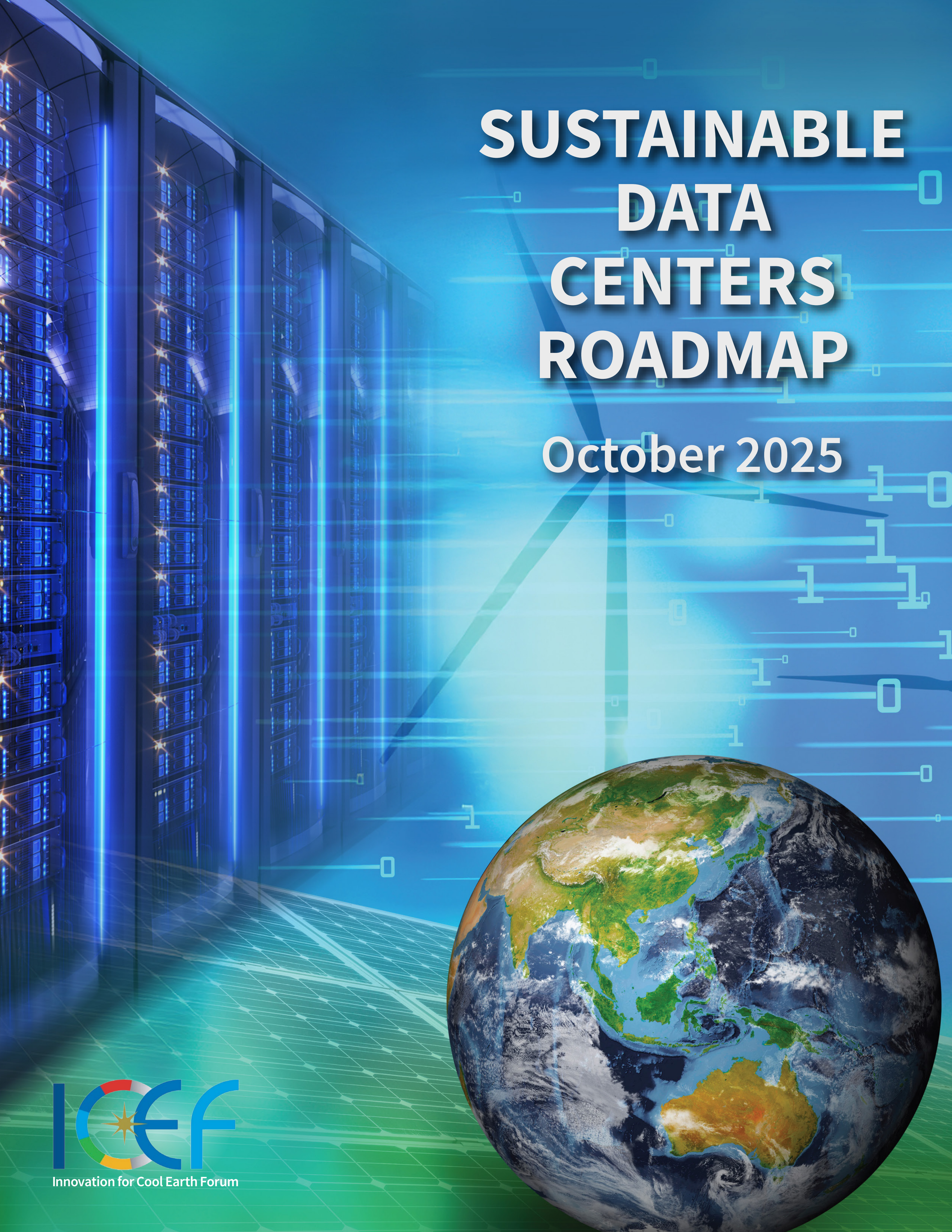


SUSTAINABLE DATA CENTERS ROADMAP

October 2025

Innovation for Cool Earth Forum

PRAISE FOR THE ICEF ROADMAPS

The ICEF roadmaps "provide essential information on different aspects of our energy systems and how they might change over time. I recommend them for anyone interested in this challenging topic."
Vaclav Smil — energy historian and Distinguished Professor, University of Manitoba

“The ICEF Roadmaps provide important research on a wide range of technologies for helping achieve net zero emissions. They are an important resource for anyone working on these issues.”
Hoesung Lee — former Chair, Intergovernmental Panel on Climate Change

"The ICEF Roadmaps…are an excellent resource for researchers and practitioners who cross disciplinary boundaries to develop transformative solutions for climate change.”
Alissa Park — Dean, UCLA School of Engineering

“Climate challenges are vast, and AI can accelerate solutions across domains. The *ICEF Artificial Intelligence for Climate Change Roadmap (Second Edition)* delivers an unprecedented, rigorous catalog of AI applications, both clarifying today’s state of play and preparing us for the opportunities and demands of tomorrow.”
Nicole Iseppi — Director of Energy Innovation, Bezos Earth Fund

“The application of AI solutions to climate mitigation is a crucial new front in the battle against climate change. Through our work across this ecosystem, we have found no better researched or more comprehensive framework in this critical area than the *ICEF Artificial Intelligence for Climate Change Roadmap (Second Edition)*”
Uday Khemka — Chairman, The Green Artificial Intelligence Learning Network

"The ICEF Roadmaps have covered topics vital to fighting climate change, including CO_2 utilization, direct air capture and carbon mineralization. I urge anyone working on these topics to study these roadmaps, which have been foundational for our work."
Larry Linden — Trustee, Advocates for Climate Innovation

PREFACE

Investment in data centers is booming. Several forecasts project that trillions of dollars will be spent on data centers and related infrastructure in the years ahead.[1-3] These investments are fueled by the explosive growth in attention to artificial intelligence (AI), as well as data centers' central role in much of the modern economy, including in e-commerce, email systems, video streaming and more.

This surging investment is raising concerns about data centers' energy and environmental impacts. Data centers use enormous amounts of electricity, creating challenges for electric grids in many regions. That electricity can and often does produce greenhouse gases and local air pollutants. Data centers also use water, result in electronic waste and often change land use patterns. These impacts have led to growing local opposition to new data centers in many countries.

At the same time, data centers' electricity demand can help accelerate adoption of clean energy. Data center owners and operators are the world's largest purchasers of solar and wind power. Several hyperscale data center owners are investing in innovative low-carbon power technologies in the hopes of bringing these technologies to market more quickly. Work is underway to turn flexible load at some data centers into grid assets.

This year's ICEF Roadmap explores these topics. In this Roadmap, a team of 12 coauthors examine data centers' energy use, strategies for improving data centers' energy efficiency, greenhouse gas emissions from data centers, strategies for using data centers to accelerate deployment of low-carbon power, data centers' water use and government data center policies around the world, as well as related topics. The Roadmap offers specific, actionable recommendations in each chapter. Our goal is to provide a useful resource for experts and non-experts alike.

This Roadmap builds on the body of literature produced annually in connection with the ICEF conference. Previous ICEF Roadmaps have addressed the following topics:

- Artificial Intelligence for Climate Change Mitigation (Second Edition) (2024)
- Artificial Intelligence for Climate Change Mitigation (2023)
- Low-Carbon Ammonia (2022)
- Blue Carbon (2022)

- Carbon Mineralization (2021)
- Biomass Carbon Removal and Storage (BiCRS) (2020)
- Industrial Heat Decarbonization (2019)
- Direct Air Capture (2018)
- Carbon Dioxide Utilization (2017 and 2016)
- Energy Storage (2017)
- Zero Energy Buildings (2016)
- Solar and Storage (2015)

This Roadmap is a team effort. We are grateful to everyone who reviewed draft chapters and/or helped with research for this Roadmap in the past five months. Special thanks to Thomas Spencer, Jon Turnbull, Erica Thomas, Vivek Shastry, Gareth Jones, Anne-Amélie Campant, Nadia Maunsell and Jonathan Wang. Any mistakes are of course our own. We are especially grateful for the support provided by the ICEF Secretariat, the ICEF Steering Committee (including in particular its chair, Nobuo Tanaka), the New Energy and Industrial Technology Development Organization (NEDO), and our superb copy edit and design team (including in particular Dr. Kathryn Lindl, as well as Amy Jaick and James Walker at Carbon Direct).

The ICEF Innovation Roadmap Project aims to contribute to the global dialogue about solutions to the challenge of climate change. We welcome your thoughts, reactions and suggestions.

David Sandalow
Center on Global Energy Policy, Columbia University
Chair, ICEF Innovation Roadmap Project

1. Jesse Noffsinger, et al. The cost of compute: A $7 trillion race to scale data centers. 2025 Published by McKinsey & Company; Available from https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/the-cost-of-compute-a-7-trillion-dollar-race-to-scale-data-centers.

2. Vivian Lee, et al. Breaking Barriers to Data Center Growth. 2025 Published by Boston Consulting Group; Available from https://web-assets.bcg.com/pdf-src/prod-live/breaking-barriers-data-center-growth.pdf.

3. John Minnix. 255 Data Center Stats (September-2025) (Brightlio Blog). 2025 Published by Brightlio; Available from https://brightlio.com/data-center-stats/.

AUTHORS

David Sandalow
Center on Global Energy Policy, Columbia University
Chair, ICEF Innovation Roadmap Project

Colin McCormick
Georgetown University and Carbon Direct

Julio Friedmann
Carbon Direct

Alexis Abramson
Columbia Climate School

Eric Masanet
University of California, Santa Barbara

Roger Aines
Lawrence Livermore National Laboratory

Alp Kucukelbir
Columbia University

Varun Sivaram
Emerald AI and Council on Foreign Relations

Ayse Coskun
Emerald AI and Boston University

Swasti Jain
Emerald AI and Columbia University

Angela Yuan
University of Cambridge

Minjue Wu
Global Sino Institute

We are deeply grateful to the Ministry of Economy, Trade and Industry (METI) and New Energy and Industrial Technology Development Organization (NEDO), Japan for launching and supporting the ICEF Innovation Roadmap Project of which this is a part.

Cite report as: David Sandalow, Colin McCormick, Julio Friedmann, Alexis Abramson, Eric Masanet, Roger Aines, Alp Kucukelbir, Varun Sivaram, Ayse Coskun, Swasti Jain, Angela Yuan, Minjue Wu, Sustainable Data Centers Roadmap (ICEF Innovation Roadmap Project, October 2025). Available at https://www.icef.go.jp/roadmap/.

TABLE OF CONTENTS

FIVE MAIN MESSAGES

1. **With a data center construction boom underway globally, the months and years ahead will be a critical time for data center sustainability.** Many decisions with respect to the construction and operation of data centers will have lasting impacts on natural resources and the environment.

2. **The energy and environmental impacts of data centers vary dramatically depending on their siting, design, management and other factors.**

    - Well-located and well-managed data centers can help accelerate deployment of low-carbon power by serving as anchor customers for innovative clean energy technologies, de-risking investments in renewables projects and enabling grid flexibility.

    - Poorly-located and poorly-managed data centers can have significant negative energy and environmental impacts, including greenhouse gas emissions, local air pollution, and water stress in surrounding areas.

    - Owners and operators of next-generation data centers are accelerating the adoption of liquid cooling and intelligent thermal management systems, while exploring opportunities such as free cooling and heat reuse.

3. **Smart siting is key to reducing the energy, water and carbon emissions impacts of data centers.** Locations with the potential for additional low-carbon electricity and ample water resources are especially valuable. When commercial, infrastructure and strategic factors lead to data centers being sited in locations that are suboptimal from a sustainability standpoint, technology options and management practices can minimize adverse impacts.

4. **Data center water use is tiny globally in relation to other sectors but can be very significant locally.** When a data center draws significant power from water-intensive generation sources, such as coal or nuclear plants, the data center's indirect (off-site) water use often exceeds its on-site water use.

5. **Data concerning data centers' environmental impacts are poor, including in particular data concerning greenhouse gas emissions and water use.** Research institutions, standards bodies, NGOs, government agencies, data center operators and equipment manufacturers should converge on common metrics for reporting data center energy use, water use and emissions.

FINDINGS AND RECOMMENDATIONS

Finding 1

With a data center construction boom underway globally, the months and years ahead will be a critical time for data center sustainability. Many decisions with respect to the construction and operation of data centers will have lasting impacts on natural resources and the environment.

Recommendations

1-1. Data center owners and operators should integrate energy and environmental concerns centrally into the planning for new data centers and operations of existing data centers.

1-2. Governments should require disclosure of energy and environmental impacts in connection with data center construction and operations, using standardized metrics. (See Recommendation 11-1.)

1-3. Governments should require data centers to meet minimum energy and environmental standards. These standards should address topics such as operational energy efficiency, water use efficiency, low-carbon power and greenhouse gas emissions. (The standards could focus on data centers in particular or be part of broader standards applicable to multiple sectors.)

Finding 2

Attitudes about data centers vary greatly from place to place and within jurisdictions. Many governments are eager to attract new data centers, typically for a combination of economic and strategic reasons. However, local communities around the world are expressing growing concerns and, in some cases, strong opposition to new data centers.

Recommendations

2-1. Data center owners and operators should engage collaboratively with local communities throughout the life-cycle of a project, from site selection to post-construction operations. This engagement should include communication of both the expected benefits to the community from the data center and potential risks (including those related to grid strain, water resources and local air pollution).

2-2. Data center owners and operators should work collaboratively with local communities in areas near data centers to implement measures that protect residential quality of life.

Finding 3

The energy and environmental impacts of data centers vary dramatically depending on their siting, design, management and other factors.

Well-located and well-managed data centers can help accelerate deployment of low-carbon power by serving as anchor customers for innovative clean energy technologies, de-risking investments in renewables projects and enabling grid flexibility.

Poorly-located and poorly-managed data centers can have serious negative energy and environmental impacts, including greenhouse gas emissions, local air pollution and water stress in surrounding areas.

Recommendations

3-1. Data center owners and operators should make energy and environmental factors a priority in data center siting and management decisions. Governments should make *energy and environmental factors a key factor in data center permitting decisions.*

3-2. *Utilities, regulators and data center operators should support accessible impact analyses on grid, environment, water and electricity pricing to enable informed local community decisions.*

Finding 4

Smart siting is key to minimizing environmental impacts of data centers. Locations with the potential for additional low-carbon electricity and ample water resources are especially important. When economic and strategic factors lead to data centers being sited in locations that are suboptimal from a sustainability standpoint, a range of technology options and management practices can partly reduce adverse environmental impacts.

Recommendations

4-1. *Data center owners and operators should prioritize current and future availability of low-carbon electricity, freshwater and other resources in data center siting decisions.*

4-2. *Data center owners and operators should select technologies and adopt management practices that minimize adverse environmental impact.*

4-3. *Governments should fast-track approvals for well-located, well-designed and well-managed data centers. Priority factors should include availability of low-carbon power and ample water supplies, as well as the potential for power flexibility. Rating agencies and other stakeholders should recognize and credit the energy and environmental performance of such data centers.*

Finding 5

In 2024, data centers used roughly 1.5% of electricity globally. The figure was much higher in some countries and regions, including roughly 4-5% in the United States, 3% in the European Union, 22% in Ireland and 25% in Northern Virginia. Some new data centers in planning and under construction will use extraordinary amounts of electricity (thousands of megawatts and more), straining electric grids.

Finding 6

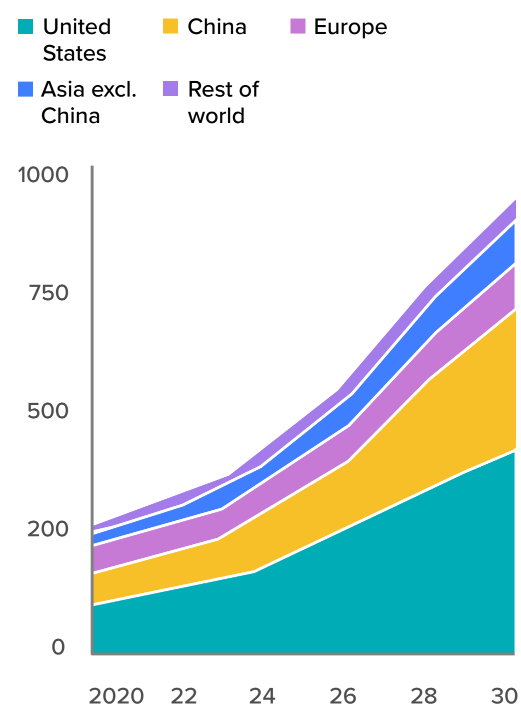
Scenarios of future power demand for data centers vary widely and depend on assumptions related to market growth, technological progress, adoption of efficiency measures and future policies. Recent International Energy Agency (IEA) scenarios suggest that global power demand for data centers could reach 1.8-3.4% of total power demand by 2030. In the United States (home to roughly 45% of global data center capacity), similar scenarios suggest that power demand for data centers could reach 6.7-12% of total power demand by 2028.

Recommendations

6-1. Data center operators should prioritize energy efficiency, including the use of advanced cooling, other highly-efficient equipment and highly-efficient algorithms whenever possible.

6-2. Governments, research institutions and philanthropies should support efforts to develop and disseminate best practices for modeling data center energy use, including inter-model comparisons and open models.

Global data center electricity consumption (TWh) by world region



Source: IEA, Energy and AI (April 2025)

Finding 7

In 2024, data centers were responsible for roughly 0.3% of global greenhouse gas emissions. This figure is likely to increase in the years ahead. The rate of increase will depend on several factors, including in particular the extent to which new data centers use low-carbon power.

Recommendations

7-1. Data center operators should include grid carbon intensity as a key siting consideration and site data centers in the lowest-emitting grid regions as much as possible.

7-2. Data center operators and utilities should work together to identify a mix of new low-carbon power generation technologies to add to the grid to meet rising data center load. This should include renewable power, nuclear power and retrofits of existing fossil fuel power plants with carbon capture and storage.

7-3. Data center operators and utilities should prioritize measures that enhance data center load flexibility when determining the amount of new generation required in connection with new data centers.

7-4. Governments should require disclosure of data centers' life-cycle greenhouse emissions, aligned with standardized protocols. Data center operators should transition to low-global-warming-potential refrigerants, procure carbon-free electricity and set targets for embodied carbon reduction in building materials and equipment. Industry and policymakers should collaborate on marginal emissions accounting to ensure real-world reductions.

Finding 8

The scale of data center growth creates opportunities to bring innovative clean energy technologies to market. Advanced market commitments and direct investments from large data center operators offer an important opportunity to de-risk first-of-a-kind technologies and accelerate deployment of low-carbon power.

Recommendations

8-1. Electricity regulators should enable advanced market commitments that allow multi-buyer participation, recognize hourly matching and credit verifiable flexible-load performance.

8-2. Data center owners and operators should seek additional opportunities for catalytic investments and advanced market commitments that accelerate production of clean power, clean goods and environmental services (such as water reclamation and CO_2 removal).

8-3. Data center owners and operators should pay close attention to immediate local environmental impacts at facilities where advance market commitments offer the potential for longer-term, global clean energy benefits.

Finding 9

Many data centers have traditionally been inflexible loads on electric grids. If this continues, it will create significant stress on many grids, causing reliability problems and increasing greenhouse gas emissions. However, many data centers have significant opportunities to manage their loads with more flexibility, which could improve reliability. Data center load flexibility can also reduce future emissions by minimizing or avoiding construction of new fossil-based generation capacity and enabling better renewables integration.

Recommendations

9-1. Utilities and independent power producers should deploy advanced control tools to accelerate interconnection and grid studies and to operate flexible portfolios. These tools include model-predictive control, enhanced forecasting and, where appropriate, artificial intelligence (AI).

9-2. Data centers should employ power flexibility technologies (including software and energy storage) to reduce impacts on the power grid, especially during periods of high demand and grid stress.

Finding 10

Data center water use is tiny globally in relation to other sectors but can be very significant locally.

Global agriculture water use is roughly 12,000 times greater than global data center water use. The process of producing one hamburger consumes roughly the same amount of water as 19,000 ChatGPT-3 queries.

However, large data centers can have significant impacts on water availability and quality in some regions. Operators can mitigate adverse water impacts from data centers through site selection, choice of cooling technologies, reclaiming water and other measures.

High-quality data on water use at data centers are extremely limited.

Recommendations

10-1. Data center owners and operators should share site-specific water use and consumption data proactively and invite third-party review. If necessary, governments should require disclosure of this information.

10-2. Before siting data centers, data center owners and operators should assess likely water impacts, including in particular by consulting with local stakeholders. In water-scarce regions, companies should consider several steps to reduce likely water impacts, including advanced cooling technologies that require little to no water, water reclamation and reuse, maximizing non-thermal power supplies, and buying building materials and chips with low-water footprints.

Finding 11

Data concerning data centers' environmental impacts are poor, including in particular data concerning greenhouse gas emissions and water use. Absence of data hampers understanding of the sector's impacts, modeling of its potential trajectories and good decision-making. Facility-level disclosures provide significantly more benefit than company-level disclosures. The IEA Energy and AI Observatory is an important repository for information related to data centers.

Recommendations

11-1. Research institutions, standards bodies, non-governmental organizations (NGOs), government agencies, data center operators and equipment manufacturers should converge on common metrics for disclosing data center energy use, water use and emissions.

11-2. Governments should establish dedicated units to monitor data center energy and water impacts. International collaborations, such as the IEA's Energy and AI Observatory and the Clean Energy Ministerial (CEM), should expand their work on data centers, supported by sustained funding.

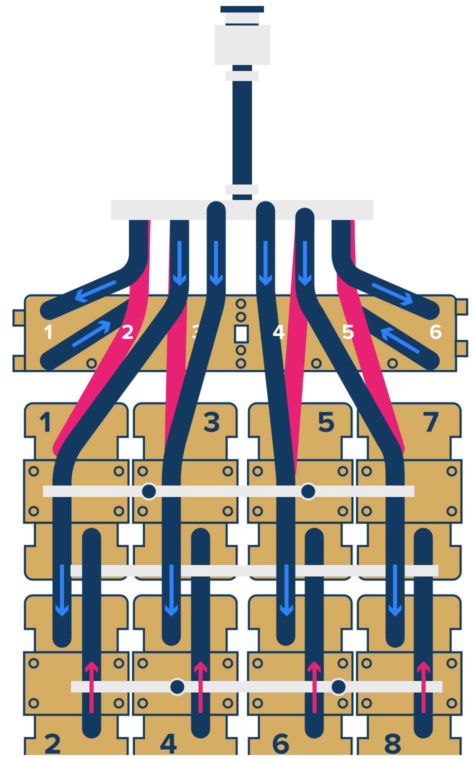
11-3. Research institutions, governments and NGOs should establish knowledge sharing platforms for developing and sharing best practice datasets, energy models and scenario frameworks to rapidly improve the rigor and accuracy of data center energy models.

Finding 12

Roughly 95% of data centers today predominantly use air-based cooling. However high-performance computing (HPC) and graphics processing unit (GPU)-heavy workloads are driving adoption of liquid cooling methods, as well as intelligent thermal management. These liquid cooling methods use substantially less water and often significantly less energy per kilowatt of information technology (IT) load than air-based cooling methods, which rely heavily on evaporation and high fan power.

Recommendations

Direct-to-chip cold plate liquid cooling for high-heat-density data centers.



12-1. Data center developers and operators should (1) select locations that enable use of free cooling, heat reuse or access to non-potable water and renewable energy, (2) install climate-appropriate cooling systems and (3) establish facility-level energy and water performance targets and publish sustainability metrics annually.

12-2. National policymakers should establish the market conditions and regulatory frameworks necessary for broad adoption of energy- and water-efficient cooling technologies.

12-3. Standards organizations, such as the American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE), the International Standards Organization (ISO) and the Open Compute Project (OCP), should establish uniform testing protocols and certification pathways to validate performance of new data center cooling technologies—especially liquid cooling, rear-door heat exchangers and high-efficiency refrigerants.

Finding 13

For data center waste heat to be useful, data centers must be located near heat hosts, such as district heating systems. Opportunities for beneficial use of data center waste heat are greatest in cold climates. Other than in those regions, data center waste heat is unlikely to be used beneficially at a significant scale.

Recommendations

13-1. Data center operators should adopt high-temperature liquid-cooling systems—such as direct-chip or immersion cooling—that achieve exit temperatures of 45-70 °C, enabling effective heat reuse in applications, such as district heating.

13-2. National and subnational governments should require feasibility studies for heat reuse when permitting large new data centers and offer incentives to deploy heat reuse systems. National and subnational governments should consider 10-20% heat reuse mandates for new data centers.

13-3. Heat host industries (e.g., district heating utilities, hospitals, laundries, greenhouses and industrial processes) should actively engage with data-center operators to explore using waste heat for 24/7 applications.

Finding 14

Governments have a wide range of policies to address the energy and environmental impacts of data centers, including regulatory standards, fiscal incentives and disclosure requirements. Power usage effectiveness (PUE) is the most common metric used by governments in data center energy and environmental policies. PUE is an important but narrow metric that only measures one aspect of a data center's energy use. Other metrics, such as carbon usage effectiveness (CUE), water usage effectiveness (WUE) and energy reuse effectiveness (ERE), are important to fully reflect data centers' energy and environmental impacts. Economy-wide policies (not focused on data centers) play an important role in data centers' energy use and environmental impacts.

Recommendations

14-1. Governments should collect and share data on data centers' energy use and environmental impacts.

14-2. Governments should use a broad set of metrics when regulating data centers' energy use and environmental impacts, not just PUE.

14-3. When governments procure data center services, they should require vendors of data center services to disclose their energy use, water use and greenhouse gas emissions.

14-4. IEA Member governments should expand the IEA's Energy and AI Observatory, devoting additional resources to monitoring and reporting on data centers' energy use and environmental impacts, as well as policy trends with respect to data centers around the world. The Clean Energy Ministerial (CEM) should expand its work on data centers under its power sector and AI initiatives.

Finding 15

Data centers produce a small fraction of global e-waste, but the amount of such waste will increase as data center capacity grows in the years ahead. If not handled correctly, hazardous materials in e-waste can leach into soil and water and pose serious environmental and health risks. Recycling and reuse programs can help significantly in managing e-waste challenges. However, data privacy and security concerns, as well as a lack of standardization and transparency, lead to the destruction of many data center components that could be repurposed.

Recommendations

15-1. Governments should implement and harmonize standards for reuse, refurbishment and recycling of e-waste, including safe sanitization of data-bearing IT equipment, and should strengthen extended producer responsibility rules.

15-2. Data center operators should refurbish, resell or donate retired equipment through certified recyclers, reduce equipment turnover with preventative maintenance, and optimize infrastructure using virtualization, cloud computing or shared systems.

15-3 Manufacturers should design modular, repairable and recyclable equipment; provide spare-part support and clear recyclability labeling; and minimize use of hazardous materials.

EXECUTIVE SUMMARY

Chapter 1: Data Center Energy Use

The energy use of data centers can be roughly divided into two categories: information technology (IT) equipment and infrastructure equipment. IT equipment energy use refers to the electricity consumed by servers, storage arrays and/or network switches. Infrastructure energy use refers to energy consumed by non-IT equipment that provides space conditioning (such as cooling) and ensures reliable power for the data center. Ideally, the IT equipment uses most of the electricity consumed by a data center, as it performs the data center's main revenue-earning function. This concept is reflected in a key metric used to assess data center efficiency: power usage effectiveness (PUE)—the ratio of a data center's total energy usage to the energy used by its IT equipment.

Data center electricity demand has risen sharply in recent years due to increasing digitalization, greater connectivity and the artificial intelligence (AI) boom. According to recent estimates, global electricity demand from data centers doubled from around 200 TWh in 2017 to more than 400 TWh in 2024. Data centers' share of global electricity demand increased from 0.9% to 1.5% in the same period. Growth has been concentrated in the United States, China and Europe, which collectively accounted for 86% of the world's data center electricity demand in 2024.

