings.org.