Scenarios of future data center electricity demand vary widely based on assumptions related to market growth, technological progress, energy efficiency improvements and other factors. Recent International Energy Agency (IEA) scenarios suggest that data center electricity use could reach 670-1260 TWh globally (or 1.8-3.4% of total global electricity demand) by 2030. In some locations with high data center concentrations (e.g., Ireland and Northern Virginia), data centers' share of total power demand is much higher and is placing growing stress on regional electric grids.

Chapter 2: Data Center Energy Efficiency

Chapter 2.1: Information Technology (IT) Equipment

Computing efficiency has improved by a factor of 10 billion since 1946. Since around 2005, efficiency improvements have been doubling roughly every 2.3 years.

AI workloads are creating unprecedented energy demands that fundamentally challenge traditional data center designs. Graphics processing units (GPUs) and accelerated AI processors consume 400-1000+ watts (compared to traditional processors' 50-200 watts), while requiring sustained high-bandwidth memory and storage systems that cannot easily enter power-saving modes. Chip-level efficiency gains are increasingly offset by exponential demand growth from AI applications.

Multiple innovation pathways are emerging to address these challenges. Advanced chip packaging techniques are reducing power usage, silicon photonics are enabling power improvements in networking, and next-generation semiconductors are reaching near-perfect energy conversion efficiency. Companies are developing AI-optimized server designs, exploring power oversubscription strategies and investigating edge computing to reduce data center computational loads. Although efficiency improvement rates are slowing, these emerging technologies and operational strategies specifically tailored to AI workloads offer potential for renewed progress in reducing energy consumption.

Chapter 2.2: Software

Software efficiency depends on algorithmic design, where clever approaches can reduce computational requirements by orders of magnitude. Traditional algorithms have clear "correct" outputs that enable objective efficiency comparisons, but AI systems operate differently. Modern AI architectures are delivering substantial efficiency gains using specific techniques that reduce memory requirements and accelerate processing with minimal performance degradation.

Operational flexibility represents a powerful approach to reducing data center energy consumption and carbon emissions. Checkpoint and restart technology enables AI training workloads to pause, migrate between data centers and resume during periods of abundant renewable energy or in regions with cleaner electricity grids. However, AI inference workloads face stronger constraints due to latency requirements.

Emerging AI applications are creating new challenges. AI reasoning models allocate extra computational resources during inference but can outperform models hundreds of times larger, potentially reducing overall energy consumption despite higher per-query costs. Autonomous agents and multi-agent systems are driving dramatic increases in inference demand, consuming up to 100 times more compute than traditional chatbot interactions through continuous operation and complex reasoning chains. The rapid pace of AI development often prioritizes performance over energy optimization, while the lack of standardized benchmarks for measuring AI efficiency complicates efforts to balance computational capability with sustainable energy consumption.

Chapter 2.3: Cooling Technologies

Data centers' rising thermal loads are outpacing the capabilities of traditional air-based cooling. Air-based cooling technologies rely on circulating conditioned air to remove heat from server environments. Enhancements like hot aisle containment, evaporative cooling and air-side economization improve efficiency, but these systems struggle to manage rack power densities above ~30 kW. This has driven a transition toward more efficient and scalable cooling solutions, including direct-to-chip and immersion liquid cooling systems. These liquid-based systems offer significantly higher thermal efficiency by transferring heat through fluids with greater capacity than air. They can cool racks exceeding 100 kW and minimize water use.

Case studies from hyperscale operators including Meta, Microsoft and Amazon Web Services (AWS) illustrate diverse strategies—from precision evaporative systems and machine learning (ML)-optimized heating, ventilation and air conditioning (HVAC) to large-scale heat reuse via district heating networks. These operators are testing and deploying new approaches, including modular systems and AI-enabled controls that are enhancing resilience, adaptability and deployment speed.

Barriers to adopting advanced cooling technologies include lack of standardization, retrofit limitations, uncertain cost structures, regulatory hurdles, and undervalued water and carbon impacts. Overcoming these barriers will require a collaborative, multi-stakeholder approach.

Chapter 2.4: Heat Reuse

Most of the energy in a data center ultimately gets turned into heat. This heat can be used in beneficial ways if the data center's cooling system generates return heat with high enough temperatures. Modern liquid cooling systems offer 40-70 °C return heat, which is excellent for powering district heating systems in cold regions. Using data center waste heat for district heating systems requires collaboration between data center owners and operators and local stakeholders, such as municipal governments.

Other potential uses for data center waste heat include industrial drying, agriculture and direct air capture (DAC) of carbon dioxide (CO_2). These uses generally require heat of at least 70 °C and locations very close to data centers. They have not gone beyond pilot stages. Heat reuse by data centers has the greatest potential in cold regions.

Text Box: Data Center Energy Efficiency Metrics

By far the most widely-used metric for data center energy efficiency is “power usage effectiveness” or PUE. First introduced in 2007 by the Green Grid, PUE is the ratio of a data center’s total energy use to the energy use of its IT equipment. However, PUE does not capture the efficiency of IT equipment or the efficiency of IT workloads in a data center. Other data center energy efficiency metrics include “IT power usage effectiveness” (ITUE)—the ratio of total energy into IT equipment to the energy use of IT compute components alone; “total power usage effectiveness” (TUE)—ITUE multiplied by PUE; “data center energy productivity” (DCeP)—useful work produced divided by total energy consumed by the data center; “server energy productivity” (SEP)—the energy consumption of a server in relation to the share of compute work the server is performing; and “IT Work Capacity (ITWC)—work per unit energy.

Chapter 3: Data Center Greenhouse Gas Emissions

Chapter 3.1: On-Site Greenhouse Gas Emissions (Scope 1)

Some equipment at data centers directly emits greenhouse gases (Scope 1 emissions). Diesel generators for backup power emit CO_2. Cooling equipment and fire suppression systems can both leak hydrofluorocarbons (HFCs).

While data concerning these emissions are limited, data center operators can mitigate emissions with a variety of strategies. Natural-gas-fired generators for backup power can marginally reduce emissions if used to replace diesel generators. Drop-in diesel fuel replacements, such as hydrotreated vegetable oil (HVO), biodiesel (FAME) and synthetic paraffinic fuels, can provide deeper emissions reductions, but supplies are not universally available. Battery energy storage systems can substantially reduce emissions if charged using low-carbon power. Carbon capture is not a good candidate for mitigating emissions from on-site backup power systems because these systems have an intermittent nature and relatively small capacity.

Replacing current high global-warming-potential (GWP) refrigerants and fire suppression agents (such as R-134a and HFC-227ea, respectively) with lower-GWP alternatives can significantly reduce emissions from these refrigeration and fire suppression systems.

Chapter 3.2: Power Supply Greenhouse Gas Emissions (Scope 2)

Modern data centers consume large amounts of electricity, which can lead to high greenhouse gas emissions from power generation (Scope 2 emissions). These emissions are the largest contribution to overall lifecycle greenhouse gas emissions of the data center industry. Scope 2 emissions for data centers globally were approximately 180 megatons of CO_2 ($MtCO_2$) in 2024 (roughly 0.5% of global CO_2 emissions). This could double or triple by 2030.

Strategies for reducing Scope 2 emissions from data centers include using highly efficient cooling systems and other equipment (especially helpful on high-emissions grids) and locating on low-carbon grids with the capacity for additional wind, solar, hydroelectric, geothermal and/or nuclear power. Data centers that draw power from low-emissions grids can have Scope 2 emissions almost 100x smaller than those that draw power from high-emissions grids, so siting decisions are especially important. Implementing load flexibility by using on-site clean generation and storage and actively managing the timing of power consumption from the grid can help significantly to reduce Scope 2 emissions on grids with large daily variations in emissions intensity (such as those experiencing a solar “duck curve”).

Many low-carbon power generation technologies are being considered as part of grid capacity expansion to meet new data center load, including variable renewables (wind and solar) firmed by storage, enhanced geothermal, gas-fired generation with carbon capture and storage (CCS), solid oxide fuel cells, nuclear fission (both conventional and small modular reactors) and nuclear fusion. The key factors that determine the most appropriate technologies are dispatchability, technology readiness, location flexibility and costs. In addition to these strategies to address Scope 2 location-based emissions, many data center operators seek to further reduce their Scope 2 market-based emissions by procuring renewable energy, typically through power purchase agreements (PPAs) and the retirement of electricity energy attribute certificates (EACs). Changes to the Greenhouse Gas Protocol Scope 2 guidance (currently under development) will likely restrict, but not eliminate, the ability to pursue this strategy.

Chapter 3.3: Embodied Greenhouse Gas Emissions (Scope 3)

Data center construction can produce significant embodied greenhouse gas emissions (Scope 3 emissions). In data centers that consume mostly very low-carbon power (e.g., from renewables or nuclear), embodied emissions can exceed 40% of a data center’s total greenhouse gas emissions and may dominate the lifetime greenhouse gas emissions footprint.

The largest share of embodied greenhouse gas emissions at a data center comes from IT hardware manufacturing, both for chips and complementary systems, such as memory. This is chiefly due to the fluorinated gases (F-gases) used in manufacturing. These F-gases can be extremely strong greenhouse gases—between 100 and 24,000 times more potent than CO_2. After IT hardware, steel and cement production have the largest share of data center embodied greenhouse gas emissions.

Builders and operators of data centers could significantly reduce embodied greenhouse gas emissions with existing technologies. Innovative approaches, both in design and within supply chains, have the potential to reduce embedded emissions further. Materials substitution, carbon capture, green hydrogen and extending the life of IT equipment all have roles to play.

Chapter 4: Accelerating Low-Carbon Power with Artificial Intelligence (AI) Data Centers

Data centers powering AI workloads are among the fastest-growing sources of new power demand, often concentrated in regions where grid capacity is already constrained. Without deliberate intervention, this growth could entrench reliance on fossil fuel generation and increase costs for consumers. With the right strategies, however, AI can serve as a catalyst for clean energy deployment, driving innovation, financing and system reliability.

Aligning AI's growth with low-carbon power requires a coordinated toolkit. Advanced market commitments can bring forward clean, dependable generation. Demand flexibility enables data centers to shift workloads across time and geography, thereby turning consumption into a valuable grid asset. In parallel, strategic siting steers new facilities toward renewable-rich regions, which eases dependence on fossil capacity. At the same time, AI can optimize grid operations, improve forecasting and accelerate discovery of next-generation energy technologies. Taken together, these approaches couple rapid AI expansion with a cleaner, more reliable power system.

Realizing this opportunity will require coordinated action by technology providers, utilities, regulators and governments. Clear pathways to scale clean power, stronger partnerships to reduce the cost of reliable low-carbon supply, and thoughtful siting of new AI campuses can ensure growth supports both grid stability and community needs. Through timely intervention, AI can be harnessed not only to advance digital innovation but also to reinforce affordability and sustainability on a global scale.

Chapter 5: Data Center Water Use

Water consumption by data centers is tiny globally in relation to other sectors but can be very significant locally. For example, global data centers consume less than 0.008% of the water consumed by agriculture, and the water consumed to produce one hamburger roughly equals the water needed for 19,000 ChatGPT-3 queries.

However, in water-scarce regions, data centers can create significant pressure on water resources. Data centers are frequently clustered in dry areas due to land and energy availability. Two-thirds of data centers built or in development in the United States since 2022 are in areas with high water stress. Data centers can also degrade water quality if managed poorly.

Data center water use can be divided into three categories: on-site water use (Scope 1), water used off-site by power plants providing electricity to a data center (Scope 2) and embodied emissions in the data center's supply chain (Scope 3). Scope 1 water use is chiefly for evaporative cooling. As cooling technologies improve, data centers' Scope 1 water use will decrease. Data centers' Scope 2 water use can be significant, especially when power is drawn from grids with heavy coal or nuclear power. Data centers' Scope 3 water use comes mainly from the water used in producing concrete, steel and chips. Data on data center water use are poor—improving data quality and access should be a priority for companies and governments.

Text Box: Electronic Waste (E-Waste)

Data centers produce a small percentage of the total e-waste today, but data center e-waste will grow in the years ahead. The AI boom is projected to generate 1.2-5 million metric tons of e-waste cumulatively between 2020 and 2030. When improperly disposed of, the hazardous substances in e-waste pose environmental and health risks. Circular economy strategies, including reuse, refurbishment and material recovery, can offer substantial e-waste reductions. However, security concerns and a lack of standards, transparency and oversight (especially with international shipments) mean most data center e-waste is not recycled. Reducing data center e-waste requires public-private partnerships to establish global standards, producer responsibility laws and norms, and greater priority for circular design and responsible end-of-life management.

Chapter 6: Government Policy

Governments around the world are paying increasing attention to data centers' energy consumption and environmental impacts. The European Union and several EU countries have policies aligned with their net-zero greenhouse gas emissions goals. China, Japan and other Asian governments have energy efficiency standards and policies to encourage data centers to use renewable power. US federal policies change dramatically from administration to administration, with the current US administration emphasizing new data center construction as a high priority while deemphasizing environmental protection.

Standards for PUE are the most common policy tool. Water usage effectiveness (WUE) standards and disclosure obligations with respect to energy and/or water use are becoming more common as well. Some jurisdictions have imposed moratoria on new data center construction to address grid strain and other local impacts. Economy-wide policies not focused on data centers play an important role in data centers' energy use and environmental impacts.

Evaluating the impacts of data center policies can be challenging. Facility-level data are scarce; operators have strong incentives to improve energy efficiency whether or not policies require it; voluntary initiatives from leading firms can make it difficult to determine whether improvements in energy and environmental performance are the result of government policies; and rebound effects mean energy efficiency standards can lead to more compute rather than less energy consumption. Governments should prioritize robust data collection, adopt a broader and more sophisticated set of performance metrics beyond PUE, and facilitate the rapid buildout of clean power capacity to meet data centers' growing energy needs.

Text Box: Industry Initiatives

Many companies in the data center industry have made ambitious voluntary sustainability commitments. Hyperscalers, such as Amazon, Microsoft, Google and Meta, have pledged 100% renewable energy, net-zero carbon, water positivity and circular economy practices. Voluntary industry-driven initiatives, such as The Green Grid, the EU Code of Conduct, the Climate Neutral Data Centre Pact and the iMasons Climate Accord, have spearheaded use of standardized metrics, accelerated adoption of efficiency innovations and benchmarks, and expanded the focus to climate neutrality and embodied carbon.

Text Box: Local Opposition

Local opposition to new data centers is rising globally, driven by concerns about grid strain, electricity costs, water consumption, noise, land use and other factors. One report found that opposition has delayed or blocked more than $64 billion in data centers in the United States in the past two years. Data center projects have also been delayed or blocked in other countries, including Chile, Ireland, the United Kingdom, the Netherlands and Singapore. Local opposition often conflicts with national and regional policies that encourage data center development for economic growth and security reasons. To address community concerns, data center operators should engage local stakeholders throughout the project lifecycle, clearly communicating a project's potential benefits and risks and implementing measures to protect residential quality of life.

1 Data Center Energy Use

Eric Masanet

A. Background

i. What is a data center?

Data centers can be thought of as the brains of the internet. These facilities process, store and communicate digital information, providing myriad services including website hosting, email, online shopping, banking, stock trading, maps and navigation, gaming and much more. According to the American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE), a data center can be defined as any dedicated space with information technology (IT) equipment loads greater than 10 kW and a floor area with an IT power density greater than 20 W/ft^2.[a,1]

In recent years, the data center sector has expanded rapidly to provide an increasingly salient societal service: the computations that make artificial intelligence (AI) and its many applications possible.[2,3] Data centers range vastly in size from small rooms with a few racks of IT equipment requiring tens of kW to massive dedicated buildings with thousands of racks requiring many hundreds of MW (Figure 1-1).

a While most cryptocurrency mining centers meet this definition, they are outside the scope of this report (see Box 1-2).

While approaches vary, data centers are commonly categorized as follows[2]:

- **Internal data centers:** Run by organizations internally, for their own use. These can range in size from large standalone buildings to small rooms and closets located in existing buildings. Such data centers can also be referred to as enterprise data centers.

- **Retail data centers:** Include both co-location and wholesale data centers built by real estate companies. Co-location data centers host multiple tenants, each of whom leases dedicated space to install and operate their own IT equipment, with cooling, power and other services provided by the data center. A wholesale data center is leased in its entirety to a single tenant.

- **Cloud data centers:** Offer cloud services (i.e., hosted software, hardware, infrastructure and platforms) to consumers who, instead of building and maintaining these systems themselves, pay the cloud provider for remote access and usage.

- **Edge data centers:** Tend to be smaller data centers located close to consumers to reduce latency.

- **High-performance computing (HPC) data centers:** Generally dedicated to hosting supercomputers for scientific computing purposes.

- **Hyperscale data centers:** Built and operated by companies that deploy internet services and platforms at massive scales. Examples include Google, Amazon Web Services (AWS), Microsoft, Oracle, Meta and Apple.[4]

In reality, these categories can have significant overlap. For example, Meta is classified as a hyperscaler but operates its data centers in internal fashion.[5,6] Many cloud providers are also hyperscalers, such as Google Cloud, Microsoft Azure or AWS, and these companies can also operate using space leased at co-location data center facilities.[7] Spaces within co-location data centers can be operated as edge data centers by some clients. Still, such classifications are typically useful for understanding key energy and environmental attributes of data center operators, such as greenhouse gas emissions from power consumption or type(s) of cooling technologies associated with each category[8,9] (see Chapter 2.3 of this Roadmap).

Figure 1-1. *The many shapes and sizes of data centers: (a) Meta's hyperscale data center campus in Odense, Denmark, which contains more than 200,000 m2 of floor area*[10]*; (b) an example of a tenant-leased space in a co-location data center with secured access*[11]*; (c) a high-performance computing (HPC) data center occupying around 13,000 m*[12,13]*; and (d) a small data center room in an office building, occupying less than 100 m*[14]*.*

The total number of data centers globally is unknown. This uncertainty is largely attributable to internal/enterprise data centers that are operated mostly by non-tech companies and have little visibility outside of the organization. For example, recent estimates by Lei et al. (2024)[15] suggest there may be around 1.4 million small- and medium-sized internal data centers operating in the United States alone. However, such data centers are also expected to host a small share of the overall total number of data center IT racks compared to—typically much larger—retail, cloud and hyperscale data centers. For example, Lawrence Berkeley National Laboratory (LBNL) estimated that all types of internal data centers represented only around 15% of servers installed in the United States in 2023, which is currently the world's largest data center market.[2,3] Conversely, LNBL estimated that around 75% of all US servers were located in co-location, cloud and hyperscale data centers that same year. Therefore, when it comes to energy use, while internal data centers are still important (especially from the firm-level perspective), the retail, cloud and hyperscale markets will drive national and global data center energy use moving forward.

While estimates vary, more is known about the number of retail, cloud and hyperscale data centers globally. For example, Synergy Research Group (2025) estimated that the number of large data centers operated by hyperscale providers increased to 1136 at the end of 2024. More than half of these are located in the United States. Online trackers of co-location data centers put their current global numbers around 6600-9000 globally.[16-18] However, these numbers may change quickly, as the sector is expanding rapidly in response to the ongoing AI boom.[19] For example, Synergy Research Group (2025) expects to see 130-140 additional hyperscale data centers coming online each year for the next several years.

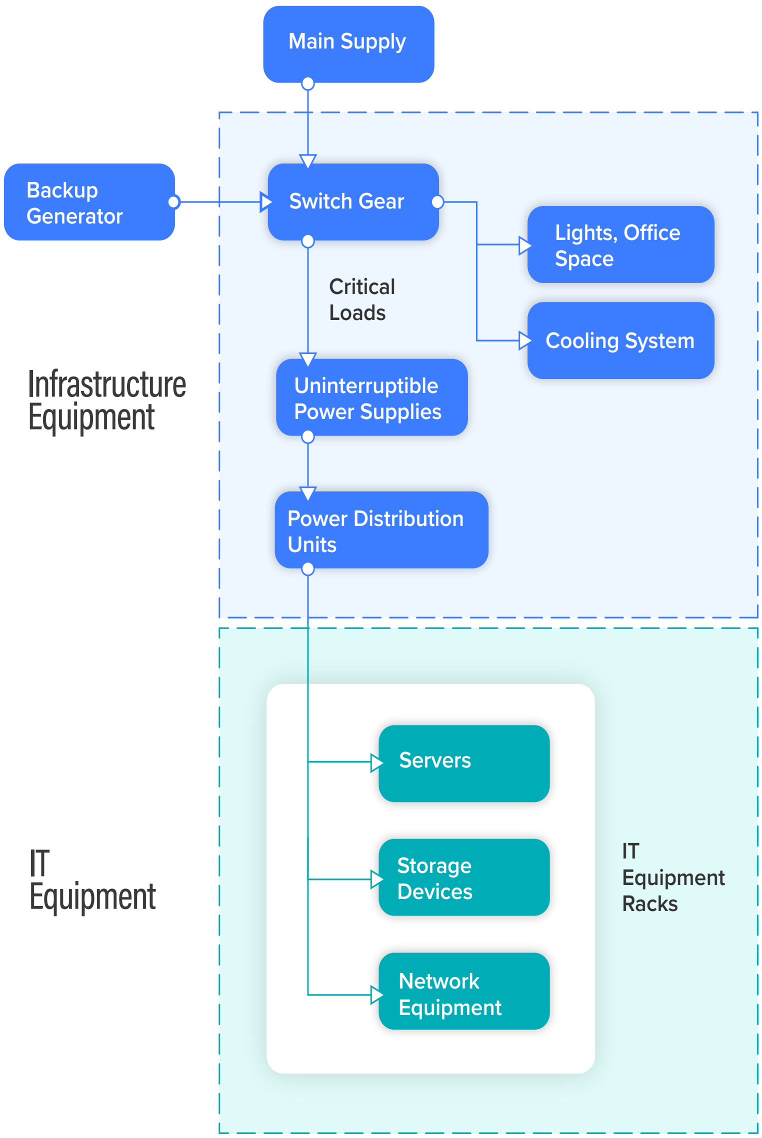
ii. How do data centers use energy?

The energy use of data centers can be roughly divided into two major categories: IT equipment and infrastructure equipment (Figure 1-2).[20] IT equipment energy use refers to the electricity consumed by data center IT equipment racks, which contain servers, storage arrays and/or network switches (Figure 1-3).[21] Servers provide the computations and memory necessary for software applications, such as websites, online ordering, and AI model training and inference. Storage arrays are used for storing files, such as documents, databases and streaming videos, and can comprise hard disk drives (HDDs), solid-state drives (SSDs) and even tape storage drives (for archival storage). Network switches enable data communications between racks and to and from the data center via the global internet.

Infrastructure energy use refers to energy consumed by non-IT equipment that provides space conditioning and ensures reliable power within the data center. Space conditioning involves cooling to keep the IT equipment from overheating and humidification or dehumidification to avoid electrostatic discharge, corrosion and condensation that would damage the IT equipment. Data centers use many different types of cooling equipment, the choice of which depends on local factors. Such factors include ambient climates, energy and water costs, building envelope, and rack power densities,[9,22] which are further discussed in Chapter 2.3 of this Roadmap.

Power provision equipment typically consists of transformers and switchgear, uninterruptible power supplies (UPSs), power distribution units (PDUs), and backup generators that provide power during grid outages. All infrastructure equipment runs on electricity except backup power generators, which require fuels such as diesel or natural gas for operation. (See Chapter 3.1 of this Roadmap.)

Figure 1-2. Schematic diagram of data center electricity use (derived from Brown et al., 2007[20])



Typical data center electrical components

Ideally, most electricity consumed by a data center would be used by the IT equipment itself, which performs the main revenue-earning function of the data center. When infrastructure uses a high fraction of the total electricity consumed, a data center is considered relatively inefficient. This concept is captured by a key metric used to assess data center efficiency: power usage effectiveness (PUE), which is defined as the ratio of a data center's total energy usage to the energy used by its IT equipment (see Box 1-2).

Figure 1-4 provides a breakdown of the most recent estimates of IT equipment electricity use versus infrastructure electricity use for all data centers in world regions[2,3,23,24] for which such estimates are available. Note that these types of data will typically be estimates due to widespread lack of empirical and reported data on data center energy use, which is a topic addressed later in this chapter. On average for these regions, infrastructure is estimated to account for 30-33% of electricity use, whereas IT equipment is estimated to account for 67-70% electricity use.

However, these proportions can differ substantially at individual data centers depending on their purpose, types of IT and cooling equipment, efficiency levels, climate and other factors, as shown in Figure 1-5.[3]

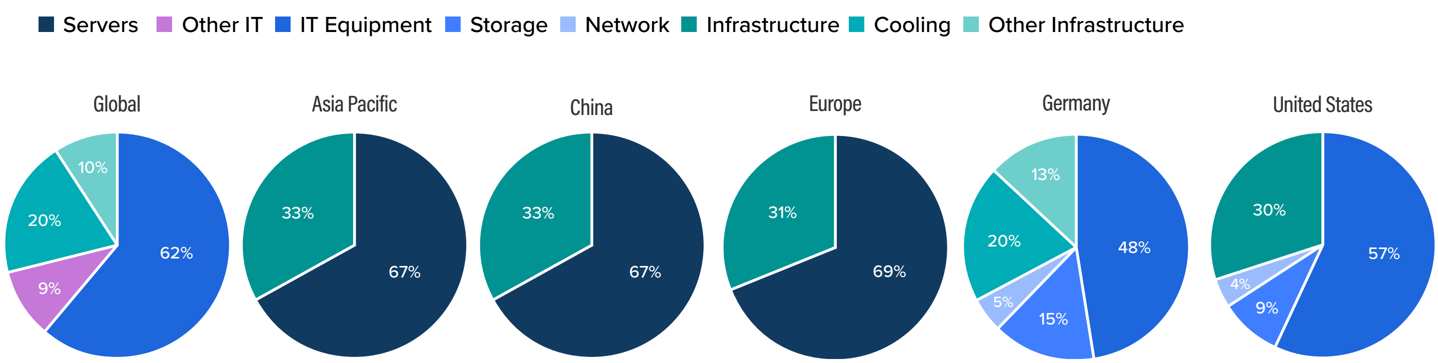
The share of overall data center electricity use attributable to infrastructure has steadily declined over the past two decades. This decline is primarily due to concerted efforts by data center operators to improve the energy efficiency of their cooling systems (as discussed in Chapter 2.3 of this Roadmap) and to adopt more efficient electrical equipment, such as high-efficiency UPS systems. For example, in 2007 LBNL estimated that infrastructure equipment consumed as much power as IT equipment in

Figure 1-3. A typical data center information technology (IT) equipment layout with individual IT racks arranged in rows. In this depiction, each row contains 13 racks.

US data centers, resulting in an estimated nationwide average PUE of 2.0. However, by 2023, thanks to the aforementioned efficiency gains, infrastructure equipment consumed less than half as much power as IT equipment, with an estimated nationwide average PUE around 1.43.[2,20]

Servers generally account for the vast majority of energy use on the IT equipment side (Figure 1-5). The energy use characteristics of each major type of IT and infrastructure equipment are discussed briefly below.

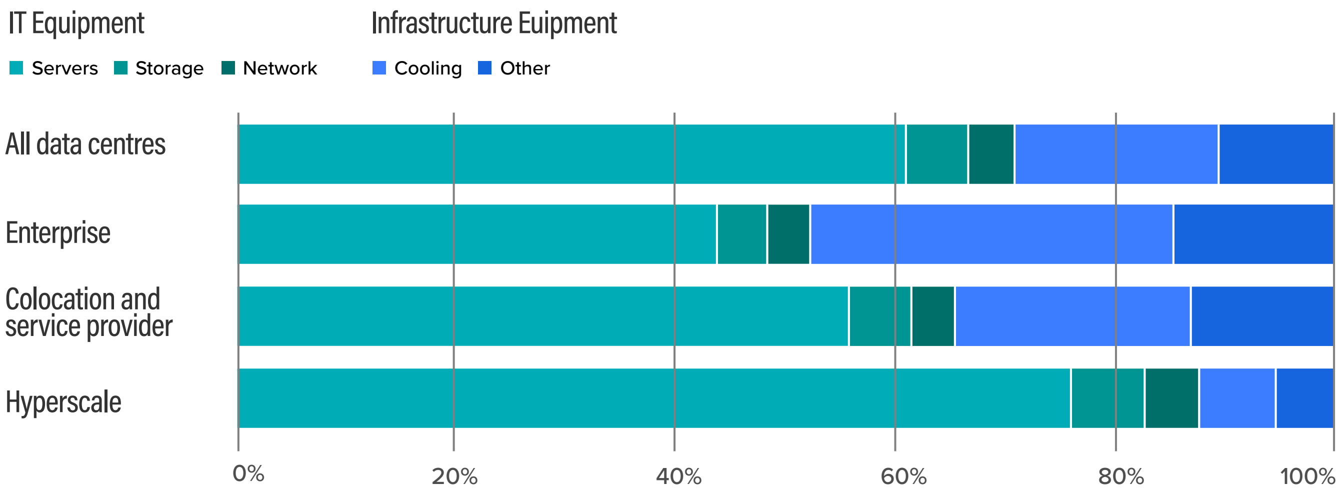
Figure 1-4. Estimated average IT equipment vs. infrastructure equipment energy use for different regions and countries.



ii(a). Servers

Servers are generally the single largest consumer of electricity in most data centers. While there are several different types of servers used in data centers, over the years, the average power requirements of servers has been growing steadily as has the total number of servers installed in data centers globally. For example, LBNL estimates that the total stock of servers in the United States doubled from 2014 (around 15 million) to 2023 (around 30 million).[2,25] This implies a compound annual stock growth rate (CAGR) of around 8%. While absolute stock numbers are not available at the global level, IEA estimated that the worldwide stock of servers grew at an even faster CAGR of 10% between 2015 and 2024.[3]

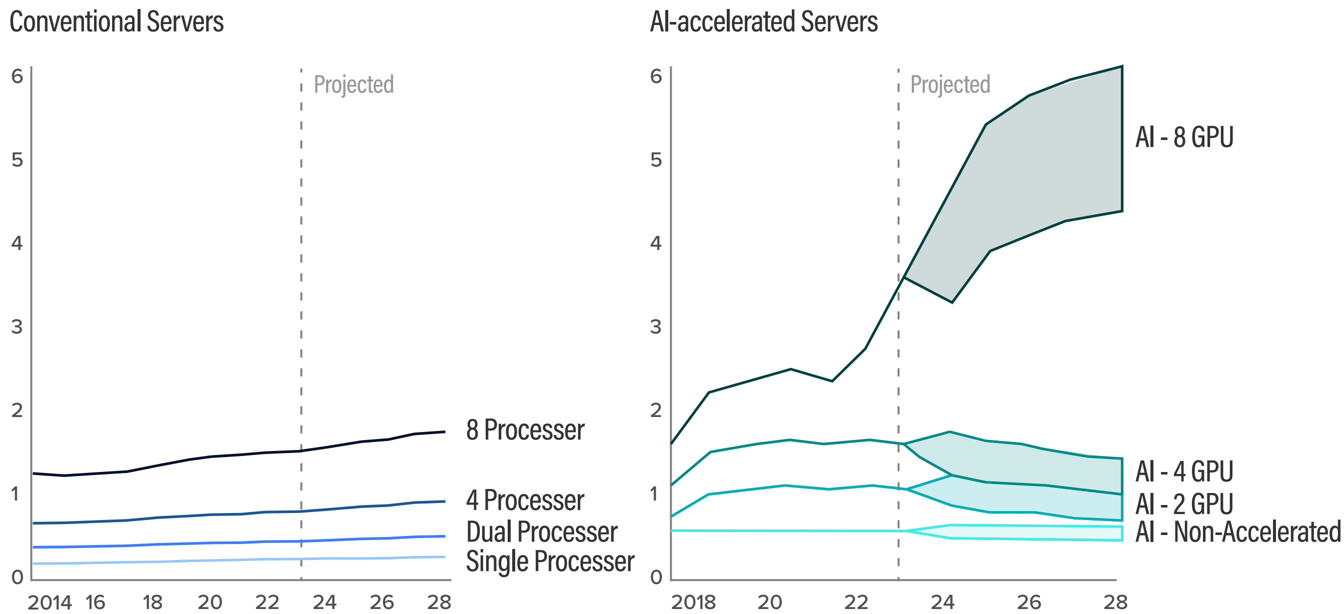
Figure 1-5. International Energy Agency (IEA) estimates of shares of electricity consumption by data center and equipment type, 2024[3]



The average power draw has increased over the past decade for all types of data center servers according to stockwide average estimates from LBNL (Figure 1-6).[2] While the data in Figure 1-6 are for the United States, global average power draws are expected to be similar given that servers are commodity products. Conventional servers are single, dual, 4-, and 8-processor servers (Figure 1-6a). These types of servers have historically been associated with traditional data center services, such as cloud computing, database hosting, websites, online commerce and so on. Despite steady energy efficiency gains at the component level, the average power requirements of conventional servers have increased due to a combination of higher wattage central processing units (CPUs), greater onboard memory and storage capacities, and higher utilization levels compared to the past. Accelerated servers (Figure 1-6b) are those that are additionally equipped with accelerated processors, such as graphic processing units (GPUs), which are required for machine learning (ML) and AI model training and inference applications (see Chapter 2.2 of this Roadmap). While accelerated servers are not new, their global stocks have expanded rapidly with the proliferation of AI-focused data centers.

Stocks of accelerated servers can consume far more power per unit compared to stocks of conventional servers (Figure 1-6). As a result, IEA estimates that while accelerated servers accounted for less than 5% of global servers in 2024, their operations accounted for around 20% of worldwide data center electricity demand that same year (accounting for associated cooling, storage and network energy use).[3] The shift to more accelerated servers will have major implications for server energy use in the coming years. For example, IEA estimates that around 8% of the global server stock will be accelerated servers by 2030, whereas LBNL estimates that 21-32% of the US server stock will be accelerated servers by 2028 (8-12 million accelerated servers out of 37 million total servers in 2028).[2]

Figure 1-6. Average per-server operational power draw (kW) of US stocks of different server types for conventional and AI-accelerated servers, historical and projected.. Source: Shehabi et al (2024).[2]



Finally, LBNL's projections for the future power draw of all server types suggest that the average power draws of conventional servers and of the largest AI accelerated servers will continue to rise for the next several years (Figure 1-6). These trends, coupled with expanding server stocks and shifts to AI-accelerated servers, suggest strong growth in global data center energy use in the near term (see Section C).

ii(b). Storage devices

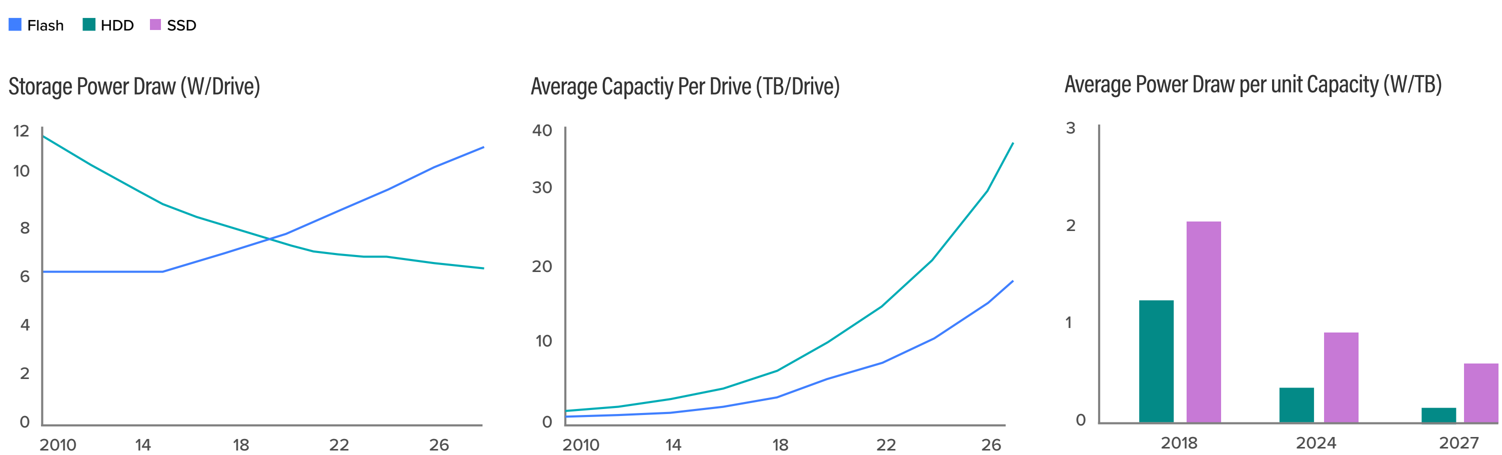
Some servers contain sufficient onboard storage to support their applications. Many others, however, require external storage arrays to store files and data that may be accessed either frequently or infrequently. The two major external storage array types are HDDs and SSDs, although tape storage is still utilized in some data centers for infrequently accessed files, as well as for backup purposes, due to its longevity.[26]

The choice between HDDs and SSDs is typically driven by application requirements, costs and performance. In general, HDDs are cheaper than SSDs and offer higher storage densities, but they have lower data transfer speeds, have higher cooling loads and can be less durable in shock-prone environments. As a very mature technology, HDD arrays can also be expanded and upgraded easily and cost-effectively. In contrast, SSDs are faster and more durable, offering performance advantages in critical applications, while being quieter and running cooler than HDDs. The primary disadvantage of SSDs is their higher cost.[27]

On average, storage is typically the second largest consumer of electricity in the IT rack, albeit at much smaller scales than servers (Figure 1-5).[2] Both LBNL (2024)[2] and IEA (2025)[3] estimate that, as of 2023, HDDs accounted for more than 80% of data center storage capacity with the remaining share attributable to SSDs. The estimated power draws and capacities of both types of storage are summarized in Figure 1-7a and 1-7b, respectively.

For both types of drives, capacities have been increasing steadily as storage companies continue to innovate. However, a different story emerges for HDDs compared to SSDs. The former are also decreasing in power draw per drive, even as capacities increase, leading to steady reductions in energy intensity (the average watts of power draw per terabyte of storage). The latter have seen increases in power draw per drive as capacity per drive has increased, indicating lower energy efficiency gains compared to HDDs. On a watts per terabyte basis, by 2027 HDDs are expected to exhibit substantially lower energy intensities than SSDs, which means they may be the more energy efficient choice for applications where speed and performance are not mission critical (Figure 1-7c).

Figure 1-7. *Trends in estimated power draw per drive, capacity per drive and power draw per terabyte for hard disk types in Figure 1-8 are reasonable for other world regions drives (HDDs) and solid-state drives (SSDs), 2010-2027, derived from panels (left) and (center). Source: Shehabi et al (2024).*[2]

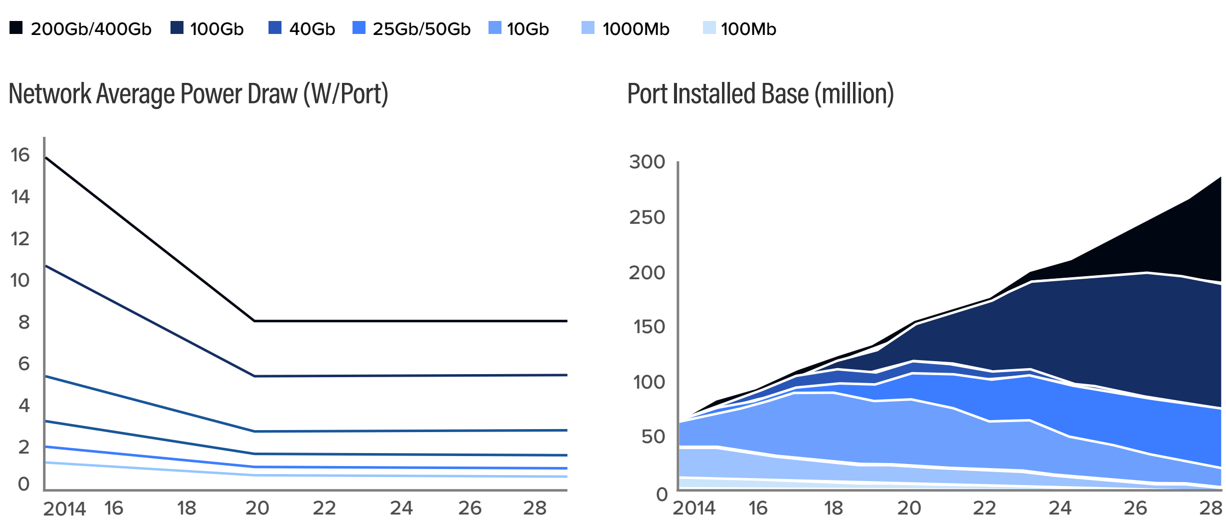


ii(c). Network switches

Network switches generally represent the smallest share of IT equipment energy use by a significant margin. For example, network switches comprised only around 6-7% of estimated national data center IT equipment electricity use in the United States (2023) and Germany (2024).[2,24] Similar to servers and storage, the overall electricity use of network switches is driven by the total stocks of different network switch types and the average per-unit power draws of each switch type. At a high level, network switches fall into two categories: ethernet switches and InfiniBand switches. Ethernet switches have been the dominant class with historical port speeds ranging from 0.1-100 GB/s and are primarily associated with conventional servers. InfiniBand switches are a recent class of ultra-high speed switches (with port speeds up to 400 GB/s presently) that are increasingly used in high-performance and AI computing applications.

Data from market firms such as International Data Corporation (IDC),[28] which are used by some analysts in data center energy analyses, provide network switch shipments in terms of network port numbers and speeds. Figure 1-8 summarizes estimates by LBNL (2024)[2] of the average power draw of different port types over time (left panel)—categorized by port speed—and the corresponding total stock of different ports in the United States (right panel). Similar data are not available for the rest of the world, but IEA (2025)[3] has estimated that both the power draws and the shares of different port types in Figure 1-8 are reasonable for other world regions.

Figure 1-8. Estimated average power draws of different types of network ports and shares of each port type in the United States, 2014-2028. Source: Shehabi et al (2024).[2]



The total electricity use of network switches is a function of the power per port and the total number of ports installed in data centers (as seen in Figure 1-8). While steady per-port power reductions were seen from 2014 to 2020, since that time LBNL's estimates have held constant. Meanwhile, the number of US ports has risen rapidly, with shares steadily shifting to faster port speeds over time as new switch innovations become available. In the United States, the net effect has been around a three-fold increase in network port electricity use from 2014 to 2023[2]; however, overall shares of network switch energy use remained small due to growth in IT electricity use overall.

As more AI-accelerated servers are deployed, the number of InfiniBand ports connecting AI clusters is also expected to increase rapidly. Due this growth, InfiniBand ports are expected to account for 45% of electricity use of all US network switches by 2028[2], underscoring how shifts to AI-accelerated servers will increase not only server electricity use, but network switch electricity use, as well. LBNL (2024)[2] estimates that InfiniBand switches require roughly 11 W/port, which is similar to the high per port power draws of the fastest ethernet switches (Figure 1-8, left panel).

ii(d). Power chain

The electricity use associated with the data center power chain generally arises from conversion and operating losses within transformers, switchgear, UPSs and PDUs. Collectively, these losses can amount to around 10% of total data center energy use on average (Figure 1-4). However, their scales can also vary significantly by data center type (Figure 1-5), depending on the equipment efficiency levels the data center has adopted. For example, Table 1-1 summarizes typical efficiency ranges of the UPS systems different types of US data centers employ, which generally represent the largest losses in the data center power chain. While the combined losses associated with electricity transformation and distribution equipment generally amount to less than 10% of total data center energy use, losses can be much less for the most efficient data centers.[8,20]

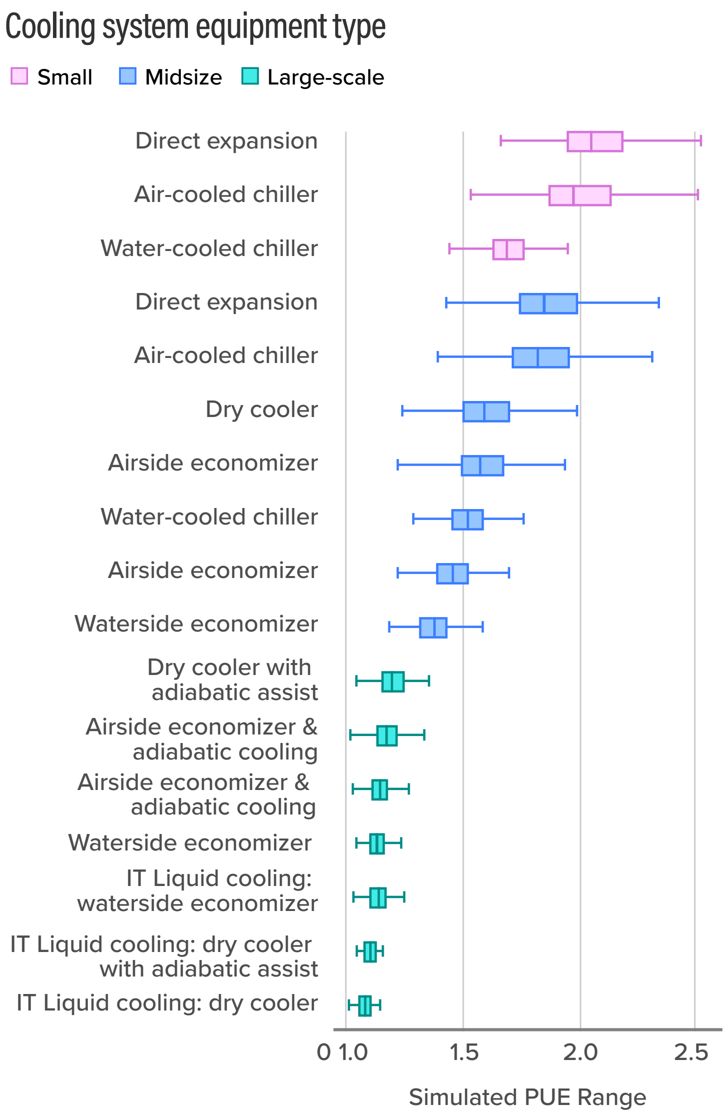
Table 1-1. Ranges of uninterruptable power supply (UPS) efficiences by US data center type. Shehabi et al (2024).[2]

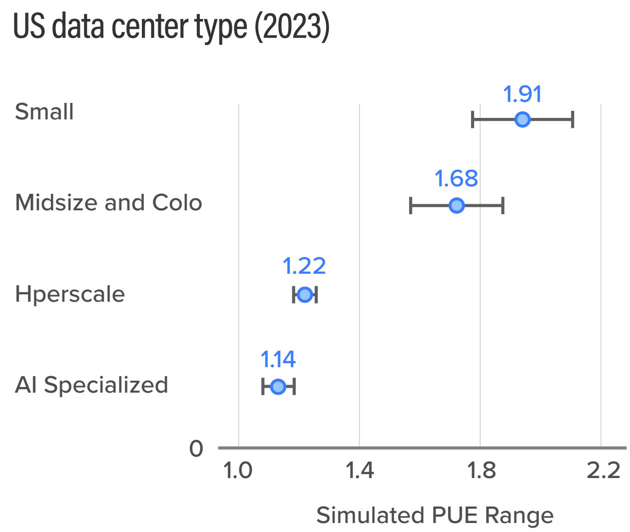
Data Center Type	UPS Efficiencies
Commercial edge Enterprise branch Small and medium businesses Telco Edge	77 - 85%
Comms service providers Internal data centers Colocation	80 - 94%
Hyperscale AI data centers	90 - 99%

ii(e). Cooling

Cooling systems generally account for the largest share of infrastructure energy use in most data centers (Figure 1-2). The energy use of cooling systems depends on many factors, including the types of cooling equipment employed, their efficiency levels, data center temperature set points, local climate and more, as discussed in Chapter 2.3 of this Roadmap. A common proxy for the efficiency of cooling systems is PUE (see Box 1-2), given that cooling systems typically consume the largest share of energy for infrastructure at a data center (Figures 1-4 and 1-5). The most efficient cooling systems generally deliver low PUE values, whereas the opposite is true for less efficient cooling systems. Figure 1-9[2] depicts the wide range of PUE values associated with different common cooling systems and approaches (left panel), estimated PUE ranges for different US data center types as of 2023 (upper right panel), and IEA's[3] estimated regional averages for 2024 (lower right panel). Higher PUEs are generally associated with today's small, midsized and co-location data centers, which typically rely on air-cooled racks with chiller-based systems or even direct-expansion (DX) cooling systems in smaller spaces. The listed technologies include liquid cooling systems, which are increasingly deployed to cool high power-density racks at AI data centers and can generally achieve very low PUE values when well designed. However, as AI data center campuses become increasingly large, with some announced data centers set to consume gigawatts of power, even low PUE values can still translate into very large electricity loads.[29]

Figure 1-9. Typical power usage effectiveness (PUE) ranges by cooling system equipment type, U.S. data center type in 2023 and world region in 2024.





World region (2024)

Power Usage Effectiveness (PUE)	
World	**1.41**
North America | US	**1.32**
Central and South America	**1.70**
Europe	**1.45**
Africa	**1.81**
Middle East	**1.92**
Asia Pacific | China	**1.5**

B. How Much Energy Do Data Centers Currently Use?

i. Estimation approaches

Unlike some other sectors, there are no comprehensive energy statistics available for all data centers at national or global levels. While some large operators report the annual electricity use of their data centers, others only report electricity use for their entire organizations (which can include office buildings, research labs, retail spaces and more), and some operators report no data at all.[29] Even when such data are collected from major operators by governments, as in Ireland,[30] the many small internal data centers operated by non-tech companies are often excluded. While the share of global electricity use attributable to internal data centers is small compared to hyperscale, co-location and AI data centers, as discussed earlier in this chapter, their electricity use is still substantial in an absolute sense. While the reporting situation may change with emerging policies that will require more disclosures (see Chapter 6 of this Roadmap), historically, the only way to quantify the energy use of all data centers at different spatial scales has been to use estimation methods.

In general, estimation methods fall into the following broad categories[2,31,32]:

- **Bottom-up calculations:** Estimate total stocks of different types of IT equipment by region and/or data center type, assume average power draws for each, and apply PUE assumptions (e.g., as discussed in Section A.2) to arrive at grand totals of energy use. Typically considered the "gold standard" of estimation methods due to its rigor and explanatory power, [29,33] its main drawback is that many (and often commercial) data are required for comprehensive estimates.

- **Top-down calculations:** Rely on energy consumption data that are collected by governments, reported by companies and/or are associated with data center populations (e.g., advertised critical IT capacities of co-location data centers) within a given region. While such aggregated data are partially available, this method risks omitting segments of the market that do not report (e.g., internal data centers and non-reporting operators).

- **Extrapolation-based calculations:** Develop estimates by applying growth rates to the estimates of previous years (which can be bottom-up, top-down or even extrapolative in nature), wherein proxies, such as internet traffic, investments or past energy growth, are typically used for growth rates. This method can introduce large uncertainties for multi-year extrapolations and lacks explanatory depth.

Each method has advantages and disadvantages, and each method is prone to uncertainties. While examples of all methods exist in the literature, the most rigorous and transparent analyses generally use bottom-up calculation approaches, such as recent US national and global analyses published by LBNL (2024)[2] and IEA (2025),[3] respectively. Kamiya and Coroama (2025)[31] offer a detailed discussion of estimation methods.

ii. Estimated current energy demand

Summaries of the most recent and authoritative estimates of data center energy demand are provided in Figures 1-10 and 1-11.[a]

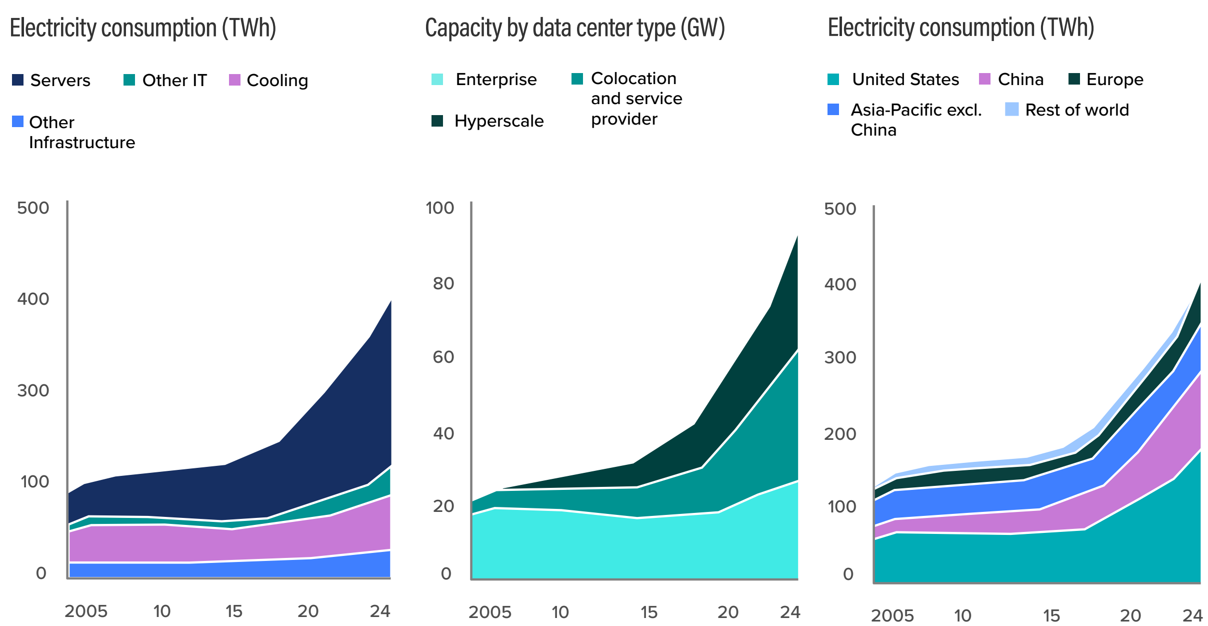
At the global level, IEA (2025)[3] recently estimated that data center energy demand increased roughly 3-fold over the past two decades, rising to 416 TWh in 2024, or around 1.5% of global electricity use that same year (Figure 1-10, left panel). However, much of that growth has occurred since around 2017, after a long period of relatively modest electricity demand growth.[33] The post-2017 rise was primarily due to growth in cloud computing, increasing online media consumption and, importantly, increasing deployments of AI computing globally. These trends are reflected by shifts in installed IT power capacity away from internal/enterprise data centers and toward hyperscale, cloud and co-location data centers (Figure 1-10, center). Over the same time period, substantial growth occurred in the world's three largest data center markets: the United States, Europe and China—with the latter experiencing the greatest relative growth due to large investments in AI data centers[34] (Figure 1-10, right panel).

Figure 1-11 summarizes best-available country- and region-level results by market segment and/or equipment type (when available) and for the most recent historical years reported. In terms of individual countries, the United States and China represent by far the greatest estimated shares of recent data center energy demand, at 8 and 5 times the scale, respectively, of Germany, which is the next largest country at around 20 TWh in 2024 (Figure 1-11). As a region, however, Europe is estimated by IEA (2025) to be the third largest data center energy user, at 68 TWh in 2024, followed by the rest of Asia Pacific (i.e., excluding China) at 48 TWh in 2024. Within Europe, data centers in Germany consume 2-4 times the energy of the next largest countries, which include France, the Netherlands and Ireland. Also notable in Figure 1-11 are the very small current scales of data center energy use in the large world regions of Africa, the Middle East, and Central and South America, whose collective energy use (around 5 TWh in 2024) is comparable to that of Ireland alone.

a Global estimates and those for all other countries besides the United States and Europe are from IEA (2025),[3] which relies primarily on bottom-up methods. Estimates for the United States are from LBNL (2024),[2] which also relies on bottom-up methods. Estimates for Europe (except Germany) are from Kamiya and Bertoldi (2024),[23] which aggregates data from different bottom-up, top-down and other types of studies. Estimates for Germany are from Murzakulova et al. (2025),[24] which uses bottom-up methods.

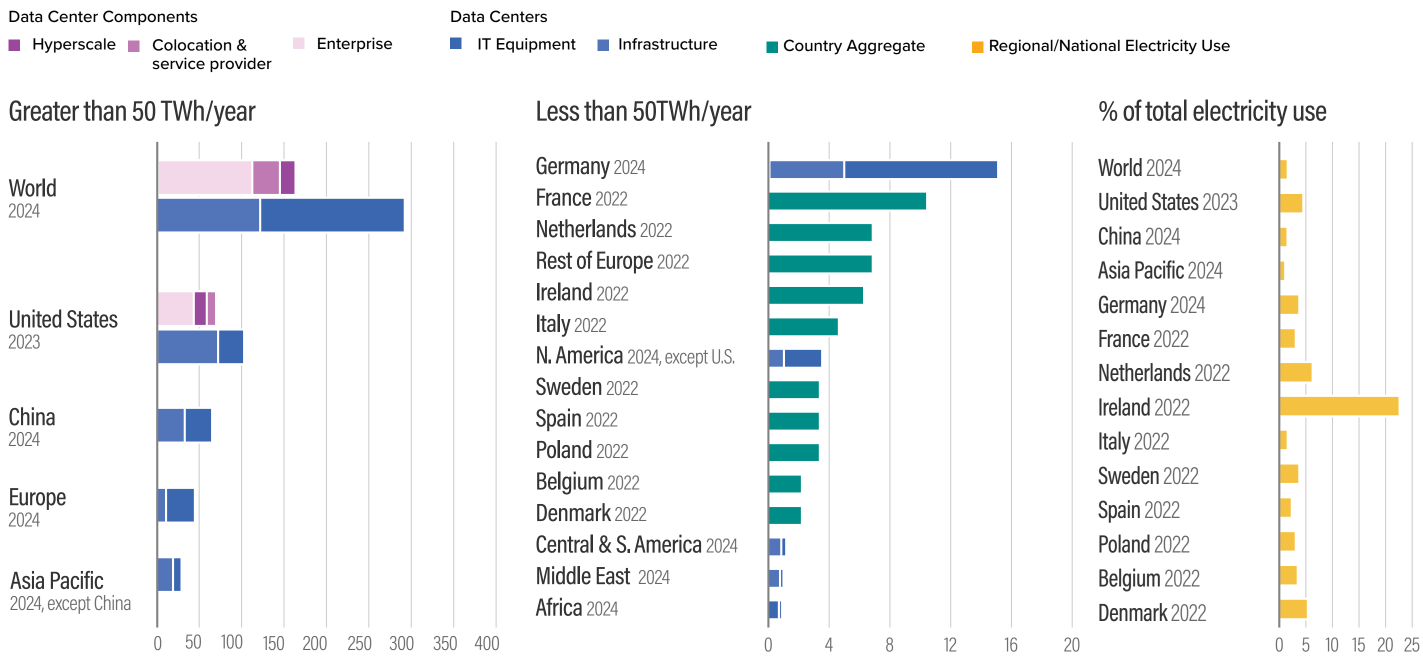
Expressed as a share of total national or regional energy use, a different perspective on data center energy use arises (Figure 1-11). Namely, while the United States and China account for the largest absolute amounts of total data center electricity use globally, data centers comprised only around 4% and 1% of their national electricity consumption in 2024, respectively.[3] Conversely, while Ireland's total data center electricity use is small from global perspectives, at the national level, data centers are a major source of electricity use, comprising 22% of Ireland's national electricity use in 2024.[35] Therefore, the importance of data center electricity use can also vary by country, even if country-level electricity use is a small contributor to global totals.

Figure 1-10. *Estimates of global historical (left) energy use by data center components, (middle) installed power capacity by data center type and (right) data center energy use by major world region, 2005-2024. Source: IEA (2025).*[3]



At subnational scales, data centers can affect electricity grids even more acutely. For example, in the City of Santa Clara, California-in the heart of the U.S. Silicon Valley-data centers account for around 60% of total electricity use, leading to electricity rate increases for all inhabitants to support grid infrastructure projects.[36] In the U.S. state of Virginia, which hosts an estimated 13% of reported data center capacity globally,[37] data center electricity use comprised around 25% of total electricity demand,[38] making data centers a critical focus of utility planning, rate structuring, energy policy, and community activism activities in recent years.

Figure 1-11. Most recent country- and region-level estimates of data center energy demand grouped by greater and less than 50TWh/year. Sources: IEA (2025),[3] Shehabi et al. (2024),[2] Kamiya and Bertoldi (2024),[23] and Murzakulova et al.



iii. What about carbon emissions?

Accurate estimates of the operational carbon emissions associated with the world's data centers are typically precluded by several factors. First, Scope 1 emissions data[3] that are specific to the onsite and/or backup generators used at data centers are rarely reported. Second, the specific locations of many of the world's data centers, especially internal/enterprise data centers, are unknown. Third, even when the specific locations of data centers are known, analysts must often use region- or country-level emissions factors for Scope 2 emissions.[4] Fourth, when some companies report Scope 2 emissions, they report only "market-based" emissions that reflect renewable energy credits and virtual power plant agreements rather than "location-based" emissions that reflect actual emissions from their local grids. Market-based emissions tend to be smaller than location-based emissions. (Scope 1, 2 and 3 emissions accounting is discussed in Chapter 3 of this Roadmap.)

However, recent estimates can offer a rough sense of scale on regional and global data center carbon emissions.

In 2023 in the United States, according to LBNL, Scope 2 emissions from data centers were 61 billion kg (61 Mt) of carbon dioxide equivalents (CO_2e). This figure is based on an estimate of 176 TWh of electricity (4.4% of total US electricity consumption) and assumed spatial concentrations of data centers within the country.[2] For context, total 2023 emissions from the US power sector amounted to 1454 billion kg (1454 Mt) CO_2e,[40] meaning data centers accounted for around 4.2%. These results suggest that US data centers are located in power grids that deliver, on average, slightly lower CO_2 emissions intensities than the national average.

At the global scale, IEA (2025)[b] estimated that data centers accounted for around 180 Mt of CO_2 emissions in 2024 from the consumption of electricity, which excluded emissions from backup power generation. As a share of total combustion-related emissions globally, this value translates to around 0.5%.

These results suggest that, while data center electricity demand is poised to grow rapidly, thus far it has made relatively small contributions to global CO_2 emissions compared to other energy-intensive sectors. However, this may change in the near future depending on the electric power sources that are employed moving forward to meet increasing data center electricity demand.

C. Where Is Data Center Energy Use Headed?

The rapid increase in data center electricity demand in the past several years (Figure 1-10), coupled with many recent announcements of future AI data center expansions, has led to the appearance of numerous recent studies that project future data center energy use at national and global levels. These studies have employed a wide range of different methods, data sources, analysis periods and assumptions in their projections, with no two studies being directly comparable. Due to these different approaches, there is wide variance in their results, as reported by Kamiya and Coroama (2025).[31]

However, the recent IEA (2025)[3] report provides robust insights on possible future trajectories of data center energy demand, as well as their drivers, the latter of which are summarized in Figure 1-12. Between 2024 and 2030, IEA estimates that the major drivers of data center energy demand growth will be major expansions of both conventional server stocks and energy-intensive AI-accelerated server stocks (see Figure 1-6), increased wattage of AI-accelerated servers, and expanded stocks of storage and network switch arrays to accompany expansions in compute capabilities. However, IEA also estimates that the energy demand growth attributable to the above

b Scope 1 emissions for a data center come from equipment on site, such as backup generators and refrigeration equipment.[39]

c Scope 2 emissions for a data center come from the generation of electricity purchased by the data center.[39]

drivers may be partially offset through reduction in data center PUEs associated with the shift to more hyperscale and AI data centers, which are increasingly adopting low-PUE liquid cooling methods (see Chapter 2.3 of this Roadmap).

As a result of these drivers, in its base case scenario, IEA estimates that global data center electricity demand could increase by a factor of 2.3 between 2024 and 2030, from 416 TWh to 946 TWh. The majority of this estimated growth is associated with large increases in the electricity demand of conventional and AI-accelerated servers (Figure 1-13; left panel) and plays out predominantly in the United States and China data center markets (center panel). In the base case, Scope 2 carbon emissions (right panel) are projected to rise rapidly alongside electricity demand growth, increasing from 180 Mt of CO_2 emissions in 2024 to around 320 MtCO_2 by 2030. Emissions in IEA's base case stop growing and instead begin to decline slightly after 2030, as lower-carbon electricity sources begin to penetrate. Future power grids are discussed in Chapter 4 of this Roadmap.

Figure 1-12. *International Energy Agency (IEA) estimates of the major drivers of global data center energy demand growth from 2024-2030. Source: IEA (2025).[3]*

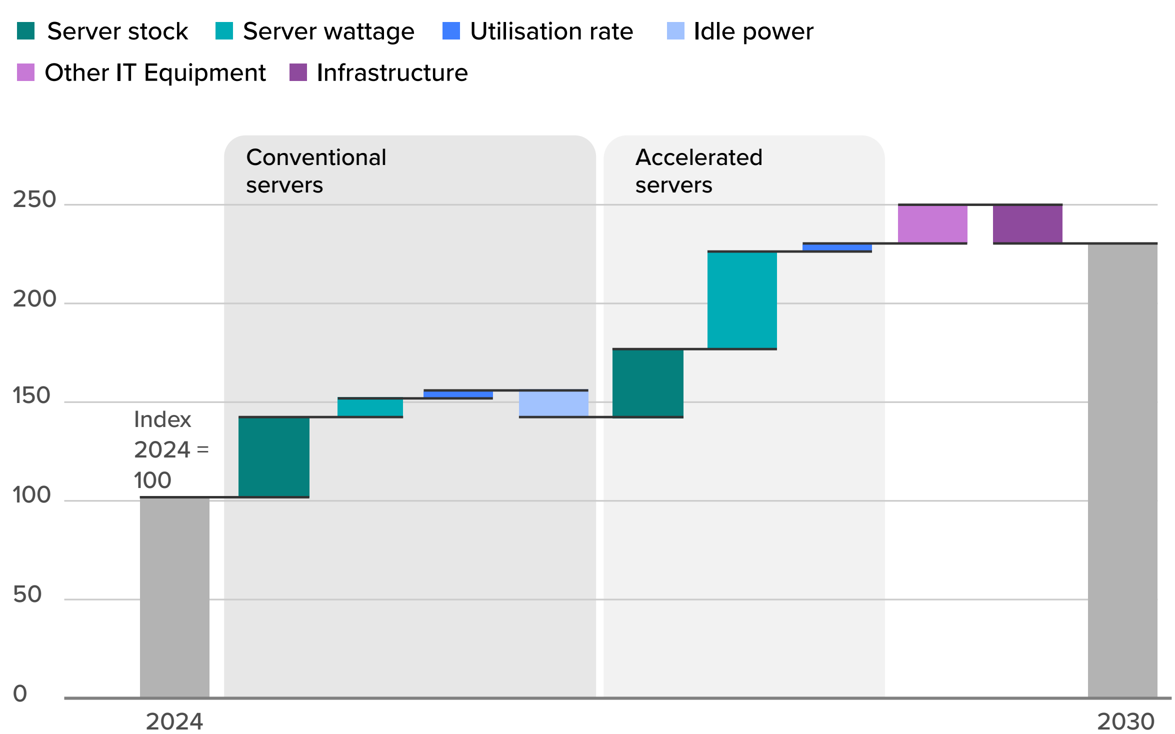
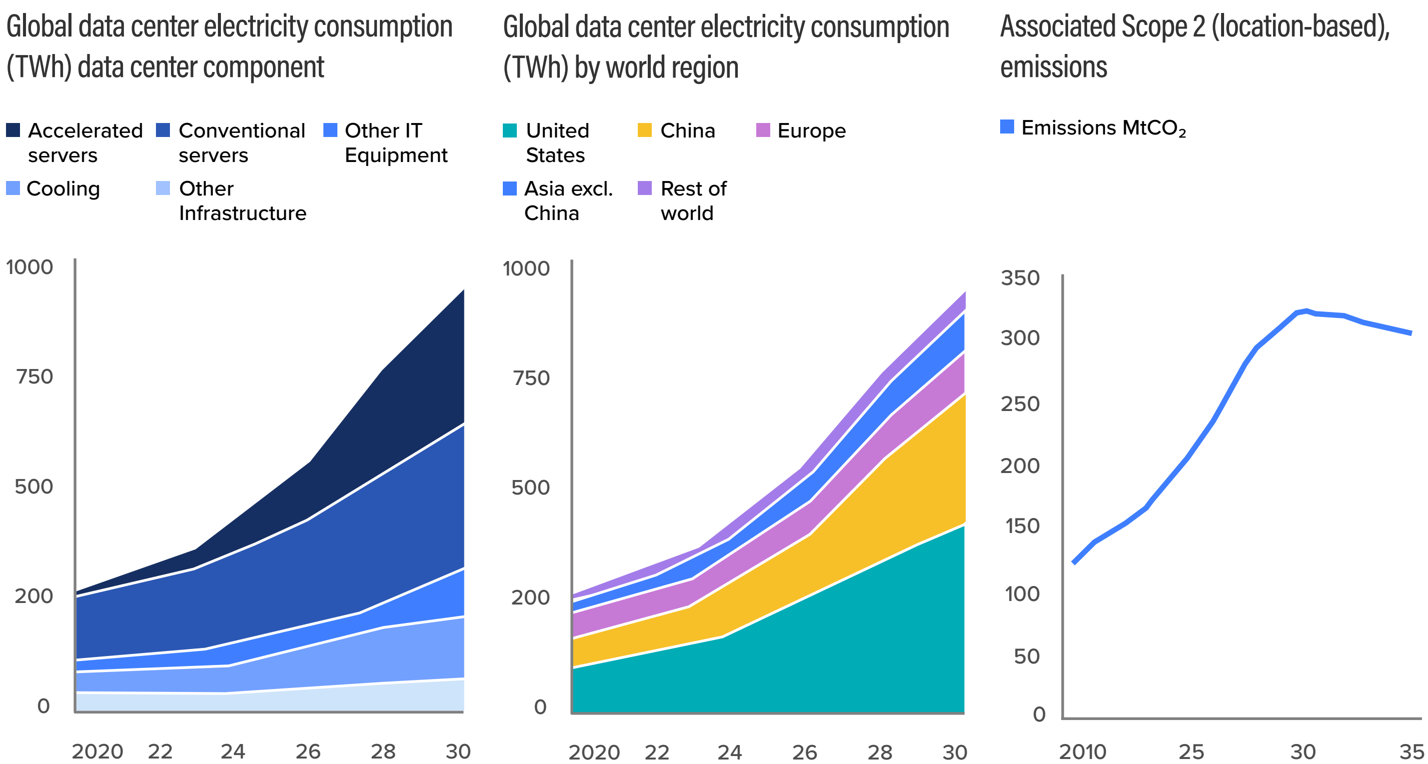


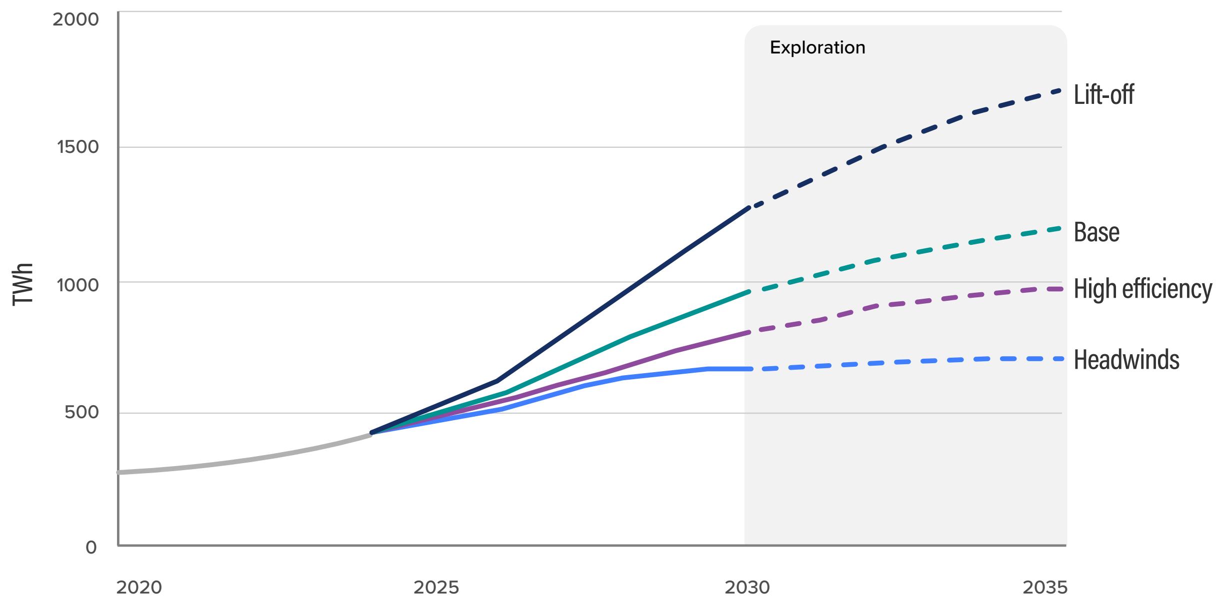
Figure 1-13. *International Energy Agency (IEA) (2025)3 Base Case scenario results for (left) global data center electricty demand by data center component, (center) global data center electricty demand by world region and (right) Associated Scope 2 (location-based) carbon emissions.*



In light of substantial estimation uncertainties, IEA (2025)[3] also considered bounding scenarios that correspond to "bullish" and "bearish" outlooks for growth of AI computing applications. In its "lift-off" scenario, IEA considered stronger AI adoption and increased global demand for digital services, leading to greater deployment of data center facilities compared to the base case. In its "headwinds" scenario, IEA considered plausible constraints to AI's growth, including difficulties financing further investments, strong local constraints on data center development and delayed data center construction due to electricity supply chain constraints. These two extremes result in vastly different outcomes, with data center electricity demand reaching around 1700 TWh by 2035 in the "lift off" scenario and leveling off at around 700 TWh by 2035 in the constrained "headwinds" scenario. These scenarios are summarized in Figure 1-14.

Scenario exercises such as these can shed light on possible future trajectories and what drives them, which can help inform early decisions that may help steer AI and data center energy demand toward the most sustainable outcomes. Such interventions include more efficient hardware and software, more efficient cooling systems, greater use of renewable electricity, reuse of waste heat, and more as discussed in later chapters of this Roadmap. As such, future analyses, especially those meant to guide policy decisions, should develop and implement scenario exercises.[47]

Figure 1-14. *International Energy Agency (IEA) global scenarios for the future of global data center electricity use (TWh), which depict a broad range of potential outcomes depending on market growth, efficiency gains and potential bottlenecks. Source: IEA (2025)*[2,3]



Box 1-1

Jevon's Paradox?

The recent rise in global data center electricity use has been predominantly driven by demand for AI computing. This trend is expected to continue through at least the end of the decade.[2,3] This rising electricity use has occurred despite steady and well-documented gains in the energy efficiency of AI hardware and software that have occurred in parallel (see, e.g., Patterson et al. 2023[41]) and despite growing adoption of the most energy-efficient cooling systems in many AI data centers.

The fact that total electricity use has risen despite large energy efficiency gains has led some analysts to invoke the concept of "Jevon's Paradox," which is attributed to economist William Stanley Jevons. Jevon's Paradox proposes that increasing the efficiency of a resource's use can lead to greater consumption of that resource over time, not less. The theory is that, as efficiency increases, the costs of using a resource are lowered, leading to more demand and expanded use of the resource. Scholars are beginning to research Jevon's Paradox in the context of AI computing. This assessment will require careful analysis of technology trends, potential rebound effects, and the extent to which AI substitutes for or complements demand for societal services.[42,43]

However, Jevon's Paradox has also been hotly debated in the energy efficiency community for years. Some leading scholars suggest that efficiency is a critical feature of sustainable growth, leading to enormous societal energy and cost savings over time while expanding access to energy services. They also argue that Jevon's Paradox ignores other growth factors, such as economic development, and that it paints energy efficiency as a negative force rather than a positive one.[44,45]

With respect to current increases in electricity demand for AI, some analysts have noted that AI technologies are in an early growth phase and many AI businesses are expanding principally due to competition for market leadership, not due to energy efficiency improvements.[46] Finally, there are many examples of energy efficiency improvements that have led to overall energy demand reductions, and despite rebound effects, such as the case of energy-efficient lighting technologies in the United States. Therefore, Jevon's Paradox is not the universal axiom that many view it to be.

While debates about Jevon's Paradox are not likely to be settled soon, most analysts agree that energy efficient hardware, software, cooling systems and data center operations will be critical for ensuring that AI data centers develop in the most sustainable ways possible. For example, in its "high efficiency" scenario, IEA estimates that energy efficiency could reduce the world's data centers' electricity consumption by nearly 20% by 2035 compared to its base case. Indeed, many gigawatts of future data center capacity builds have already been announced, not in response to historical energy efficiency gains, but rather for business, economic growth and national security reasons. Ensuring that these planned builds are as energy-efficient as possible will be critical for minimizing their environmental and social impacts.

Box 1-2

Cryptocurrency mining

Cryptocurrency mining refers to the use of specialized computing equipment that performs calculations—most commonly "proof of work (POW)"—to validate blocks of transactions in a blockchain ledger.[48] Operators of this equipment can earn rewards for each successful validation, which is typically paid in cryptocurrency, hence the term "cryptocurrency mining."

This specialized equipment, referred to as mining "rigs," can consume substantial amounts of electricity. For example, the global network of mining rigs for Bitcoin, the largest POW cryptocurrency in the world, was estimated to require around 200 TWh per year (as of August 28, 2025).[49] This is an amount that exceeds the total estimated electricity use of all US data centers in 2024 (183 TWh).[3]

Due to these high power demands, the majority of cryptocurrency mining occurs in dedicated facilities that can each contain many thousands of mining rigs. Just like the IT racks in conventional data centers, these mining rigs also generate large amounts of heat that must be removed. Cooling techniques range from simple facility ventilation fans to cutting-edge liquid cooling systems, depending on the operator.[2,50] Although data on the geographical distribution of cryptocurrency mining facilities are scarce, as of 2022, the majority of Bitcoin mining operations were estimated to be located in the United States, China, Kazakhstan, Canada, Russia and Germany.[51]

The highly specialized and physically distinct nature of cryptocurrency mining means that it is often regarded as its own sector, distinct from either conventional or AI data centers. For further information on the energy impacts of cryptocurrency mining, see Shehabi et al. (2024), IEA (2025), and CBECI (2025).

Besides POW, there are several other methods for validating cryptocurrency transactions, such as proof of authority (POA), proof of burn (POB), proof of capacity (POC) and proof of stake (POS).[52] Of these, POS—used by Ethereum and some other cryptocurrencies—is the most common. The energy requirements of POS cryptocurrencies are much lower than those of POW cryptocurrencies,[48] due to more efficient validation algorithms that can be run on standard computing equipment. CBECI (2025) estimates that the global Ethereum network consumes around 4.4 GWh of electricity per year, which is 45,000 times lower than the global Bitcoin network.

Cryptocurrency mining is outside the scope of this Roadmap.

D. Recommendations

Data center energy use has risen rapidly, and according to many projections, may continue to do so in the years ahead. However, our ability to understand and manage this potential growth depends on having models, datasets and scenarios that can accurately estimate where that growth may be headed, which factors may drive that growth and what interventions and improvements can bend the curve toward more sustainable outcomes.

To support such a vision, the following recommendations could improve the depth and quality of the knowledge base on data center energy use:

1. *Governments and regulatory agencies should* ***develop public data repositories on the energy use and characteristics of data centers*** *at national and sub-national scales, including via mandatory data collection initiatives from data center operators with strict data quality and measurement and verification protocols.*

2. *Data center operators should* ***improve reporting and transparency on the energy use, peak power demand, operating characteristics and other environmental attributes (e.g., water consumption) of data centers*** *to improve the empirical knowledge base for data center energy analysts.*

3. *Governments, philanthropies and research institutions should* ***organize and convene forums to establish best practice analysis methods and data sharing initiatives*** *to rapidly improve the state of science for estimating and projecting data center energy use.*

4. *Researchers should* ***conduct regular inter-model comparisons of data center energy models and scenarios*** *to understand model differences, identify potential improvements, and establish and coalesce on best practices.*

5. *Governments, research institutions and international organizations should* ***convene forums specifically aimed at developing and disseminating scenario narratives for exploring future growth pathways for AI and data centers****, including associated frameworks and analysis approaches; similar to the Shared Socioeconomic Pathways,[47] such scenarios would enable inclusion of data center futures in climate change mitigation scenarios.*

6. *Governments and research institutions should **develop and disseminate models and datasets of data center energy use in developing and emerging economies**, which have historically been overlooked by the research community but whose data center electricity use may grow in the future.*

7. *Research institutions should **reinforce the need for best practices in analyses adopted by policymakers**, given wide variance in results associated with low-quality studies.*[31]

8. *Governments and research institutions should **improve approaches to identifying existing and planned data centers**, such as through in-country sources and open shared location datasets such as the IEA AI Observatory. Improved understanding of the spatial patterns of data centers is important for assessing potential local impacts and proactively designing policies that avoid or minimize those impacts.*

9. *Governments and companies should **support research to better understand the CO_2 emissions of AI data center power sources**, focusing on more granular, grid-scale modeling of emissions. This includes closely tracking announced investments in cleaner power technologies, as well as tracking those investments when they ultimately come online, for more accurate forward-looking scenarios.*

E. References

1. American Society of Heating Refrigerating and Air-Conditioning Engineers (ASHRAE) Standing Standard Project Committee (SSPC). ANSI/ASHRAE 90.4-2019: Energy Standard for Data Centers (ISSN 1041-2336); ASHRAE, Atlanta, Georgia, https://webstore.ansi.org/standards/ashrae/ansiashrae902019 (2019).

2. Arman Shehabi, Alex Newkirk, Sarah J Smith, Alex Hubbard, Nuoa Lei, Md Abu Bakar Siddik, Billie Holecek, Jonathan Koomey, Eric Masanet & Dale Sartor. 2024 United States Data Center Energy Usage Report (Report #: LBNL-2001637); Lawrence Berkeley National Laboratory (LBNL), Berkeley, California, https://escholarship.org/uc/item/32d6m0d1 (2024).

3. International Energy Agency (IEA). Energy and AI; IEA, Paris, France, https://www.iea.org/reports/energy-and-ai (2025).

4. Synergy Research Group. Hyperscale Data Center Count Hits 1,136; Average Size Increases; US Accounts for 54% of Total Capacity; Synergy Research Group, Reno, Nevada, https://www.srgresearch.com/articles/hyperscale-data-center-count-hits-1136-average-size-increases-us-accounts-for-54-of-total-capacity (2025).

5. Mary Zhang. Top 250 Data Center Companies in the World as of 2024; Dgtl Infra LLC, New York, New York, https://dgtlinfra.com/top-data-center-companies/ (2024).

6. Synergy Research Group. Synergy Updates Its Ranking of the World's Top 20 Hyperscale Data Center Locations; Reno, Nevada, https://www.srgresearch.com/articles/synergy-updates-its-ranking-of-the-worlds-top-20-hyperscale-data-center-locations (2025).

7. Apple. 2024 Environmental Progress Report; Apple Inc., Cupertino, California, https://www.apple.com/environment/pdf/Apple_Environmental_Progress_Report_2024.pdf (2024).

8. Nuoa Lei & Eric Masanet. Statistical analysis for predicting location-specific data center PUE and its improvement potential. Energy 201, 117556 (2020). https://doi.org/10.1016/j.energy.2020.117556.

9. Nuoa Lei & Eric Masanet. Climate- and technology-specific PUE and WUE estimations for U.S. data centers using a hybrid statistical and thermodynamics-based approach. Resources, Conservation and Recycling 182, 106323 (2022). https://doi.org/10.1016/j.resconrec.2022.106323.

10. Dan Swinhoe. Meta/Facebook to expand Odense data center campus in Denmark; Data Center Dynamics (DCD), London, United Kingdom, https://www.datacenterdynamics.com/en/news/metafacebook-to-expand-odense-data-center-campus-in-denmark/ (2021).

11. Wikimedia Commons. Photos taken of the SDTPA deployment (File:Sdtpa wmf-8.jpg; Creative Commons Attribution-Share Alike 3.0 Unported license); San Francisco, California, https://commons.wikimedia.org/wiki/File:Sdtpa_wmf-8.jpg (2010).

12. Jon Weiner. Energy Department to Invest $16 Million in Computer Design of Materials; Lawrence Berkeley National Laboratory (LBNL), Berkeley, California, https://newscenter.lbl.gov/2016/08/16/energy-department-invest-16-million-computer-design-materials/ (2016).

13. Better Buildings Initiative & Lawrence Berkeley National Laboraotry. Retro-commissioning at LBNL's National Energy Research Scientific Computing Center; US Department of Energy (DOE), Berkeley, California, https://betterbuildingssolutioncenter.energy.gov/showcase-projects/retro-commissioning-lbnls-national-energy-research-scientific-computing-center (Accessed August 2025).

14. Bill Tschudi. The Other Half of the Problem - Server Closets and Small Server Rooms; Lawrence Berkeley National Laboratory (LBNL), Berkeley, California, https://datacenters.lbl.gov/sites/default/files/ServerclosetLBNL11-12-13.pdf (2013).

15. Nuoa Lei, Mohan Ganeshalingam, Eric Masanet, Sarah Smith & Arman Shehabi. Shedding light on U.S. small and midsize data centers: Exploring insights from the CBECS survey. Energy and Buildings339, 115734 (2025). https://doi.org/10.1016/j.enbuild.2025.115734.

16. Datacenters.com. Homepage: Find a Data Center: Search 7,074 Data Centers; Englewood, Colorado, https://www.datacenters.com/ (Accessed August 2025).

17. Data Center Map (DCM). Homepage: Explore The Leading Global Data Center Database!; Egtved, Denmark and Valrico, Florida, https://www.datacentermap.com/ (Accessed August 2025).

18. Baxtel. About Us Page: We are a Market Intelligence and Advisory Firm; Denver, Colorado, https://baxtel.com/about-us (Accessed August 2025).

19. Theara Coleman. 2023: the year of the AI boom; The Week US (Future US Inc), New York, New York, https://theweek.com/tech/2023-ai-boom (2023).

20. Richard Brown, Eric Masanet, Bruce Nordman, Bill Tschudi, Arman Shehabi, John Stanley, Jonathan Koomey, Dale Sartor, Peter Chan, Joe Loper, Steve Capana, Bruce Hedman, Rebecca Duff et al.Report to Congress on Server and Data Center Energy Efficiency: Public Law 109-431. (Ernest Orlando Lawrence Berkeley National Laboratory, Berkeley, CA (US), United States, 2007), https://www.osti.gov/biblio/929723; https://doi.org/10.2172/929723.

21. Nuoa Lei & Eric R. Masanet. Global Data Center Energy Demand and Strategies to Conserve Energy in Data Center Handbook: : Plan, Design, Build, and Operations of a Smart Data Center, Second Edition (ed Hwaiyu Geng P.E.) pgs 15-26 (2021, https://doi.org/10.1002/9781119597537.ch2).

22. Nuoa Lei, Jun Lu, Arman Shehabi & Eric Masanet. The water use of data center workloads: A review and assessment of key determinants. Resources, Conservation and Recycling 219, 108310 (2025). https://doi.org/10.1016/j.resconrec.2025.108310.

23. George Kamiya & Paolo Bertoldi. Energy Consumption in Data Centres and Broadband Communication Networks in the EU (ISSN 1831-9424); Publications Office of the European Union, Luxembourg, Luxembourg, https://data.europa.eu/doi/10.2760/706491 (2024).

24. Zhanat Murzakulova, Deniz Öztürk, Rafael Noster, Carla Groß, Lars Sorge, Svitlana Chebotarova, Simon Kieweg, Max Schulze, Jessica Stüer, Ralph Hintemann, Simon Hinterholzer, Severin Beucker, Peter Radgen et al. Stand und Entwicklung des Rechenzentrumsstandorts Deutschland; German Federal Ministry for Economy and Climate Protection, Berlin, Germany, https://www.bundeswirtschaftsministerium.de/Redaktion/DE/Publikationen/Technologie/stand-und-entwicklung-des-rechenzentrumsstandorts-deutschland.pdf (2025).

25. Arman Shehabi, Sarah Smith, Dale Sartor, Richard Brown, Magnus Herrlin, Jonathan Koomey, Eric Masanet, Nathaniel Horner, Inês Azevedo & William Lintner. United States Data Center Energy Usage Report (LBNL-1005775); Lawrence Berkeley National Laboratory, Berkeley, California, https://escholarship.org/uc/item/84p772fc (2016).

26. Tim Danton. Return of the tape: the unlikely renaissance of tape storage and how IBM Research is leading the way; TechFinitive,

London, United Kingdom, https://www.techfinitive.com/features/return-of-the-tape-the-unlikely-renaissance-of-tape-storage-and-how-ibm-research-is-leading-the-way/ (2024).

27. Amazon Web Services Inc. (AWS). What's the Difference Between an SSD and a Hard Drive?;Seattle, Washington, https://aws.amazon.com/compare/the-difference-between-ssd-hard-drive/(Accessed August 2025).

28. International Data Corporation (IDC). Worldwide Quarterly Network Infrastructure Tracker;Needham, Massachusetts, https://my.idc.com/getdoc.jsp?containerId=IDC_P23118 (Accessed August 2025).

29. Eric Masanet, Nuoa Lei & Jonathan Koomey. To better understand AI's growing energy use, analysts need a data revolution. Joule 8, 2427-2436 (2024). https://doi.org/10.1016/j.joule.2024.07.018.

30. Central Statistics Office (CSO) Ireland. Data Centres Metered Electricity Consumption 2023 (Online ISSN: 2811-5422); CSO Ireland, Cork, Ireland, https://www.cso.ie/en/releasesandpublications/ep/p-dcmec/datacentresmeteredelectricityconsumption2023/keyfindings/ (2024).

31. George Kamiya & Vlad C. Coroamă. Data Centre Energy Use: Critical Review of Models and Results; International Energy Agency (IEA) Technology Collaboration Programme on Energy Efficient End-Use Equipment (4E TCP) Efficient, Demand Flexible Networked Appliances (EDNA) Platform (EDNA-IEA 4E TCP), Paris, France, https://www.iea-4e.org/wp-content/uploads/2025/05/Data-Centre-Energy-Use-Critical-Review-of-Models-and-Results.pdf (2025).

32. David Mytton & Masaō Ashtine. Sources of data center energy estimates: A comprehensive review. Joule 6, 2032-2056 (2022). https://doi.org/10.1016/j.joule.2022.07.011.

33. Eric Masanet, Arman Shehabi, Nuoa Lei, Sarah Smith & Jonathan Koomey. Recalibrating global data center energy-use estimates. Science 367, 984-986 (2020). https://doi.org/10.1126/science.aba3758.

34. Meaghan Tobin. China Is Spending Billions to Become an A.I. Superpower; The New York Times, Taipei, Taiwan, https://www.nytimes.com/2025/07/16/technology/china-ai.html (2025).

35. Central Statistics Office (CSO) Ireland. Data Centres Metered Electricity Consumption 2024 (On-line ISSN: 2811-5422); CSO Ireland, Cork, Ireland, https://www.cso.ie/en/releasesandpublications/ep/p-dcmec/datacentresmeteredelectricityconsumption2024/ (2025).

36. Khari Johnson. Crackdown on power-guzzling data centers may soon come online in California;CalMatters, Sacramento, California, https://calmatters.org/economy/technology/2025/02/data-center-crackdown-to-protect-california-electricity-rates/#:~:text=Last%20month%20the%20municipal%20utility%20company%20in,up%2060%25%20of%20energy%20from%20the%20utility (2025).

37. Joint Legislative Audit & Review Commission (JLARC). Data Centers in Virginia; JLARD, Richmond, Virginia, https://jlarc.virginia.gov/landing-2024-data-centers-in-virginia.asp (2024).

38. Lisa Martine Jenkins & Phoebe Skok. Mapping the data center power demand problem, in three charts; Latitude Media, Somerville, Massachusetts, https://www.latitudemedia.com/news/mapping-the-data-center-power-demand-problem-in-three-charts/ (2024).

39. World Resources Institute (WRI) and World Business Council for Sustainable Development (WBCSD). The Greenhouse Gas Protocol: A Corporate Accounting and Reporting Standard (Revised Edition); WRI and WBCSD, Washington, DC and Geneva, Switzerland, https://ghgprotocol.org/sites/default/files/standards/ghg-protocol-revised.pdf (2004).

40. US Environmental Protection Agency (EPA). Inventory of U.S. Greenhouse Gas Emissions and Sinks: 1990-2023 (EPA 430-R-25-003); US EPA, Washington, DC, https://library.edf.org/AssetLink/145ky510ew61fk1tq5c2klp5kq5yp33j.pdf (2025).

41. David Patterson, Joseph Gonzalez, Urs Hölzle, Quoc Le, Chen Liang, Lluis-Miquel Munguia, Daniel Rothchild, David R. So, Maud Texier & Jeff Dean. The Carbon Footprint of Machine Learning Training Will Plateau, Then Shrink. Computer 55, 18-28 (2022). https://doi.org/10.1109/MC.2022.3148714.

42. Alexandra Sasha Luccioni, Emma Strubell & Kate Crawford. From Efficiency Gains to Rebound Effects: The Problem of Jevons' Paradox in AI's Polarized Environmental Debate. arXiv:2501.16548 (2025). https://doi.org/10.48550/arXiv.2501.16548.

43. Cyrus Moulton. What is Jevons Paradox? And why it may — or may not — predict AI's future;Northeastern Global News, Northeastern University, Boston, Massachusetts, https://news.northeastern.edu/2025/02/07/jevons-paradox-ai-future/ (2025).

44. Kelly Vaughn. Jevons Paradox: The Debate That Just Won't Die; Rocky Mountain Institute (RMI), Basalt, Colorado, https://rmi.org/blog_jevons_paradox/ (2012).

45. David Owen. The Efficiency Dilemma; The New Yorker, New York, New York, https://www.newyorker.com/magazine/2010/12/20/the-efficiency-dilemma (2010).

46. Katie Fehrenbacher. Data center electricity demand is not a crisis: report; Axios Pro, Arlington, Virginia, https://www.axios.com/pro/climate-deals/2025/02/05/data-center-electricity-demand-crisis-report (2025).

47. Amy Luers, Jonathan Koomey, Eric Masanet, Owen Gaffney, Felix Creutzig, Juan Lavista Ferres & Eric Horvitz. Will AI accelerate or delay the race to net-zero emissions? Nature 628, 718-720 (2024). https://doi.org/10.1038/d41586-024-01137-x.

48. Nuoa Lei, Eric Masanet & Jonathan Koomey. Best practices for analyzing the direct energy use of blockchain technology systems: Review and policy recommendations. Energy Policy 156, 112422 (2021). https://doi.org/10.1016/j.enpol.2021.112422.

49. Cambrige Bitcoin Electricity Consumption Index (CBECI). Bitcoin network power demand; University of Cambridge Judge Business School, Cambridge Centre for Alternative Finance, Cambridge, United Kingdom, https://ccaf.io/cbnsi/cbeci (Accessed August 2025).

50. The White House. Climate and Energy Implicatons of Crypto-Assets in the United States;Washington, DC, https://bidenwhitehouse.archives.gov/wp-content/uploads/2022/09/09-2022-Crypto-Assets-and-Climate-Report.pdf (2022).

51. Cambrige Bitcoin Electricity Consumption Index (CBECI). Bitcoin Mining Map; University of Cambridge Judge Business School, Cambridge Centre for Alternative Finance, Cambridge, United Kingdom, https://ccaf.io/cbnsi/cbeci/mining_map (Accessed August 2025).

52. Cryptopedia Staff. Blockchain Consensus Mechanisms Beyond PoW and PoS; Cryptopedia, Gemini Trust Company, LLC, New York, New York, https://www.gemini.com/cryptopedia/blockchain-consensus-mechanism-types-of-algorithm (2023).

2.1 Information Technology (IT) Equipment

Alp Kucukelbir

Data centers are complex facilities that include information technology (IT) and infrastructure equipment, often distributed across several buildings. IT equipment typically comprises electronics housed in server racks, networked to one another and powered by electrical infrastructure. Each rack contains general purpose and/or accelerated processors, connected to memory chips and short- and long-term storage. Electrical infrastructure converts electricity to the appropriate levels for IT equipment operations.

A data center running artificial intelligence (AI) workloads operates like a massive, high-speed postal sorting facility designed to handle an enormous volume of specialized packages. The servers act as sorting stations where incoming data packets (like letters and parcels) are processed by workers (processors) who must quickly analyze, categorize and route information. Memory serves as temporary holding areas where packages wait to be processed, while storage systems function as vast warehouses storing reference materials needed for sorting decisions. The networking infrastructure resembles conveyor belts that move packages between stations.

As described in Chapter 1, AI workloads are particularly energy intensive; processing AI packages requires cross-referencing millions of previous deliveries simultaneously—each "package" (data point) must be compared against vast databases of patterns and examples before determining where it should go next.

A. Defining and Measuring Energy Efficiency

Computing energy efficiency has improved by an extraordinary factor of 10 billion over 5 decades, from ENIAC's (Electronic Numerical Integrator and Computer) 150-kilowatt power consumption in 1946 to modern processors achieving billions of operations per watt-hour. This transformation followed Koomey's Law, which demonstrated that computations per watt-hour doubled every 1.6 years during the "golden age" from 1946 to 2000. The improvement was driven by transistors shrinking while maintaining nearly constant power density, creating exponential performance gains with stable or decreasing power consumption.[1]

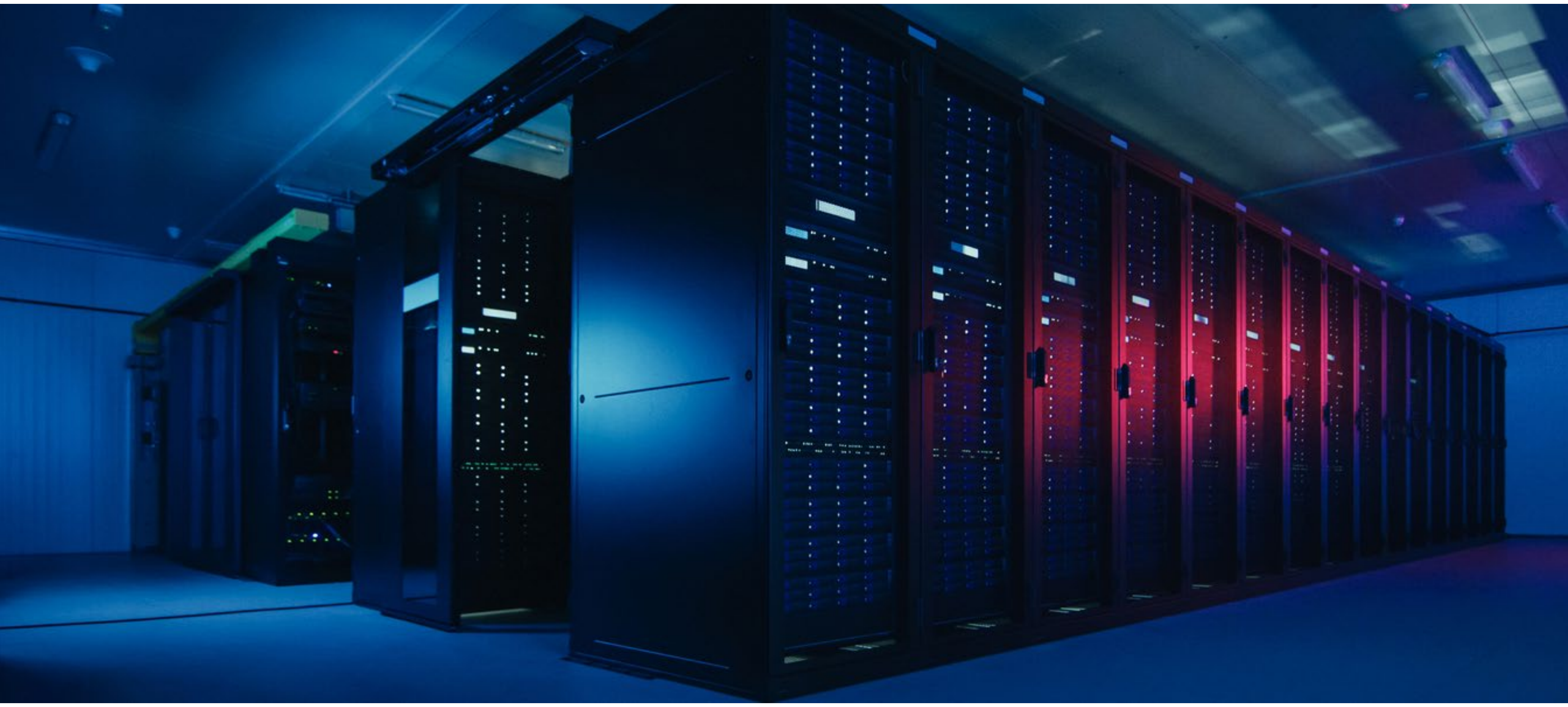


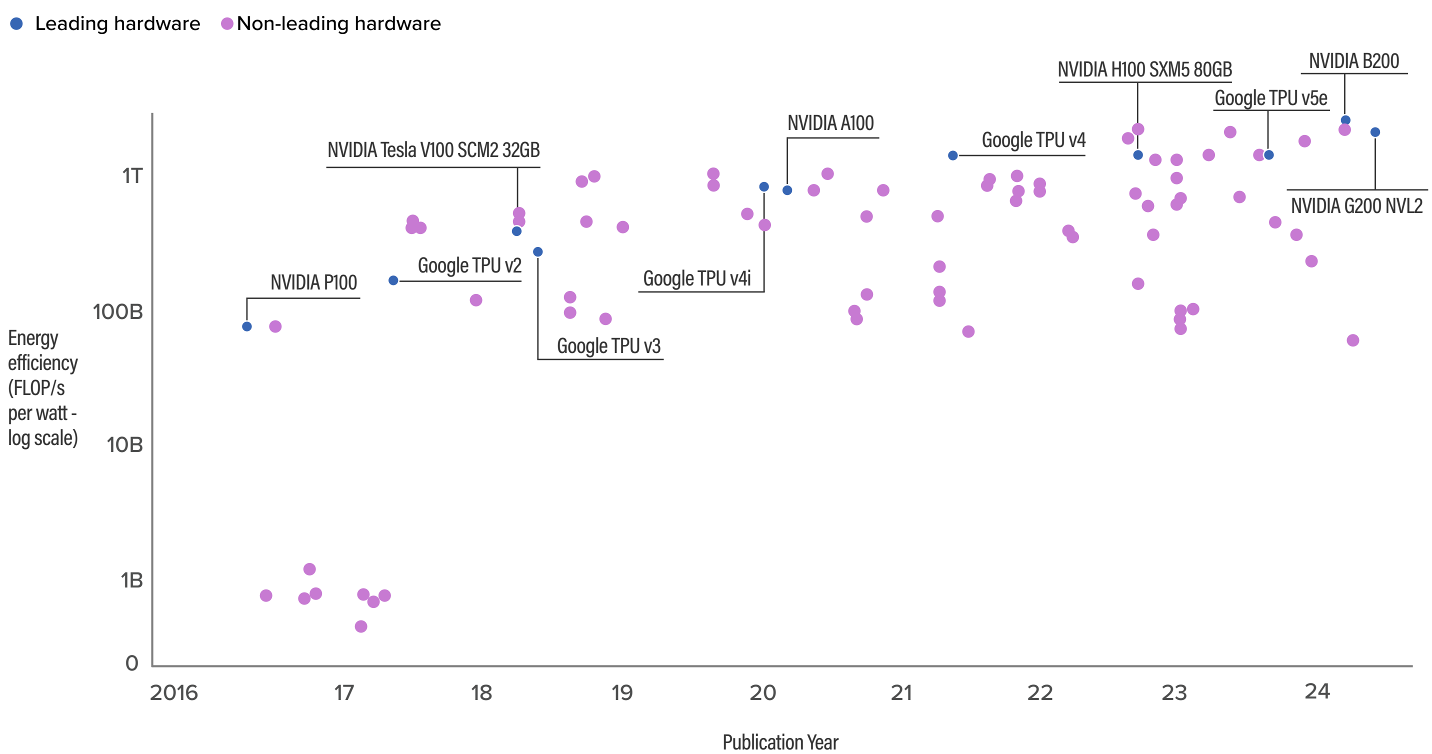
Figure 2.1-1. *Image of a data center.*

Around 2005, this golden age ended as Dennard scaling broke down due to quantum effects at atomic scales. The industry pivoted to architectural innovations, with recent advances showing continued but slower progress—efficiency now doubles every 2.3 years instead of every 1.6 years.[2] The last decade has seen exceptional technological progress. Processors advanced through smaller manufacturing nodes (22 nm to 3 nm), multi-core architectures and dynamic power scaling. Memory, storage and networking designs evolved. Critically, ARM processors and specialized AI chips challenged x86 dominance with superior performance-per-watt.

Improvements in the energy efficiency of IT equipment components have slowed but not stopped. These improvements are expected to continue in the years ahead. (See the second part of this chapter.) Yet when these component-level improvements are viewed at the system scale, a different picture emerges. Chip-level efficiency gains are increasingly offset by exponential demand growth from AI and other workloads. Despite these energy efficiency improvements, carbon emissions at data centers continue to rise.[3]

Companies, researchers and engineering consortia have yet to align on a comprehensive way to measure IT equipment energy efficiency in AI workloads. Epoch.AI, a research institute, tracks hardware efficiency trends using FLOPS per watt—a traditional metric applied to scientific computing involving many "floating point operations per second" (FLOPS). While this metric does not provide a complete picture due to IT equipment diversity and the nature of AI workloads, it serves as a good approximation. The MLCommons Power Work Group recently released a benchmark intended to include applications from data centers to mobile devices, from training to inference.[4] Stanford's Human-Centered AI 2025 Index Report, using Epoch.AI data, calculates a 40% reduction in energy consumption in AI-specific hardware per FLOP over the past three years, indicating that progress may be slowing down even further (Figure 2.1-2).[5]

Figure 2.1-2. *Energy efficiency of leading machine learning hardware, 2016-24.*

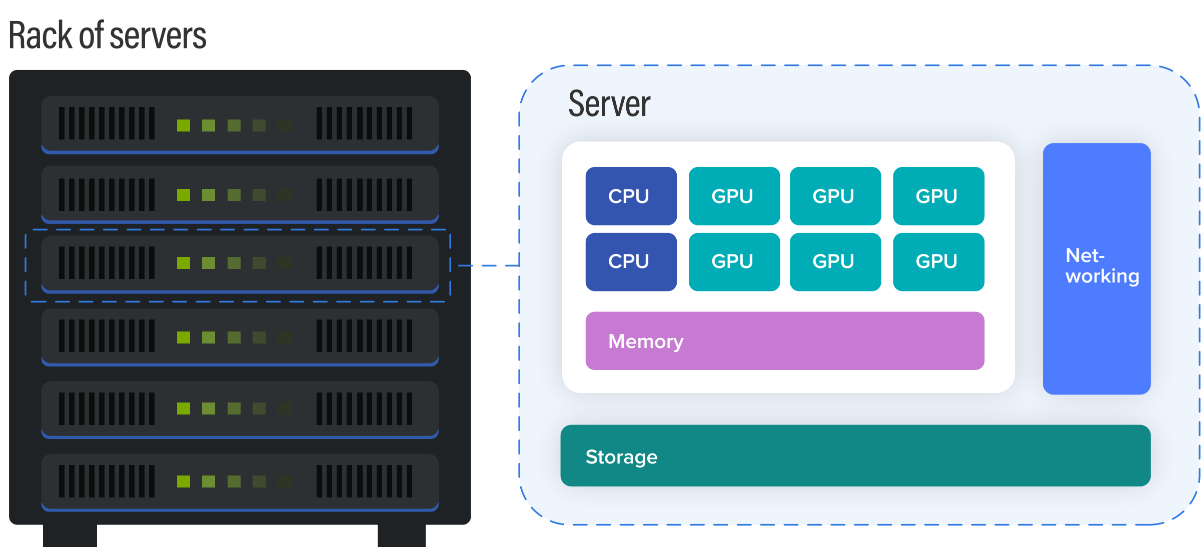


B. Components

Chapter 1 discussed energy use trends at the level of entire servers, showing how average power draw has increased over time. Below we discuss the components within those servers and how each contributes to overall energy use. Figure 2.1-3 presents a sketch of typical components found in a server. Each component consumes different amounts of energy in different ways.

Central processing units (CPUs) handle general-purpose computing tasks with moderate power draw (50-200 W), while accelerated processors, like graphics processing units (GPUs) and custom chips like tensor processing units (TPUs), excel at parallel processing but consume much more electricity (400-1000+ W). Memory provides fast temporary storage for active data, consuming negligible power that scales with capacity. Storage devices, such as solid-state drives (SSDs), provide persistent data storage, with SSDs using less power than older mechanical drives. Networking components move data between systems, with power consumption varying dramatically based on speed and distance requirements.

Figure 2.1-3. *Components of a server.*



i. Computation: processors and memory

Chapter 1 shows that servers consume the majority of power in data centers. Processors and memory are a primary source of energy usage within servers. Historically, data centers relied mainly on CPUs, which ran at roughly 50-200 W per chip. GPUs for AI have been drawing 400-700 W, with newer (2024+) chips expected to run at over 1000 watts.[6] GPUs contain thousands of smaller cores compared to CPUs' dozens of larger cores, meaning far more transistors are actively switching and consuming power simultaneously during parallel computations.

Major chip designers are pursuing different energy efficiency strategies. NVIDIA claims its most recent processor architecture uses up to 25x less energy relative to its own previous generation chips during AI inference.[7] Advanced Micro Devices (AMD) is focusing on using fewer GPUs to achieve similar outputs, claiming its recent offerings have a 30% advantage over competitors.[8] Intel targets energy savings with open system standards, claiming a 40% average power reduction relative to competitors.[9] Hyperscalers are targeting energy efficiency by specializing for AI applications. Google claims its most recent accelerated processor TPUs are 67% more energy efficient than previous versions.[10] Amazon Web Services similarly claims a 40% improvement in energy efficiency for its own accelerated processors.[11] Microsoft, Meta, IBM and other hyperscalers appear to be aiming to achieve similar outcomes.

Startup companies are developing novel processor architectures with promising energy efficiency performance. Untether AI recently demonstrated a 3-6x improvement across multiple machine learning (ML) Commons benchmark categories.[12] Groq claims to achieve 10x energy efficiency improvement through a novel chip design.[13]

The software that trains AI models and runs inference requires large amounts of memory. These chips consume energy as data are written to and read from memory. AI systems leverage higher memory bandwidth (the rate at which data are transferred) to accelerate AI training and inference. But higher bandwidth typically correlates with higher power consumption. Researchers and industry are working toward novel architectures to increase bandwidth while minimizing energy usage.[14]

ii. Storage

Chapter 1 distinguishes energy usage between servers and storage. Below we discuss storage technology within the context of servers (for short- to medium-term information needs during processing).

SSDs provide the high-speed data access required for training large AI models. During training, AI systems must process massive data sets containing billions of examples, which requires continuous, rapid data transfer from storage to processing units. For AI inference operations, SSDs enable rapid access to AI models and any reference data needed for generating responses.

AI workloads create unique energy demands for SSDs due to their sustained high-throughput patterns. During training, SSDs experience bursty read and write patterns as data sets are streamed repeatedly, keeping the drives in active states rather than allowing them to enter lower-power idle modes. In contrast, memory usage is read-heavy, continuous and sustained during inference.

Current SSD technology is almost twice as fast as the previous generation but consumes 50-100% more energy during peak operations.[15] Memory chip manufacturers, such as Samsung and Micron, are developing new designs that maintain performance while reducing energy consumption by half.[16,17]

iii. Networking

Data center networking controls how information flows throughout the system. Switches direct data to their destinations—some locally within server racks, others regionally within building sections. Load balancers ensure that no single server becomes overwhelmed. Together, these devices aim to provide fast and reliable information flow.

Similar to memory and storage requirements, AI workloads require high bandwidth networks. These expand beyond traditional enterprise networking standards. AI data centers' bandwidth requirements can range from several gigabits per second to terabits per second.[18]

Traditional enterprise networking equipment consumes dramatically less power per port than specialized AI networking hardware, with power differences 3-10x higher for AI equipment. Despite this increase, networking remains under 5% of total data center power consumption due to the massive scale of AI compute requirements. Networking equipment manufacturers are adopting technologies, such as co-packaged optics, to increase bandwidth while reducing power consumption, with projections of up to an 80% reduction in energy usage.[19,20]

C. Innovations and Forecasted Efficiency Gains

Data centers present significant opportunities for improving energy efficiency through both established best practices and emerging innovations. While conventional strategies benefit all data center operations (see Table 2.1-1 below), AI workloads demand specialized approaches, including high-bandwidth networking, sustained storage performance and power oversubscription tailored to inference patterns. Hardware advances in electronics, photonics and power distribution can improve energy efficiency, while tailoring server designs and data center operations to specific AI workloads—training versus inference—offers further energy savings. Although the rate of improvement in energy efficiency appears to be slowing, the innovations below may lead to new upticks in progress.

Table 2.1-1. Conventional data centers typically operate with legacy equipment at low utilization rates, basic power distribution systems, minimal virtualization, reactive operational practices and outdated hardware refresh cycles.

Category	Best Practice	Baseline
Electronics Manufacturing	Select ENERGY STAR certified servers and equipment	Standard servers without efficiency certifications
	Deploy solid-state drives (SSDs) over mechanical drives	Traditional spinning disk drives
	Use efficient memory technologies (e.g., DDR5, low-voltage)	Older DDR3/DDR4 standard voltage memory
Processors	Enable processor power management (e.g., C-states, P-states)	Processors running at constant maximum frequency
	Implement dynamic voltage/frequency scaling	Fixed voltage and frequency operation
Storage	Storage virtualization and deduplication	Direct-attached storage with redundant data
	Tiered storage with automated data placement	Single-tier storage systems
	Implement storage power management	Always-on storage arrays
Networking	Deploy energy-efficient network switches	Legacy switches without power scaling
	Network virtualization and consolidation	Physical dedicated network infrastructure
	Use adaptive link rate and port power management	Fixed-rate network interfaces
Power Conversion	High-efficiency power supplies (>90% efficiency)	Standard 80% efficient power supplies
	High-efficiency uninterruptible power supply (UPS) systems (>95%)	Traditional UPS systems at 85–90% efficiency
	Power factor correction (>0.95)	Uncorrected power factor (0.7–0.8)
Operations	Real-time power monitoring and management	Manual monitoring with monthly readings
	Automated workload scheduling and migration	Static workload placement

i. Electronics manufacturing and packaging

Chip packaging is the process of enclosing semiconductors in protective materials and connecting them to external circuits through pins. It provides physical protection, electrical connections and thermal management for delicate silicon chips.

Advanced packaging techniques, such as Taiwan Semiconductor Manufacturing Company's (TSMC's) chip-on-wafer-on-substrate (CoWoS), are delivering a 30%

reduction in power consumption compared to traditional packaging through reduced interconnect distances and improved thermal management.[21] The company is targeting a doubling of its production capacity to 75,000 wafers in 2025.[22]

AI-designed chips represent the next frontier of chip designs with better power efficiency.[23] For example, researchers at Oregon State University leveraged AI to design chips from new materials, achieving a 6x improvement in energy efficiency.[24]

ii. Photonics innovation

Silicon photonics innovation is revolutionizing data center networking. This technology uses light (photons) instead of electricity to transmit data through silicon-based optical components like waveguides and modulators. It achieves better power efficiency because photons do not generate resistive heat during transmission and can carry data over longer distances without signal amplification, unlike electrical signals that require constant power to overcome resistance and maintain signal integrity.[25]

Companies are leading research and development (R&D) efforts in photonics. Intel's co-packaged optics demonstrated a 67% power reduction over pluggable optical transceivers.[26] Ayar Labs' TeraPHY chiplets claim 4-8x greater power efficiency than traditional interconnects,[27] while NVIDIA's co-packaged solutions claim 3.5x better power efficiency.[28]

iii. Power conversion improvements

Power electronics are circuits that convert and control electrical power between different voltage levels or frequencies or from AC to DC. Gallium nitride (GaN) and silicon carbide (SiC) semiconductors have wider bandgaps than silicon, allowing them to operate at higher voltages, temperatures and switching frequencies with lower resistance. This enables smaller, lighter power converters with less energy lost as heat during conversion.

GaN and SiC adoption is already driving efficiency gains in data centers.[29] These technologies are showcasing up to 98% energy conversion efficiency[30] versus the 85-90% baselines seen with older power electronics. Both technologies are in active development, with pathways to an increased range of operation and reliability.[31]

iv. Data center operations

AI workflows are creating specific opportunities for operating data center IT equipment. Microsoft reports that the average and peak power usage in AI inference are bounded. This implies that AI data centers used solely for inference offer substantial headroom for power oversubscription, allowing the deployment of 30% more servers per AI inference data center with minimal performance loss.[32]

v. Server design

Server design innovations center on liquid cooling advances that dramatically improve energy efficiency. (Chapter 2.3 explores this topic in detail.) Other strategies include Microsoft Azure's efforts to use existing components to assemble server racks, leading to a net 8% reduction in emissions.[33] Amazon Web Services (AWS) has focused on standardized yet modular equipment to help retrofit existing data centers.[34]

Box 2.1-1

Moving computation to the edge

Some AI use cases offer opportunities to avoid data centers altogether. So-called "edge AI" processes data locally on devices (smartphones, industrial robots, autonomous vehicles) rather than communicating with data centers. Motivated by the ability to work offline, with low latencies and strong data privacy guarantees, running AI workloads "on the edge" can also reduce energy consumption.

Edge AI workflows eliminate data transmission to data centers, thereby reducing their computational load. Specialized processors, such as field-programmable gate arrays (FPGAs) and edge-optimized AI chips, can effectively run AI inference workflows.[35] The tradeoff is that edge AI devices are typically more expensive and less energy efficient than traditional counterparts. Multiple companies, such as DEEPX, Hailo and Axelara, are actively developing energy efficient and affordable edge AI processors.

While AI training will certainly remain at data center levels, forecasting how AI inference demand might shift toward the edge is challenging.[36] Rising data center costs, regulatory pressure for data privacy and real-time applications may push toward edge AI adoption. However, increasingly larger model sizes and edge AI device costs may continue to drive AI inference demand in data centers.

D. Recommendations

1. *Companies, industry standard setters and engineering consortia should* ***align on common metrics for calculating and reporting the energy efficiency of IT equipment****. Data center operators should support such efforts since energy efficiency directly impacts their operating costs.*

2. *Governments and educational institutions should* ***develop and distribute resources to assist non-technical audiences in understanding and analyzing the energy requirements of IT equipment.***

3. *Data center operators, utilities and government agencies should* ***consider the nature of AI computation workloads in designing, provisioning, operating and regulating data centers.*** *Differences between AI training and inference should be paramount during decision making.*

4. *Governments, utilities and industry consortia should* ***advance knowledge sharing platforms, case study data, diagnostic tools and training materials related to improving the energy efficiency of IT hardware*** *from procurement, operations and management perspectives.*

5. *Data center operators should* ***conduct, support and publish AI inference demand forecasts.*** *Trends between centralized data center computations and edge applications should be emphasized.*

6. *Data center operators should* ***redouble efforts to maximize the energy efficiencies of existing IT hardware****, including the adoption of efficient equipment, virtualization, zombie server identification initiatives, and refresh cycles optimized for energy efficiency.*

E. References

1. Jonathan Koomey, Stephen Berard, Marla Sanchez & Henry Wong. Implications of Historical Trends in the Electrical Efficiency of Computing. IEEE Annals of the History of Computing 33, 46-54 (2011). https://doi.org/10.1109/MAHC.2010.28.

2. Alberto Prieto, Beatriz Prieto, Juan José Escobar & Thomas Lampert. Evolution of computing energy efficiency: Koomey's law revisited. Cluster Computing 28, 42 (2024). https://doi.org/10.1007/s10586-024-04767-y.

3. International Energy Agency (IEA). CO_2 emissions associated with electricity generation for data centres by case, 2020-2035; IEA, Paris, France, https://www.iea.org/data-and-statistics/charts/co2-emissions-associated-with-electricity-generation-for-data-centres-by-case-2020-2035 (2025).

4. MLCommons. MLCommons Power Working Group Presents MLPerf Power benchmark at IEEE HPCA Symposium; MLCommons, Dover, Delaware, https://mlcommons.org/2025/03/ml-commons-power-hpca/ (2025).

5. Nestor Maslej, Loredana Fattorin, Raymond Perrault, Yolanda Gil, Vanessa Parli, Njenga Kariuki, Emily Capstick, Anka Reuel, Erik Brynjolfsson, John Etchemendy, Katrina Ligett, Terah Lyons, James Manyika et al. The AI Index 2025 Annual Report; AI Index Steering Committee, Institute for Human-Centered AI, Stanford University, Stanford, California, https://hai.stanford.edu/ai-index/2025-ai-index-report (2025).

6. Karthik Ramachandran, Duncan Stewart, Kate Hardin, Gillian Crossan & Ariane Bucaille. As generative AI asks for more power, data centers seek more reliable, cleaner energy solutions (Deloitte Insights); Deloitte Center for Technology, Media & Telecommunications, London, United Kingdom, https://www.deloitte.com/us/en/insights/industry/technology/technology-media-and-telecom-predictions/2025/genai-power-consumption-creates-need-for-more-sustainable-data-centers.html (2024).

7. Kristin Uchiyama. NVIDIA Blackwell Platform Arrives to Power a New Era of Computing; NVIDIA, Santa Clara, California, https://nvidianews.nvidia.com/news/nvidia-blackwell-platform-arrives-to-power-a-new-era-of-computing (2024).

8. Advanced Micro Devices (AMD). AMD Instinct MI300 Series Accelerators; AMD, Santa Clara, California, https://www.amd.com/en/products/accelerators/instinct/mi300.html (Accessed August 2025).

9. Intel Newsroom. Intel Unleashes Enterprise AI with Gaudi 3, AI Open Systems Strategy and New Customer Wins; Intel, Santa Clara, California, https://newsroom.intel.com/artificial-intelligence/vision-2024-enterprise-ai-gaudi-3-open-systems-strategy (2024).

10. Amin Vahdat. Announcing Trillium, the sixth generation of Google Cloud TPU; Google Cloud, Mountain View, California, https://cloud.google.com/blog/products/compute/introducing-trillium-6th-gen-tpus (2024).

11. CTOL Editors - Yasmine. AWS Unveils Next-Gen AI Chips with Trainium 3 and Ultra Servers at re:Invent 2024, But Unlikely to Challenge Nvidia's Dominance; CTOL Digital Solutions, Ennetbaden, Switzerland, https://www.ctol.digital/news/aws-trainium-3-chips-vs-nvidia-ai-hardware-reinvent-2024/ (2024).

12. Untether AI. Untether AI Announces speedAI Accelerator Cards As World's Highest Performing, Most Energy Efficient AI Accelerators According to MLPerf Benchmarks; Untether AI, Toronto, Ontario, https://www.untether.ai/untether-ai-announces-speedai-accelerator-cards-as-worlds-highest-performing-most-energy-efficient-ai-accelerators-according-to-mlperf-benchmarks/ (2024).

13. Groq. Products - Groq Is Fast AI Inference; Groq, San Jose, California, https://groq.com/products (Accessed August 2025).

14. Joo-Hyung Chae. High-Bandwidth and Energy-Efficient Memory Interfaces for the Data-Centric Era: Recent Advances, Design Challenges, and Future Prospects. IEEE Open Journal of the Solid-State Circuits Society 4, 252-264 (2024). https://doi.org/10.1109/OJSSCS.2024.3458900.

15. Anshul Rana. SSD Power Consumption: How Much Watts SSDs Consume?; StoredBits.com, https://storedbits.com/ssd-power-consumption/ (2025).

16. Samsung News. Now Available: Samsung 9100 PRO Series SSDs with Breakthrough PCIe® 5.0 Performance; Samsung, Suwon-si, South Korea, https://news.samsung.com/us/samsung-announces-9100-pro-series-ssds-with-breakthrough-pcie-5-0-performance/ (2025).

17. Alvaro Toledo. The world's fastest data center SSD for AI workloads; Micron Technology, Inc., Boise, Idaho, https://www.micron.com/about/blog/storage/ssd/revolutionize-ai-workloads-with-the-worlds-fastest-data-center-ssd. (2024).

18. Alexandra Jonker & Alice Gomstyn. What is an AI data center?; IBM, Armonk, New York, https://www.ibm.com/think/topics/ai-data-center (2025).

19. Laura Peters. Co-Packaged Optics Reaches Power Efficiency Tipping Point; Semiconductor Engineering, https://semiengineering.com/co-packaged-optics-reaches-power-efficiency-tipping-point/ (2025).

20. NVIDIA. NVIDIA Silicon Photonics; NVIDIA, Santa Clara, California, https://www.nvidia.com/en-us/networking/products/silicon-photonics/ (Accessed August 2025).

21. Taiwan Semiconductor Manufacturing Company (TSMC). TSMC Unveils Next-Generation A14 Process at North America Technology Symposium; TSMC, Hsinchu Science Park, Taiwan, https://pr.tsmc.com/english/news/3228 (2025).

22. TrendForce News. TSMC Set to Expand CoWoS Capacity to Record 75,000 Wafers in 2025, Doubling 2024 Output; TrendForce Corporation, Taipei, Taiwan, https://www.trendforce.com/news/2025/01/02/news-tsmc-set-to-expand-cowos-capacity-to-record-75000-wafers-in-2025-doubling-2024-output/ (2025).

23. Jeroen Kusters, Deb Bhattacharjee, Jordan Bish, Jan Thomas Nicholas, Duncan Stewart & Karthik Ramachandran. 2025 global semiconductor industry outlook (Deloitte Insights); Deloitte Center for Technology, Media & Telecommunications, London, United Kingdom, https://www.deloitte.com/us/en/insights/industry/

technology/technology-media-telecom-outlooks/semiconductor-industry-outlook.html (2025).

24. Steve Lundeberg. New computer chips show promise for reducing energy footprint of artificial intelligence; Oregon State University Newsroom, Corvallis, Oregon, https://news.oregonstate.edu/news/new-computer-chips-show-promise-reducing-energy-footprint-artificial-intelligence (2024).

25. Shupeng Ning, Hanqing Zhu, Chenghao Feng, Jiaqi Gu, Zhixing Jiang, Zhoufeng Ying, Jason Midkiff, Sourabh Jain, May H. Hlaing, David Z. Pan & Ray T. Chen. Photonic-Electronic Integrated Circuits for High-Performance Computing and AI Accelerators. Journal of Lightwave Technology 42, 7834-7859 (2024). https://doi.org/10.48550/arXiv.2403.14806.

26. Intel Newsroom. Intel Demonstrates First Fully Integrated Optical I/O Chiplet; Intel, Santa Clara, California, https://newsroom.intel.com/artificial-intelligence/intel-unveils-first-integrated-optical-io-chiple (2024).

27. Ayar Labs. Optical I/O Technology is Essential to Eliminate Bottlenecks; Ayar Labs, Inc., San Jose, California, https://ayarlabs.com/optical-io-products/ (Accessed August 2025).

28. NVIDIA Co-Packaged Silicon Photonics Networking Switches." NVIDIA, https://www.nvidia.com/en-us/networking/products/silicon-photonics/.

29. Gary Elinoff. GaN and SiC Transform AI Data Centers Efficiency; Electropages Media Ltd, Dorset, United Kingdom, https://www.electropages.com/blog/2024/07/gan-and-sic-transform-ai-data-centers-efficiency (2024).

30. Justin Chou. Scaling AI Data Center Power Delivery with Si, SiC, and GaN; Electronic Design, Nashville, Tennessee, https://www.electronicdesign.com/technologies/power/power-supply/article/55289191/scaling-ai-data-center-power-delivery-with-si-sic-and-gan (2025).

31. Buffolo, M., et al. "Review and Outlook on GaN and SiC Power Devices: Industrial State-of-the-Art, Applications, and Perspectives." IEEE Transactions on Electron Devices, vol. 71, no. 3, March 2024, pp. 1344-55. IEEE Xplore, https://doi.org/10.1109/TED.2023.3346369.

32. Pratyush Patel, Esha Choukse, Chaojie Zhang, Íñigo Goiri, Brijesh Warrier, Nithish Mahalingam & Ricardo Bianchini. "Characterizing Power Management Opportunities for LLMs in the Cloud" in Proceedings of the 29th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Volume 3, La Jolla, CA, USA. 207-222, https://doi.org/10.1145/3620666.3651329 (2024).

33. Jaylen Wang, Daniel S. Berger, Fiodar Kazhamiaka, Celine Irvene, Chaojie Zhang, Esha Choukse, Kali Frost, Rodrigo Fonseca, Brijesh Warrier, Chetan Bansal, Jonathan Stern, Ricardo Bianchini & Akshitha Sriraman. Enabling Sustainable Cloud Computing with Low-Carbon Server Design. IEEE Micro, 1-9 (2025). https://doi.org/10.1109/MM.2025.3572955.

34. Amazon Press Center. AWS Announces New Data Center Components to Support AI Innovation and Further Improve Energy Efficiency; Amazon.com, Inc., Las Vegas, Nevada, https://press.aboutamazon.com/2024/12/aws-announces-new-data-center-components-to-support-ai-innovation-and-further-improve-energy-efficiency (2024).

35. Péter Szántó, Tamás Kiss & Károly János Sipos. “Energy-efficient AI at the Edge” in 2022 11th Mediterranean Conference on Embedded Computing (MECO), 1-6, https://doi.org/10.1109/MECO55406.2022.9797178,(2022).

36. Steven Carlini. The Current And Future Path To AI Inference Data Center Optimization; Forbes Media LLC, Jersey City, New Jersey, https://www.forbes.com/councils/forbestechcouncil/2025/01/28/the-current-and-future-path-to-ai-inference-data-center-optimization/ (2025).

2.2 Software

Alp Kucukelbir and Minjue Wu

Algorithms instruct computers on how to solve problems. But not all algorithms are created equal. Computer scientists continuously seek cleverer algorithms to improve speed and efficiency.

Consider the task of sorting a list of random numbers from lowest to highest. One approach compares every pair of adjacent numbers and swaps them if they are in the wrong order. For a list of one thousand numbers, the processor would have to make one million comparisons. There is, however, a better way: repeatedly divide the list in half and merge the smaller lists back together in sorted order. This algorithm requires only about ten thousand comparisons—a 100x speedup.

Both algorithms produce the correct ordering of numbers, from lowest to highest. But one requires less computation and energy to do so; it is an objectively better algorithm.

Clever algorithm design, as in the example above, has enabled technological innovations such as digital audio and video, medical imaging, telecommunications, DNA sequencing, and supply chain optimization. When and how such algorithmic breakthroughs arise are hard to predict—these moments are considered pivotal in computer science.

This chapter examines how software efficiency impacts data center energy consumption, from algorithmic optimizations that reduce computational requirements to operational techniques that align artificial intelligence (AI) workloads with clean energy availability. As AI inference demand increases with autonomous agents and reasoning models, mastering both technical and operational efficiency becomes essential for sustainable data center operations.

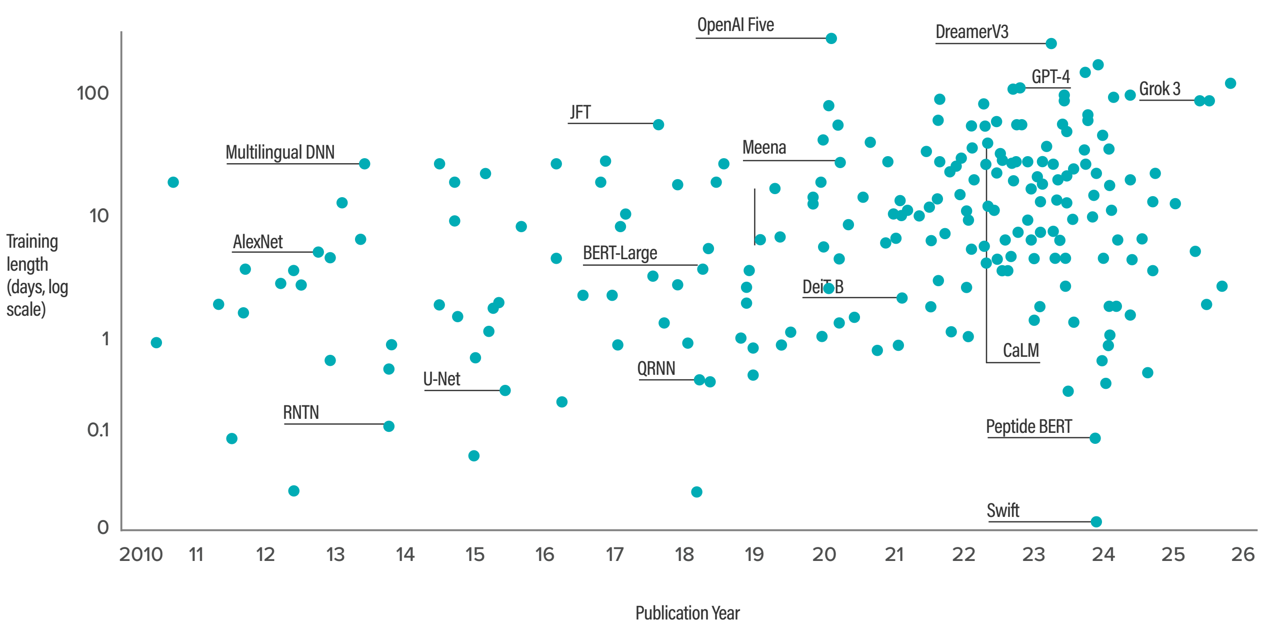
A. How Is AI Software Different?

While traditional algorithms like sorting have clear "correct" outputs we can verify, AI systems operate differently. The energy efficiency principles remain the same—better algorithms use less computation—but measuring "better" becomes far more complex. AI workflows do not exactly fit this paradigm because we do not yet have an equivalent definition of what the "correct" output of an AI algorithm should be. Despite this, there is intense research and development activity seeking to improve the computational efficiency of AI workflows, such as DeepSeek's innovations with its V3 and R1 systems.

AI systems operate in two phases: training (building the software system) and inference (using the software system). Training modern AI systems increasingly requires larger computational tasks, distributed over tens of thousands of servers, sometimes across geographies (possibly multiple data centers) and often spread over time. Recent state-of-the-art models, such as GPT-4 and Llama 3.1 were trained in dedicated data centers and are estimated to have taken around 90 days of training (see Figure 2.2-1).

This scale and complexity of AI workflows creates unique efficiency challenges. Unlike the sorting example, in which one approach was objectively 100x faster, AI efficiency requires new metrics and measurement approaches.

Figure 2.2-1. *Training length of notable AI models, 2010-present*



Reproduction of Figure 1.3-14 from Stanford HAI 2025 AI Index Report. Note: y-axis is in logarithmic scale. Training frontier models takes weeks to months, but this time is not necessarily directly related to the total amount of computation because of differences in IT equipment (see Chapter 2.1). Source: Epoch AI | Chart: 2025 AI Index report

B. What Does Efficiency Mean for AI Systems?

AI systems work by processing data using models, which are mathematical frameworks for identifying patterns in data.[1] Unlike in traditional software, we cannot verify the "correctness" of AI systems. Instead, software efficiency revolves around researching and developing new model designs and computation algorithms. These efforts are anchored around two model characteristics: size and benchmark performance. (For simplicity, this section focuses on large language models and omits audio, image, video and protein models, but similar notions apply.)

AI model size typically refers to the number of parameters—the mathematical values adjusted during training. Modern language models contain billions or even trillions of parameters, with larger models requiring more computational resources for both training and inference. Model size directly impacts memory and processor requirements, as well as energy consumption.

AI model benchmarks serve as standardized tests that allow researchers to measure and compare AI model performance across different tasks and capabilities. These evaluation frameworks assess models on specific domains like language understanding, mathematical reasoning, coding ability or factual knowledge, providing quantitative scores that enable systematic comparison between different models and training algorithms. While benchmarks do not enable true verification of accuracy (the way we can verify whether a list of numbers is indeed correctly sorted), they have become essential for tracking progress in AI development.

C. AI Model Training: Is Larger Always Better?

The dominant architecture powering modern AI application is the transformer.[2] These algorithms process a sequence of data (like text) by learning relationships between its elements (like words). Recent developments have increased the algorithmic efficiency of transformer-based approaches.

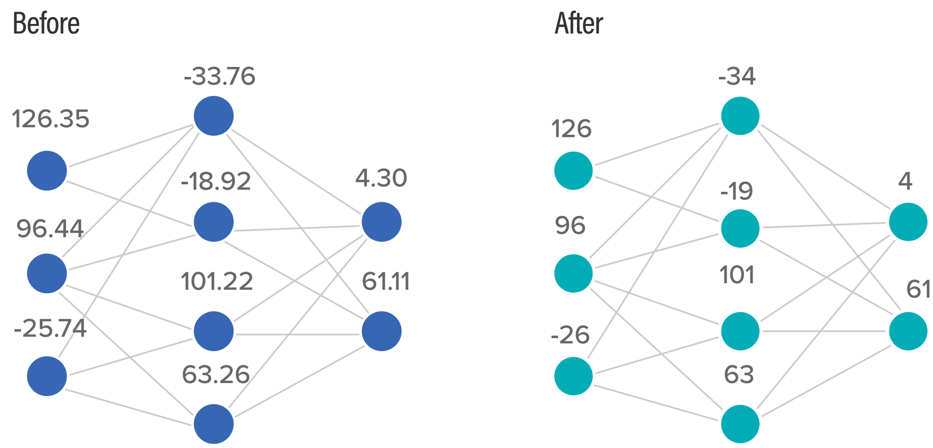
State space models (SSMs), like Mamba, process sequences more efficiently than transformers by maintaining a compressed "state" that captures relevant history, which lowers memory use.[3] Mixture of experts (MoE) architectures contain multiple specialized sub-networks ("experts") but activate only a subset for each input, dramatically increasing model capacity while keeping computational costs manageable during inference.[4] While SSMs show promise but limited adoption (notably AI21 Labs' Jamba models combining Mamba with MoE architectures)[5], MoE architectures have achieved widespread adoption among leading AI labs. GPT-4 is rumored to be MoE-based, as are the recently proposed DeepSeek-v3 and R1 models, while major 2024 releases from Mistral, xAI and Tencent employ MoE architectures.[6] Modern AI training has evolved beyond simply making models process large datasets; it now focuses on making models learn faster, use less memory, and run more efficiently.

One strategy for increasing computational training speed involves using lower-precision values in training and model construction. Mixed-precision training, for instance, accelerates training by performing operations in half-precision format, which uses fewer bits than the gold standard of 32-bit floating-point representation.[7] This approach reduces memory requirements and accelerates training while maintaining accuracy by switching to higher-precision numbers only when needed. Quantization, by contrast, speeds up inference by applying the same precision-reduction strategy to the models themselves.[8] (See Figure 2.2-2.) By switching weights and activation values from 32-bit to 8-bit or even 4-bit formats, quantization can reduce model size by up to 75%, with little to no drop in performance.[9] Another strategy to improve model efficiency is to downsize the model altogether. Distillation accomplishes this by training smaller models to mimic the behavior of larger ones.

Reinforcement learning enables more efficient training by allowing models to learn optimal reasoning patterns without requiring massive-supervised datasets. DeepSeek R1-Zero demonstrated that purely reinforcement learning-based training from a base model can achieve performance comparable to much larger traditionally trained models, while distilled versions show that sophisticated reasoning capabilities can be compressed into models as small as 1.5B parameters, outperforming conventional models many times their size.[10]

Recent innovations in AI architectures inspired by neuroscience and cognitive science have shown promise in outperforming the efficiency of current AI models. Google's Titans, which mimic human memory by combining a neural long-term memory module with attention mechanisms, outperformed both transformers and modern linear recurrent models on long-sequence processing tasks such as language modeling and common-sense reasoning.[11] Artificial Kuramoto Oscillatory Neurons (AKOrN) improve tasks such as unsupervised object discovery and adversarial robustness by replicating neuron binding synchronization dynamics.[12]

Figure 2.2-2. Quantization.



These strategies are under active development, and whether they can match the performance of higher-precision models in real-world usage remains under debate. For applications in which accuracy is paramount—such as medical diagnosis or autonomous driving—even minimal performance degradation may be unacceptable, and not all hardware has optimized support for low-precision operations, limiting deployment options.[13]

D. AI Inference: How "Reasoning" Changed Dynamics

While training efficiency focuses on one-time computational costs, inference efficiency affects every use of an AI model. Increasing the efficiency of AI inference can drive significant gains in energy efficiency, making AI deployment more sustainable and cost-effective across diverse applications. However, a recent innovation has complicated this reality beyond what this statement would suggest.

AI "reasoning" models allocate extra computational resources during inference, allowing them to reason through multiple potential responses before selecting the best answer. While this initially appears to increase energy usage, smaller models enhanced with optimized test-time compute can outperform models up to 400x larger that do not use additional computation at test time, ultimately reducing overall energy consumption.[14]

Additional innovations promise opportunities for further energy savings. Speculative decoding computes several words in parallel using a smaller "draft" model to predict likely next words, which are then verified by the main model, achieving 2x speed improvements.[15] Pruning large language models removes unnecessary parts to make them smaller and more efficient. Researchers have demonstrated that after removing 20% of the parameters, the pruned model maintains 94.97% of the performance of the original model.[16]

This computational redistribution fundamentally alters the energy requirements of AI deployments. Reasoning models have shifted the balance of "expensive training/cheap usage" dramatically: training is a one-time high-cost event, whereas inference costs accumulate over time and can surpass training costs if the model is used extensively.[17] This shift creates a new calculus for AI deployment in which organizations must weigh whether to invest heavily in training massive models that run efficiently or to deploy smaller reasoning-enhanced models that think harder during each query but may ultimately cost more at scale.

E. Reducing Emissions Through Flexible AI Computation

Beyond optimizing individual algorithms, data centers can achieve substantial efficiency gains by strategically timing and placing AI workloads. This operational flexibility represents a different layer of efficiency optimization. To improve both energy efficiency and carbon efficiency, established techniques such as checkpointing and restarting mechanisms, coupled with distributed and flexible computation, play key roles.

i. Checkpointing and restarting

Checkpoint/restart (C/R) technology is a critical foundation for fault tolerance and computational process management. Checkpointing takes snapshots of the system state at intervals, which facilitates recovery from node failures by restoring jobs to the last captured system state. The ability to restart systems from various checkpoints can then be leveraged to break down large computational tasks. Instead of executing long-running batch jobs in one go, tasks can be suspended, redistributed to different data centers, and preemptively scheduled to facilitate load balancing. C/R enables both spatial and temporal compute flexibility by allowing jobs to pause, save state, and resume elsewhere or later without losing completed work.

Originally conceived during the early development of high-performance computing (HPC), C/R technology was designed to solve problems of input/output operation speed, parallelism, and resilience.[18] HPC clusters running large-scale simulations faced risks from bugs or hardware failures that could erase hours or days of computation. Distributing the computation led to faster results, while checkpointing allowed developers to restart computation from the last snapshot.

Checkpointing mechanisms, once aimed at improving computational throughput and robustness, now have deep implications for spatial and temporal compute flexibility in modern, carbon-aware AI workflows. For AI training workloads—which operate "offline" without user-facing latency (delay between user input and AI-generated output) requirements—checkpointing allows computation to be distributed across low-carbon regions and scheduled during renewable energy peaks.[19] A large language model training job, for example, can run when renewable power is abundant, pause when grid carbon intensity rises, and restart hours later or in a different data center entirely. The system preserves all progress while optimizing for emissions reduction.

Energy efficiency, which can be defined as the amount of useful computation per watt, is necessary but not sufficient. Carbon efficiency, by contrast, depends not only on how much energy is used but also on what kind of energy powers computation and whether it yields higher emissions elsewhere on the grid.

ii. Spatial and temporal compute flexibility

Computation can be operationally distributed across different locations, taking advantage of variation in energy cost, carbon intensity and infrastructure efficiency. For instance, electricity grids in regions like the hydropower-heavy Pacific Northwest or wind-rich Northern Europe offer lower carbon footprints per unit of compute. Certain data centers are more energy efficient than others due to newer information technology (IT) equipment (see Chapter 2.1), optimized cooling systems (Chapter 2.3), or overall design maturity. Some locations may have surplus renewable energy that would otherwise be curtailed. By dynamically routing jobs to these locations, especially for workloads like AI training that do not require immediate user feedback, systems can reduce their net carbon emissions significantly without sacrificing performance.

Time-based flexibility aligns computation with renewable energy availability on the grid. This is especially effective in regions with predictable renewable peaks, such as solar energy in California, which is abundant during the daytime. Some academic and hyperscale data centers already schedule jobs to run during daylight hours to capitalize on this pattern. Unlike inference or real-time services, training workloads are throughput-driven, making them ideal candidates for time-shifted execution.

The implementation of compute flexibility differs significantly between operational compute-shifting and strategic data center siting. Operational compute-shifting involves moving workloads between existing facilities through automated systems that make sub-hourly decisions based on real-time grid conditions and data transfer costs. These systems must balance the carbon benefits against the latency and bandwidth constraints of moving large datasets, often limiting shifts to workloads with modest data requirements or pre-positioned datasets. In contrast, data center location selection operates on multi-year timescales, incorporating renewable energy potential alongside traditional siting factors like land availability, water resources, air quality permits and network connectivity. Both approaches require a sophisticated understanding of grid marginal emissions—the actual emissions impact of adding or removing load at specific times and locations. Without accounting for marginal emissions rather than average grid mix, compute-shifting efforts risk increasing carbon intensity by inadvertently displacing cleaner generation sources, underscoring the need for real-time emissions data in scheduling decisions.

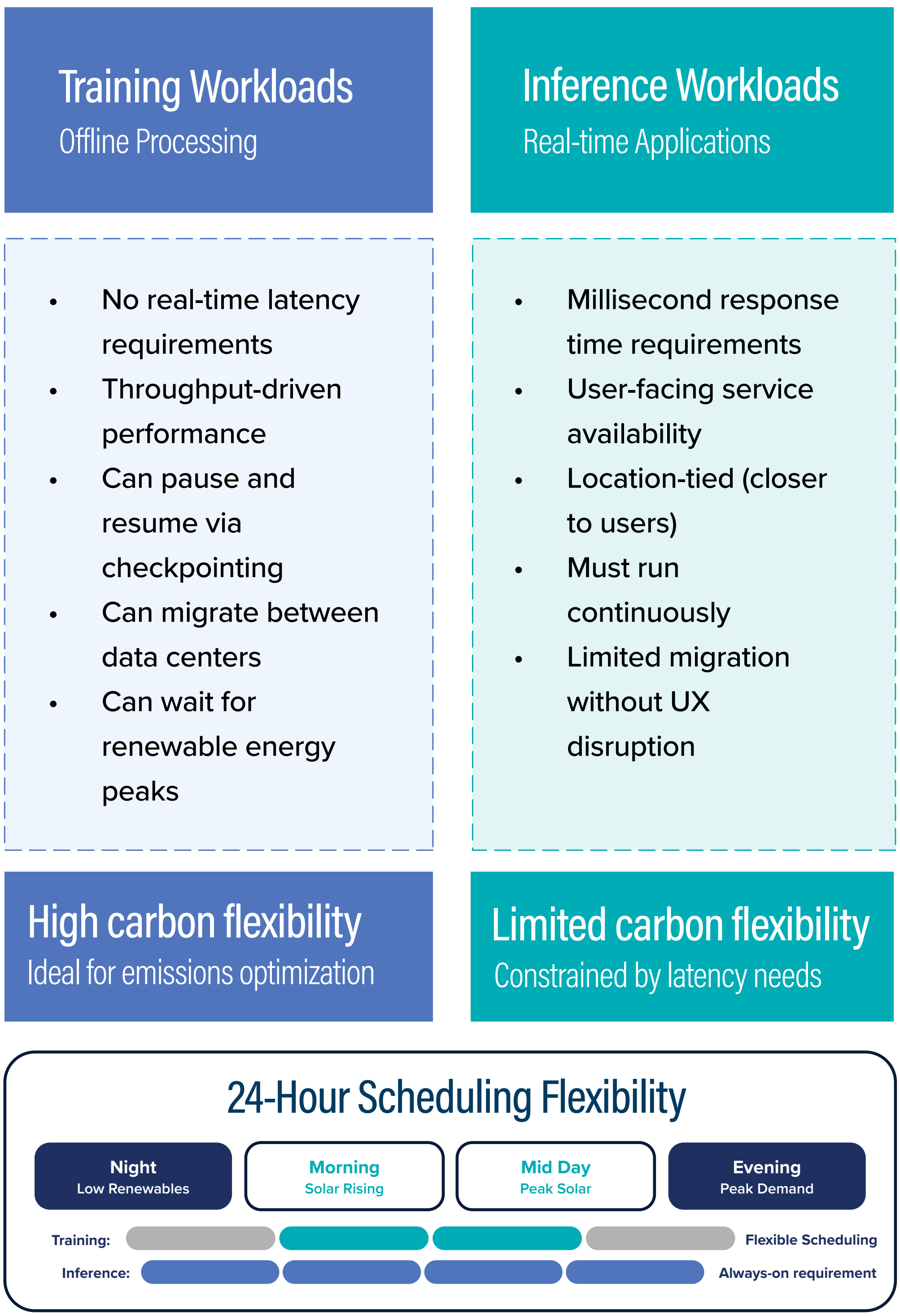
By contrast, AI inference workloads serve real-time applications like chatbots and recommendation systems that require millisecond response times. With comparatively lower data usage, AI inference workloads have greater spatial flexibility than the training phase, as computations can be transferred from busy data centers to those with more capacity. However, latency constraints demand immediate feedback, which greatly limits the temporal flexibility available to training workloads. Even for tasks with flexibility in response times—such as the “research” activities many leading AI companies offer—inference generally cannot be paused and redistributed elsewhere over a longer period without interrupting user experience. (See Figure 2.2-3.)

Data centers typically try to maintain constant availability for numerous other latency-sensitive services—video streaming, online banking, e-commerce transactions, email, and gaming—which are similarly more challenging to pause and shift among multiple data centers without disrupting user experience. While training can easily wait for cleaner energy, inference is often expected to run continuously to maintain service availability, thus limiting opportunities for carbon-aware scheduling without degrading user experience. These efficiency techniques, from algorithmic improvements to flexible scheduling, face new challenges as AI applications evolve. Emerging trends like autonomous agents are reshaping the efficiency landscape entirely.

F. An Outlook on AI Software Efficiency

Modern AI systems are increasingly contributing to data center computations. Autonomous agents and coding assistants will likely increase inference costs. Recent AI agent trends are driving substantial inference demand through autonomous workflows, multi-agent systems, and continuous reasoning. Adoption of AI agents for 2% of business tasks is projected to drive a 15% increase in inference compute.[20] AI agents are software applications that can independently complete multi-step tasks by using various tools and making decisions along the way. Agents perform extended chains of operations—planning, tool usage, environment interaction, and iterative problem-solving—requiring hundreds of inference calls per task compared with single-shot responses. Multi-agent frameworks multiply this effect as agents collaborate, negotiate and coordinate through continuous communication.

Figure 2.2-3. Flexibility of training workloads and inference workloads.



Agent workflows are expected to consume up to 100x more inference compute than traditional chatbot interactions.[21] These workflows drive a combination of always-on operations, where unlike human-initiated queries, agents run continuously for monitoring, scheduling and completing automated tasks. As agents become more capable, organizations that deploy them across more processes will drive growth in compute demand.

Box 2.2-1

Opportunities and challenges in AI energy efficiency research

AI researchers are focusing on three strategies to improve energy efficiency: (1) shorter outputs—compressing lengthy reasoning chains into concise yet effective responses; (2) smaller models—developing compact language models with strong reasoning capabilities through techniques such as knowledge distillation, model compression, and reinforcement learning; and (3) faster responses—designing efficient decoding strategies to accelerate inference.[22]

The effectiveness of inference optimizations is highly sensitive to workload geometry, software stacks and hardware accelerators.[23] For example, techniques like speculative decoding are beneficial only at low batch sizes, while mixture-of-experts models sometimes incur higher energy costs despite similar active parameters. Although appropriate application of relevant inference efficiency optimizations can reduce total energy use by up to 73% from unoptimized baselines, naive energy analyses based on floating-point operations or theoretical GPU utilization can significantly underestimate real-world energy consumption.

G. Barriers and Risks

i. Limited expertise pool

AI energy efficiency requires specialized knowledge spanning computer science, energy systems and carbon accounting—a narrow field given the complexity involved in measuring AI efficiency versus traditional algorithms.

ii. Benchmark limitations

Unlike traditional software where correctness is verifiable (as with the sorting algorithm example), AI systems lack clear "correct" outputs. Current benchmarks inadequately measure "higher-order" reasoning tasks, making efficiency comparisons challenging and unreliable.

iii. Development pace outstripping efficiency considerations

Rapid advances (DeepSeek V3/R1, transformer innovations, MoE architectures) may prioritize performance over energy optimization. The 100x increase in agent inference demand could materialize before efficiency solutions are broadly implemented. Model oversimplification risks: Efficiency techniques like quantization and distillation risk removing safety guardrails built into larger models. As in medical and autonomous driving applications, even minimal performance degradation may be unacceptable in critical systems.

iv. Model oversimplification risks

Efficiency techniques like quantization and distillation risk removing safety guardrails built into larger models. As in medical and autonomous driving applications, even minimal performance degradation may be unacceptable in critical systems.

H. Recommendations

1. *Educational institutions should* ***prepare the next generation of computer scientists and policymakers with the conceptual tools to understand and advance software efficiency****.*
    - ***Create cross-disciplinary curricula*** *on algorithmic efficiency, carbon-aware computing and AI systems engineering.*
    - ***Include energy literacy in computer science and AI degree programs****, covering both micro-level optimizations (e.g., quantization) and macro-level system design (e.g., flexible compute).*
    - ***Develop policy bootcamps or executive courses for non-technical audiences****, such as civil servants, journalists and business leaders, on emerging AI compute trends and their implications for sustainability.*
    - ***Fund open-access software efficiency toolkits and benchmarks****, especially those focusing on inference-time efficiency and emissions transparency.*

2. *Data center operators should* ***incorporate software-aware workload management and emissions-aware operations*** *into core planning.*

- ***Partner with utility companies and cloud platforms*** *to offer real-time grid carbon intensity data and renewable energy forecasts for intelligent job scheduling.*
- ***Offer time-delayed training products and pricing schemes*** *that incentivize shifts to lower-carbon AI training workflows.*
- ***Provide application programming interfaces (APIs)*** *that expose real-time energy and emissions data for AI workloads, enabling software developers to optimize code based on environmental impact.*
- ***Adopt software-aware procurement criteria*** *that favor AI models and systems with verifiable efficiency gains or emissions-conscious design.*

3. *Software development companies* ***build efficiency into the core of AI model design, training and deployment.***
    - ***Develop APIs and platforms that report model energy use,*** *offering customers transparency and options for greener usage.*
    - ***Collaborate with academia to publish standardized benchmarks and best practices for evaluating software energy use****, not just performance or accuracy.*

4. *Regulatory bodies should* ***establish policies that ensure AI progress aligns with the public interest, energy constraints and climate goals.***
    - ***Incorporate software efficiency and emissions data disclosure requirements into AI governance frameworks****, especially for high-volume models or widely deployed systems.*
    - ***Mandate transparent compute and energy reporting*** *for AI systems procured with public funding or deployed in sensitive sectors (e.g., health, education).*
    - ***Develop incentives for energy-efficient AI systems,*** *such as research and development tax credits or public procurement preferences.*

5. *Educational institutions should* ***prepare the next generation of computer scientists and policymakers with the conceptual tools to understand and advance software efficiency.***
    - ***Create cross-disciplinary curricula*** *on algorithmic efficiency, carbon-aware computing and AI systems engineering.*

I. References

1. David Sandalow, Colin McCormick, Alp Kucukelbir, Julio Friedmann, Michal Nachany, Hoesung Lee, Alice Hill, Daniel Loehr, Matthew Wald, Antoine Halff, Ruben Glatt, Philippe Benoit, Kevin Karl et al. Artificial Intelligence for Climate Change Mitigation Roadmap (Second Edition) (ICEF Innovation Roadmap Project, November 2024); https://doi.org/10.7916/2j4p-nw61 (2024).

2. NVIDIA. Large Language Models Explained (NVIDIA glossary); NVIDIA, Santa Clara, California, https://www.nvidia.com/en-us/glossary/large-language-models/ (Accessed August 2025).

3. Albert Gu & Tri Dao. Mamba: Linear-Time Sequence Modeling with Selective State Spaces. arXiv:2312.00752 (2023). https://doi.org/10.48550/arXiv.2312.00752.

4. Noam Shazeer, Azalia Mirhoseini, Krzysztof Maziarz, Andy Davis, Quoc Le, Geoffrey Hinton & Jeff Dean. Outrageously Large Neural Networks: The Sparsely-Gated Mixture-of-Experts Layer. arXiv:1701.06538 (2017). https://doi.org/10.48550/arXiv.1701.06538.

5. Chris McKay. TII Unveils Falcon Mamba 7B, A New Open-Source State Space Language Model; Maginative, Boston, Massachusetts, https://www.maginative.com/article/tii-unveils-falcon-mamba-7b-a-new-open-source-state-space-language-model/ (2024).

6. Asif Razzaq. List of Large Mixture of Experts (MoE) Models: Architecture, Performance, and Innovations in Scalable AI Solutions; Marktechpost Media, Inc., Tustin, California, https://www.marktechpost.com/2024/11/16/list-of-large-mixture-of-experts-moe-models-architecture-performance-and-innovations-in-scalable-ai-solutions/ (2024).

7. NVIDIA. Train With Mixed Precision (User's Guide | NVIDIA Docs; DA-08617-001_v001); NVIDIA Corporation, Santa Clara, California, https://docs.nvidia.com/deeplearning/performance/mixed-precision-training/index.html (2024).

8. Meegle. Quantization Vs Mixed Precision; Meegle, Culver City, California, https://www.meegle.com/en_us/topics/quantization/quantization-vs-mixed-precision#understanding-the-basics-of-quantization-and-mixed-precision (2025).

9. Robert McMenemy. How to 8-bit quantize large models using bits and bytes; AI Accelerator Institute (AIAI), San Francisco, California, https://www.aiacceleratorinstitute.com/how-to-8-bit-quantize-large-models-using-bits-and-bytes/ (2025).

10. DeepSeek-AI, Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Ruoyu Zhang, Runxin Xu, Qihao Zhu, Shirong Ma, Peiyi Wang, Xiao Bi, Xiaokang Zhang, Xingkai Yu et al. DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning. arXiv:2501.12948 (2025). https://doi.org/10.48550/arXiv.2501.12948.

11. Ali Behrouz, Peilin Zhong & Vahab Mirrokni. Titans: Learning to Memorize at Test Time. arXiv:2501.00663 (2024). https://doi.org/10.48550/arXiv.2501.00663.

12. Takeru Miyato, Sindy Löwe, Andreas Geiger & Max Welling. Artificial Kuramoto Oscillatory Neurons. arXiv:2410.13821 (2024). https://doi.org/10.48550/arXiv.2410.13821.

13. Ashish Abraham. The Ultimate Handbook for LLM Quantization; Towards Data Science, Insight Media Group, LLC, Toronto, Ontario, Canada, https://towardsdatascience.com/the-ultimate-handbook-for-llm-quantization-88bb7cb0d9d7/ (2024).

14. Runze Liu, Junqi Gao, Jian Zhao, Kaiyan Zhang, Xiu Li, Biqing Qi, Wanli Ouyang & Bowen Zhou. Can 1B LLM Surpass 405B LLM? Rethinking Compute-Optimal Test-Time Scaling. arXiv:2502.06703 (2025). https://doi.org/10.48550/arXiv.2502.06703.

15. Charlie Chen, Sebastian Borgeaud, Geoffrey Irving, Jean-Baptiste Lespiau, Laurent Sifre & John Jumper. Accelerating Large Language Model Decoding with Speculative Sampling. arXiv:2302.01318 (2023). https://doi.org/10.48550/arXiv.2302.01318.

16. Xinyin Ma, Gongfan Fang & Xinchao Wang. "LLM-pruner: on the structural pruning of large language models" in Proceedings of the 37th International Conference on Neural Information Processing Systems, New Orleans, LA, USA. 21702-21720 (Article No. 21950), https://dl.acm.org/doi/10.5555/3666122.3667072 (2023).

17. Kyle Aubrey. How the Economics of Inference Can Maximize AI Value; NVIDIA, Santa Clara, California, https://blogs.nvidia.com/blog/ai-inference-economics/

18. Kathryn Mohror Adam Moody, Franck Cappello. SCR Framework: Accelerating Resilience and I/O for Supercomputing Applications (Prepared for the 2019 R&D 100 Award Entry); Lawrence Livermore National Laboratory (LLNL), Livermore, California, https://ipo.llnl.gov/sites/default/files/2020-01/SCR-RD100final_version2.pdf (2019).

19. Andrew A Chien, Liuzixuan Lin, Hai Nguyen, Varsha Rao, Tristan Sharma & Rajini Wijayawardana. "Reducing the Carbon Impact of Generative AI Inference (today and in 2035)" in Proceedings of the 2nd Workshop on Sustainable Computer Systems, Boston, MA, USA. Article 11, https://doi.org/10.1145/3604930.3605705, (2023).

20. Ross Sandler Tom O'Malley, Trevor Young. The next wave of AI: Demand and adoption; Barclays, London, United Kingdom, http://www.ib.barclays/our-insights/3-point-perspective/the-next-wave-of-AI-demand-and-adoption.html (2024).

21. Kif Leswing. Nvidia CEO Huang says AI has to do '100 times more' computation now than when ChatGPT was released; CNBC, Englewood Cliffs, New Jersey, https://www.cnbc.com/2025/02/26/nvidia-ceo-huang-says-next-generation-ai-will-need-more-compute.html (2025).

22. Sicheng Feng, Gongfan Fang, Xinyin Ma & Xinchao Wang. Efficient Reasoning Models: A Survey. arXiv:2504.10903 (2025). https://doi.org/10.48550/arXiv.2504.10903.

23. Jared Fernandez, Clara Na, Vashisth Tiwari, Yonatan Bisk, Sasha Luccioni & Emma Strubell. Energy Considerations of Large Language Model Inference and Efficiency Optimizations. arXiv:2504.17674 (2025). https://doi.org/10.48550/arXiv.2504.17674.

2.3 Cooling Technologies

Alexis Abramson

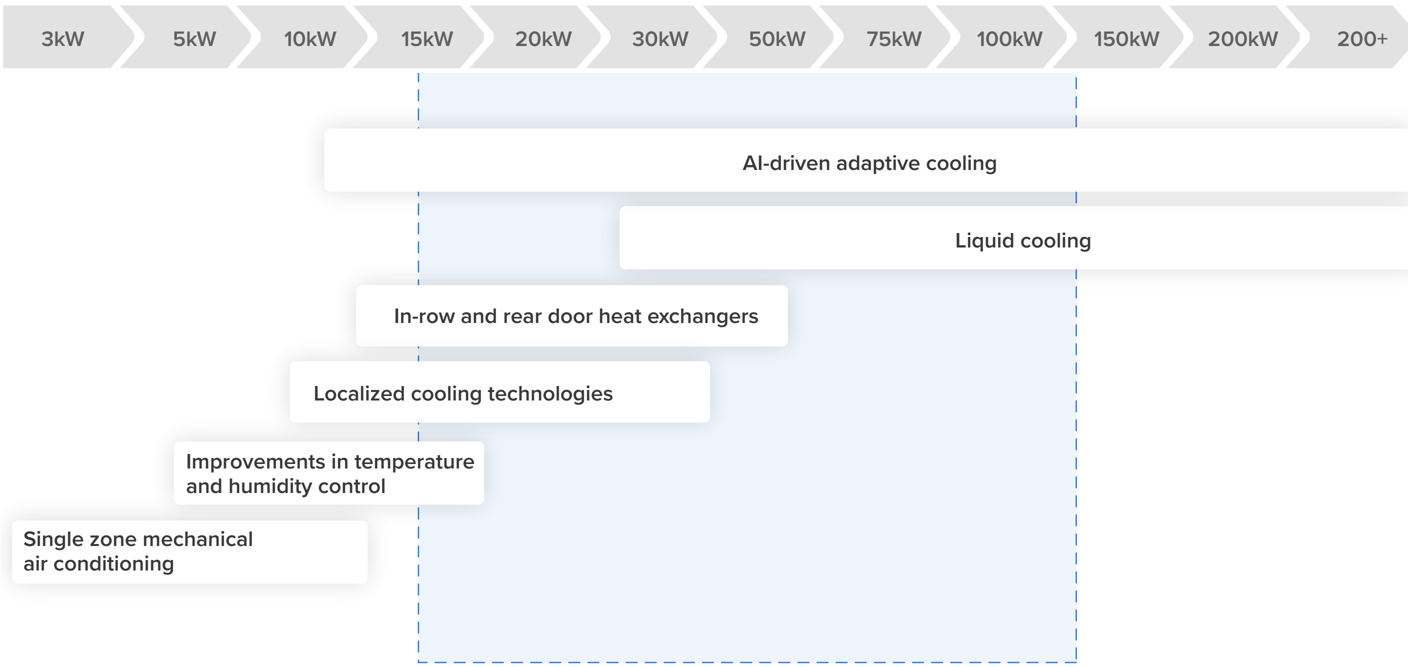
A. Overview

i. History of data center cooling

Cooling infrastructure can be a major challenge for data centers, especially as rising power densities from artificial intelligence (AI), high-performance computing (HPC) and graphics processing unit (GPU)-intensive workloads drive the need for more advanced cooling technologies. The history of data center cooling is closely tied to the evolution of computing itself. As computer systems became more powerful, their thermal output increased, necessitating more sophisticated and efficient cooling methods. The journey from basic ventilation to advanced liquid and AI-optimized cooling reveals a story of continuous innovation driven by performance, reliability and energy efficiency. The American Society for Heating, Refrigeration, and Air Conditioning Engineers (ASHRAE) has been publishing and updating guidelines for data center cooling since 2004 and every 4-6 years thereafter.[1]

Cooling data centers has a lot in common with cooling other buildings, but there are some important differences. Typical residential and commercial buildings gain heat through windows and walls, appliances (e.g., refrigerators), and even people. Circulating cool air through the building via a central air conditioning system (i.e., treating the entire space as a single thermal zone) is the typical way to remove heat. A commonly employed rule of thumb for sizing building cooling systems is 1 ton of cooling for every 400-500 ft^2, which is equivalent to ~8 W/ft^2. In contrast, data centers have no windows, minimal occupancy and highly insulated walls. Their primary heat sources are the computer servers themselves—extremely localized loads that can exceed 100 W/ft^2. (See Chapter 1(A)(ii)(a) of this Roadmap.) This concentrated heat makes broad air circulation inefficient. Instead, cooling systems must be substantially over-provisioned and precisely engineered to prevent local "hotspots" within the equipment. In the mid-20th century, early mainframe computers were housed in large

Figure 2.3-1. Understanding direct-to-chip cooling in HPC infrastructure: A deep dive into liquid cooling.



rooms and cooled primarily through ambient air and basic air conditioning systems.[3] Using direct expansion (DX), these systems pulled warm air from the single thermal zone, passed it over a refrigerant coil, and rejected the heat via condensers. Raised floors are used to distribute cool air and ceiling return vents for hot air extraction. As server density increased in the 1990s and 2000s, the inefficiencies of room-level cooling became more apparent. More sophisticated cooling systems, including Computer Room Air Conditioners (CRAC) and later Computer Room Air Handlers (CRAH), allowed for better temperature control and humidity regulation.

In the mid-2000s and into the 2010s, the proliferation of high-density server racks led to adoption of more localized cooling strategies. These servers pack a relatively high compute density, thereby requiring a significantly higher power draw. In high density racks, this power draw can vary from 15-30 kW, while in ultra high-density racks it can reach 100+ kW, leading to a much greater heating load. During this time period, computational fluid dynamics (CFD) modeling became a commonly used tool for facility-level design to simulate airflow and temperature distribution. CFD enables data center designers to optimize cooling strategies, prevent hotspots and improve overall energy efficiency. In-row cooling and rear door heat exchangers (RDHx) became popular for targeting heat removal closer to the source. These heat exchangers can be mounted on the rear door of server racks to remove heat directly before the air can re-enter the data center environment. While they have not captured a significant market share, this technology has emerged and is evolving as an attractive option to address increasingly dense and power-intensive server racks. (See Chapter 2.4.)

With the advent of AI, HPC, and GPU-heavy workloads over the past decade, traditional air cooling approaches have begun to hit their limits. Liquid cooling—through direct-to-chip cold plates, immersion cooling and two-phase systems—has seen an increase in the marketplace. These systems offer superior heat transfer and are often more energy efficient, especially for workloads associated with high-density racks where power densities range from 30-100+ kW per rack.[4]

The selection of specific cooling systems depends not only on data center needs but also on local conditions, enabling them to be more climate responsive and to optimize for performance, costs, efficiency, thermal management, water use and reliability. Globally, the selection of cooling systems can be particularly influenced by geographic climate factors:

- Hot-humid climates (e.g., Singapore): Employ a combination of robust air conditioning, often supported by chilled-water systems. Advanced systems, such as desiccant dehumidification and various forms of liquid cooling, have increasingly been adopted.

- Hot-dry climates (e.g., Phoenix): Rely on evaporative cooling, which is highly efficient and often used in conjunction with traditional air cooling as a backup. Also employed is adiabatic cooling, which uses the evaporative cooling effect to pre-cool intake air. Free cooling (see Figure 2.3-1) may be integrated in the system to take advantage of cooler outdoor temperatures. The availability of water and its potential usage in these systems is an important consideration.[5]

- Temperate climates (e.g., Frankfurt): Deploy hybrid systems that integrate free cooling to leverage moderate outdoor temperatures combined with conventional air cooling mechanical systems optimized for seasonal shifts.

- Cold climates (e.g., Sweden): Adopt free cooling via air-side or water-side economization used most of the year.[6] Ample waste heat generated by servers has also been used for heat recovery applications. (See Chapter 2.4.)

Box 2.3-1

Free cooling aka economization

In cooler climates, free cooling, also known as economization, can eliminate much of the need for energy-intensive, air-cooled mechanical systems. Leading data centers in Scandinavia, Canada and the US Pacific Northwest report significant energy savings using air-side economizers, as well as natural or readily available sources of cold water, to absorb and dissipate the heat.[7] While attractive in theory, some have found free cooling challenging to implement. For example, Microsoft's Project Natick and Subsea Cloud's "Jules Verne" pod have demonstrated how to leverage ocean water for passive cooling, but the project was halted in 2024.[8-11] Meanwhile, in orbit, space data center initiatives, such as the Hewlett Packard Enterprise (HPE) Spaceborne Computer-2 on the International Space Station, utilize the naturally cold environment of space to reject heat; however, challenges remain due to the large surface area required for the radiative heat transfer.[12]

Today, air cooling still dominates in data centers-and is the default choice for many legacy environments because it is widely understood, relatively low cost to install and maintain, and remains effective for facilities with moderate power densities. Even more, new demands of modern data centers are driving an urgent need for more intelligent, adaptable and scalable cooling innovations to sustain future performance and efficiency.

ii. Cooling and its impact

Nearly all electricity consumed by servers is ultimately converted into heat. As server densities rise, so does the cooling load (the amount of heat that must be removed), driving the need for more advanced cooling technologies. Without effective cooling, thermal stress can reduce performance and hardware lifespan, making efficient thermal management an imperative for both reliability and sustainability.

Many efficient cooling technologies, such as evaporative cooling, dramatically reduce electricity use but may also require substantial water, depending on the cooling mechanism.[3] Air conditioning systems that rely on mechanical DX refrigeration use little or no water but demand higher energy input, increasing carbon emissions when powered by fossil fuels. Optimizing for one resource—such as water or energy—can often increase the burden on another.

Power usage effectiveness (PUE) and water usage effectiveness (WUE) provide an indication of the performance of cooling technologies in data centers. PUE is a data center's total energy use divided by the energy its information technology (IT) equipment uses. A lower PUE (closer to 1.0) indicates more efficient cooling and operations. Efficient operators now report PUEs below 1.2, but the global average remains closer to 1.5.[13] WUE represents a data center's water use divided by the energy its IT equipment uses. WUE is typically expressed in liters per kilowatt-hour (L/kWh). Highly efficient data centers can achieve WUE values as low as 0.2 L/kWh or even close to 0 L/kWh, which effectively means the facility uses no water for cooling purposes.[14] Data centers that rely heavily on evaporative cooling are likely to have the highest WUEs, potentially up to 2.5 L/kWh.[15] Target WUEs for a particular data center should be considered in conjunction with PUE, local environmental conditions and water availability. As sustainability becomes a priority, operators and designers must consider both resources to fully appreciate the overall impact.

Cooling systems also contribute to carbon emissions through their use of refrigerants. Some refrigerants, such as hydrofluorocarbons (HFCs), are greenhouse gases with high global warming potential (GWP) and must be managed to avoid leakage or improper disposal. In fact, so-called F-gas leakage-the unintended release of fluorinated gases-can equate to massive amounts of carbon emissions. Various global regulations are already mandating the phase-down and eventual phase-out of high-GWP F-gases in equipment. This is pushing data centers toward lower-GWP alternatives, improved leak-detection protocols and, in some cases, wholesale replacement of legacy systems.[16,17] (See Chapter 3.1-Scope 1 Emissions.)

Lifecycle analyses of data center cooling are now more commonly conducted for quantifying not just operational energy use but also the emissions and resource consumption associated with manufacturing, deploying and disposing of cooling equipment. Such assessments have shown that advanced cooling methods like cold plates and immersion cooling (see below) can reduce greenhouse gas emissions by 15-21% over the entire lifecycle compared to conventional air cooling systems.[18] So while the increasing power demands of modern computing is essentially forcing the industry to adopt more efficient liquid cooling methods, these technologies are inherently better for the climate overall. This makes liquid cooling a compelling and necessary shift for the future of digital infrastructure. A study by Microsoft (2023) confirmed that while liquid cooling systems have higher manufacturing impacts (due to pumps, cold plates and materials), they can offer lower total lifecycle emissions in high-density environments by enabling energy savings and improved PUE.[18] The study also emphasized the need to account for cooling infrastructure replacement cycles, refrigerant GWP and end-of-life impacts when comparing technologies.

Sustainable data center cooling now requires multi-metric optimization, balancing energy, water, carbon and local resource availability, such as water scarcity (if applicable), grid capacity and renewable energy availability. This means moving beyond simply achieving the lowest PUE to performing holistic and complex assessments of energy, water, carbon, and the local environmental and social context of these systems.

B. Current Cooling Technologies and Associated Enhancements

Table 2.3-1. *Data center cooling systems and PUEs.*[6,19-25]

Data Center Cooling System	Typical PUE Range	Notes
Air-Based Cooling		
CRAC units with DX (no enhancements)	1.8–2.5	Small to medium data centers
CRAH units with chilled water (no enhancements)	1.7–2.2	Medium to large data centers
CRAC/CRAH + hot/cold aisle containment	1.5-2.0	May include hot or cold containment or combined
CRAC/CRAH + economizers	1.2-1.6	“Free cooling”; can be air-side or water-side; depends highly on local climate
CRAC/CRAH + modern enhancements	1.04–1.4	May include variable speed drives in compressors; enhanced controls
CRAC/CRAH + adiabatic cooling/pre-cooling	1.2–2.0	Depends on efficiency of primary system and local climate
Direct evaporative cooling	1.1-1.3	Highly dependent on local climate
Liquid-Based Cooling		
Direct-to-chip liquid cooling	1.05-1.2	CRAH still required for heat removal to environment
Single-phase immersion cooling	1.02–1.05	No server fans; no air-based systems
Two-phase immersion cooling	1.01–1.03	Relies on thermosiphon effect rather than pumps

i. Air-based cooling technologies

Air remains the most widely used cooling medium in data centers due to its simplicity, cost-effectiveness and compatibility with existing IT infrastructure. But liquid cooling is gaining traction. In a 2024 survey of almost 1000 data center owners and operators, 95% reported relying predominantly on air-cooled systems, but 94% also reported at least some use of liquid cooling, including 17% who reported substantial use of liquid cooling technologies.[26] Over time, air-based cooling strategies have evolved from room-level temperature control to highly localized, intelligent and climate-responsive systems that improve energy efficiency and thermal performance.

Traditional CRAC units and air handling units (AHUs) are foundational components of air-based cooling. CRACs typically use DX systems to cool air, while CRAHs and AHUs circulate chilled air from centralized systems. These systems supply conditioned air—typically cooled and sometimes dehumidified—either through raised-floor plenums or via overhead ductwork, depending on the facility design. In a raised-floor setup, cool air is pushed into the underfloor space and directed upward through perforated or directional vent tiles, which are strategically placed in front of server racks to deliver air directly to equipment inlets. In overhead systems, ductwork distributes cool air from above, typically aimed down into cold aisles.

To manage heat generated by the equipment, the systems rely on return air plenums, often located in the ceiling, to collect hot exhaust air from the back of the server racks. This hot air is then routed back to the cooling units for reconditioning and recirculation. To improve efficiency and prevent mixing of hot and cold air, many modern facilities incorporate hot aisle containment. This strategy encloses the hot aisle—where exhaust air is expelled—in a physical barrier, such as doors or ceiling panels, which channels the hot air directly into the return system. By maintaining a clear separation between hot and cold air streams, hot aisle containment enhances thermal management and allows for higher cooling efficiency.[27]

In temperate climates, air-side economizers allow filtered outside air to cool the data center when ambient temperatures are favorable, minimizing mechanical cooling. This technique, known as free cooling, can yield substantial energy savings and lower PUE. Microsoft was among the first hyperscalers to report successful deployment of air-side economization in production environments, showing that well-controlled airflow and outdoor air integration could support high-performance operations with PUEs below 1.2.[28] At Alibaba's Qiandao Lake data center (2015), a combination of natural ventilation, free cooling and computational fluid dynamics (CFD) modeling reduced energy consumption significantly, especially during cooler months.[29] Free cooling was achievable for over 90% of the year, leading to cooling cost reductions of approximately 80% and a PUE below 1.3.

Data centers have increasingly adopted evaporative and adiabatic cooling as energy-efficient mechanisms to enhance traditional air-based cooling systems. These systems leverage the natural cooling effect of water evaporation to reduce air temperature. In direct evaporative cooling, outside air is humidified and cooled as it passes through a wetted medium, then it is introduced into the data center. Adiabatic cooling, often used in indirect systems, pre-cools air or a heat exchange surface without introducing moisture directly into the data center environment. Both approaches are particularly effective in dry or temperate climates, where low humidity allows for significant cooling potential. Major operators, such as Amazon, Facebook and Google, have deployed evaporative or adiabatic cooling systems in regions like Iowa, Oregon and Utah, achieving lower PUE and reducing dependence on energy-intensive chillers.[30] These methods are often integrated into multi-mode cooling systems that shift between economization, evaporative cooling and mechanical backup based on real-time environmental conditions.

Case Study 1: Meta's MeeFog—pushing the limits of cooling with evaporative precision

In 2011, Meta (then Facebook) broke new ground with its MeeFog evaporative cooling system, a high-efficiency design that leveraged adiabatic cooling to generate an ultrafine water mist to pre-cool incoming outside air. By using high-pressure atomization to saturate air without over-humidifying it, the system enabled Meta to cool data halls without the traditional chillers used in air-based mechanical systems—even in warmer climates.

The result? A dramatic drop in cooling energy use and a PUE as low as 1.06 at facilities like Prineville, Oregon—years ahead of industry norms.[31] Crucially, the MeeFog system still maintained ASHRAE-recommended temperature and humidity ranges, demonstrating that ultra-efficient evaporative cooling could be both high-performing and reliable. Meta is continuously exploring and implementing upgrades to the MeeFog system. These upgrades are likely to push PUE and WUE even lower, while extending the viability of evaporative cooling to more variable climates.

Recent advances in AI/machine learning (ML) enable dynamic, software-driven thermal management, allowing systems to adapt cooling in real-time based on workload and environmental conditions. In 2022, Meta deployed a reinforcement learning system to optimize the performance of air-handling units located in the penthouse—the mechanical space typically situated on the roof of its data centers that houses key heating, ventilation and air conditioning (HVAC) equipment. This led to a 20% reduction in fan energy usage and improved temperature compliance.[32] In collaboration with Trane Technologies, DeepMind published an article showing actual cooling energy reductions of 9-13% in commercial HVAC systems via reinforcement learning.[33] Earlier, a 2016 DeepMind blog reported up to 40% cooling energy savings and a 15% PUE reduction in Google data centers using ML controls.[34]

ii. Liquid cooling technologies

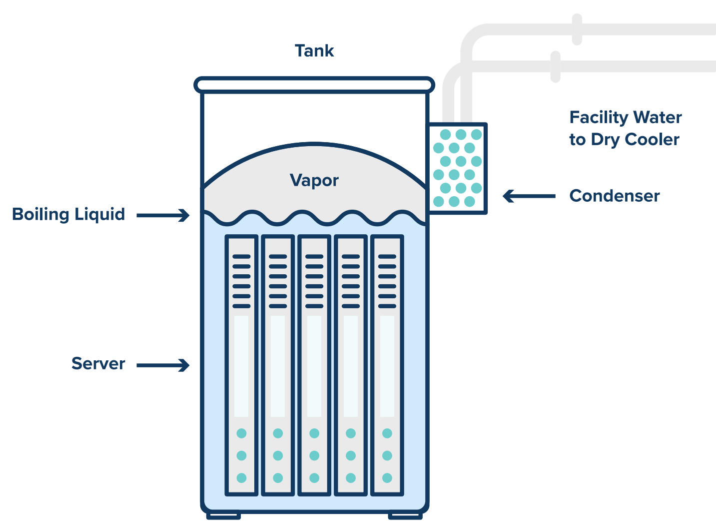
As data centers face rising rack power densities, traditional air cooling methods are struggling to keep pace. Current hyperscale racks now routinely operate above 100 kW. Looking ahead, Nvidia's roadmap anticipates a stepwise progression to approximately 600 kW by 2027 and possibly beyond megawatt-class densities by 2030.[35] A massive new data center project in Texas, backed by OpenAI and Oracle, will be relying on a liquid cooling system to manage the intense heat from its high-density AI servers. This approach involves a continuous circulation of millions of gallons of chilled water, highlighting how hyperscalers are adopting advanced cooling methods to handle extreme computing loads.[36] Sophisticated liquid cooling technologies will remain essential for managing thermal loads efficiently while improving sustainability metrics like PUE and WUE. For most liquid cooling technologies, WUE values are at or near 0.0 since closed-loop systems can be employed. This is because cooling fluids continuously circulate within a sealed, self-contained circuit without being exposed to the outside environment or losing fluid through evaporation or discharge. Thus, in contrast to water-intensive systems, such as evaporative cooling towers, adiabatic coolers, and open-loop water-cooled chillers (which rely heavily on water consumption to reject heat), liquid cooling technologies reduce or even eliminate water use while maintaining efficient thermal management.[5]

Liquid cooling systems in data centers use a range of fluids, including water-based mixtures for indirect cooling methods (e.g., cold plates and rear-door heat exchangers) and dielectric fluids for direct-contact approaches (e.g., immersion and spray cooling), depending on whether electrical isolation is required. These fluids have a much higher thermal capacity than air, and therefore liquid cooling can be thousands of times more effective at heat removal per unit volume than air-based systems.[37,38] The higher heat transfer efficiency can unlock new levels of compute performance in constrained thermal environments where there might be limitations on heat removal.

Innovations in IT infrastructure and liquid cooling are enabling inlet temperatures to be raised to 30-45 °C (up from 18-24 °C), allowing liquid loops to run at hotter temperatures and reducing mechanical cooling requirements.[39,40] Additionally, these modern IT enhancements are enabling cooling systems to reject heat even in hot climates where passive rejection of low temperature heat would not be thermodynamically feasible due to the lack of sufficient temperature differential. Operating at higher loop temperatures also increases the exergy—or useful work potential—of the waste heat, making it easier and more efficient to transfer heat for secondary uses. This opens the door to a wider range of reuse applications, including district heating, industrial processes or absorption cooling. With the addition of a heat pump, producing steam may even be possible. (See Chapter 2.4.)

The two most commonly employed advanced liquid cooling methods for data centers are direct-to-chip (DTC) (Figure 2.3-3) cooling and immersion cooling. DTC cooling uses cold plates or microchannels mounted directly on processors and GPUs to absorb heat, transferring it through the liquid loop to an external heat exchanger or a dry, air-based cooler.[2] DTC has the advantage of integrating with existing racks but the disadvantage of requiring plumbing for cooling distribution. In 2012, Google was one of the first to deploy cold plate DTC technology to support analytics workloads that pushed power densities beyond 20-30 kW, a level that air cooling alone could not handle. Google laid the groundwork for today's generation of liquid-cooled, high-density racks, providing an early real-world demonstration that chip-level cooling could be scalable, safe and cost-effective in hyperscale environments.

Figure 3.3-4. *Immersion Cooling for Data Center Cooling.*



Advances in the DTC sector are accelerating to overcome key technical barriers. For example, ZutaCore's innovative cooling technology, called HyperCool, is a waterless, two-phase, DTC liquid cooling, closed-loop system that uses a non-conductive, dielectric fluid that evaporates directly at the chip level, absorbing heat as it changes phase from liquid to vapor.[41] This approach enables effective cooling of ultra-high-density server racks (up to 250 kW per rack) and supports next-generation chips drawing over 2800 W, while reducing cooling energy use by up to 80%. Another company, Iceotope, offers a solution that combines the compact, targeted efficiency of DTC liquid cooling with the thermal stability and enclosure advantages of tank-based immersion, enabling precise heat removal while protecting IT equipment from environmental contaminants and airflow-related issues.

Case Study 2: DTC with Microsoft and AWS

In 2024, Microsoft began a phased rollout of closed-loop direct-to-chip (DTC) liquid cooling systems at its new AI-focused data centers—making a shift from evaporative cooling in response to the rising heat loads associated with high-density compute.[42] The pilot deployments will be at locations in Phoenix, Arizona and Mount Pleasant, Wisconsin.

Microsoft's goal is to evaluate these systems, considering both energy and water use holistically, to improve overall sustainability while meeting demanding thermal requirements. The deployed system uses a glycol-based coolant in a sealed loop, coupled with dry coolers and adiabatic assist, which pre-cools incoming air using a fine water mist when ambient temperatures allow. While this move resulted in a modest increase in energy consumption and PUE compared to traditional evaporative systems, it reflects a deliberate tradeoff to support long-term water resilience and environmental sustainability of Microsoft's expanding global AI data center fleet.

In 2025, Amazon Web Services (AWS) unveiled one of the most versatile and scalable thermal management strategies to date: a flexible multimodal cooling architecture engineered specifically for next-generation AI workloads. Designed and delivered in just 11 months, the hybrid system uses both air-based cooling and DTC cold plates to extract heat from central processing units (CPUs) and accelerators, circulating a non-evaporative coolant through a sealed-loop system connected to air-based heat rejection units outside the data hall.[43] Unlike immersion systems, this solution maintains compatibility with standard server form factors and supports easy retrofitting across AWS's global infrastructure. It integrates with AWS's building management system for centralized thermal optimization and enables cooling of racks exceeding 80 kW—all while dramatically reducing reliance on water.

AWS reports that their newly launched cooling system can cut cooling energy use by up to 46% during peak cooling periods based on internal comparisons with its legacy system design.[44] While specific PUE values are not disclosed, AWS noted that in 2023, its data centers achieved an average global PUE of 1.15 with the best performing facility reaching 1.04.[45]

In contrast to DTC, immersion cooling (Figure 2.3-4) submerges servers in a non-conductive dielectric fluid that directly absorbs and removes heat, providing uniform cooling to all components. Either the fluid can be circulated through a heat exchanger and returned (single phase immersion), or it can be made to boil at low temperatures on contact with hot components, leading the vapor to rise, condense on a heat exchanger, and return to the bath (two-phase immersion). The technology reduces fan power use to near zero (although pumps are now required to circulate the fluid), supports high rack densities (100-250 kW+), and can extend equipment lifespan by eliminating dust and humidity exposure. And while immersion cooling can deliver superior thermal performance and energy efficiency, its deployment often requires significant modifications to data hall infrastructure, including changes to rack configurations, service workflows and facility layout. As a result, it is typically more practical for use in new builds or purpose-designed expansions. In contrast, DTC liquid cooling is more easily integrated into existing data centers, as it aligns with conventional rack architectures and is therefore better suited for retrofits. Nonetheless, innovation in this space is growing, particularly due to the increasing demands of AI.[46] For example, Green Revolution Cooling (GRC) was one of the first companies to commercialize immersion cooling at scale and has seen deployments across more than 20 countries.[47] Another industry example, Submer, is marketing its single-phase immersion technology using a biodegradable dielectric fluid along with a forced convection heat sink developed in partnership with Intel to cool extremely dense chips, overcoming prior density and heat dissipation barriers.[48]

Many data centers combine liquid and air cooling into hybrid systems, using liquid to manage hotspots (e.g., CPUs, GPUs) and air for bulk heat removal. (See Case Study 2 above.) One additional benefit of the hybrid approach is the extended useful life of air-based cooling infrastructure by enabling it to coexist with targeted liquid systems—minimizing capital disruption. To enhance air-based cooling, rear door heat exchangers (RDHx) are increasingly used in data centers to efficiently cool high-density racks by immediately capturing the hot air that servers expel—known as server exhaust—as it leaves the back of the rack.

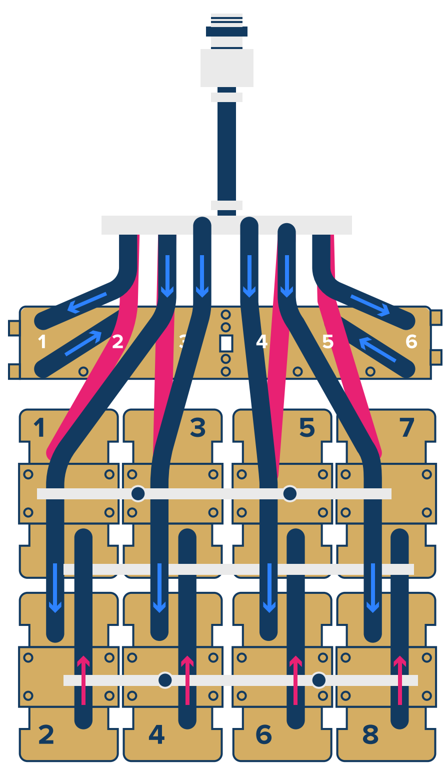


Table 2.3-3. *Direct-to-chip cold plate liquid cooling for high-heat-density data centers.*

This heat is then transferred to a chilled liquid loop, preventing it from entering the data hall and reducing the overall cooling load. RDHx systems enable higher rack densities, reduce reliance on traditional room cooling, and support energy efficiency and sustainability goals. Currently popular for retrofits and new builds alike, RDHx are evolving with smarter controls and higher capacity designs to meet demands of AI and edge computing workloads.

iii. Heat reuse in data centers

As described in more detail in Chapter 2.4, data centers convert nearly all the electricity they consume into heat—typically considered a waste product. But with growing energy and climate concerns, data center operators are increasingly viewing waste heat as a valuable resource. When captured and redirected, this thermal energy can support district heating systems, greenhouse agriculture, industrial processes or building-level heating, offering both carbon reductions and community benefits. One interesting approach is repurposing waste heat to drive adsorption cooling systems, which use heat rather than electricity to regenerate a desiccant material—providing an energy-efficient way to produce chilled water or air for additional cooling loads.[50,51]

In colder regions, the most impactful use of waste heat is often through district heating networks—centralized systems that distribute hot water or steam to buildings for space conditioning and water heating. In 2020, Meta began routing low-grade heat from its Odense data center to the local district heating utility (Fjernvarme Fyn).[52] Through heat pumps, the temperature is raised to meet residential needs. The system provides essentially “free” heat to over 7000 homes (and growing), offsetting fossil fuel use and supporting Denmark’s national climate targets. Following Meta’s lead, Microsoft is launching a heat recovery system at its new Danish data centers to supply thermal energy to the local district heating network. Scheduled for full operation by 2026, the system will deliver surplus heat that is equivalent to thousands of households’ annual demand.[53] The system forms part of Microsoft’s broader strategy to align with Denmark’s clean energy and circular economy goals.[54]

As compute demand grows, especially in temperate and cold regions, heat reuse could become a key pillar of sustainable data center design—transforming facilities from energy consumers into climate-positive infrastructure components.

iv. Other innovations and trends in data center cooling

Beyond air and liquid cooling fundamentals, a wave of emerging technologies and design philosophies are reshaping how data centers manage heat, water and infrastructure resilience. These innovations aim to further reduce energy use, minimize climate impacts and support the deployment of flexible, sustainable data center ecosystems. Technologies being further developed and tested include the following:

- **Thermosyphons** are sealed, gravity-driven heat transfer loops that use phase change (evaporation and condensation) of a working fluid to move heat with no mechanical input. One study suggests these devices have the potential to cut cooling energy use by up to 45% under optimal conditions compared to conventional air-based cooling systems operating under the same heat rejection requirements.[55] They offer near-zero-energy heat rejection and are increasingly integrated into free cooling systems for high-efficiency data centers. Thermosyphon-like technologies also can be found in various cooling solutions available today, but looking ahead, AI-optimized thermosyphons embedded with sensors and adaptive controls could dynamically tune their performance to match varying workloads and environmental conditions.

- **Adsorption cooling** uses materials like silica gel or zeolites that absorb moisture and release heat, then regenerate when heated. These systems can reuse low-grade waste heat to drive their cycles (without requiring significant electrical input)—ideal for pairing with high-efficiency CPUs and heat recovery loops. Recent studies have demonstrated effective integration of adsorption chillers with data center waste heat recovery loops.[50,51]

- **Desiccant systems** remove humidity from intake air before use of mechanical or evaporative cooling, increasing the effectiveness of both strategies and reducing microbial risk—particularly valuable in hot-humid climates. A 2023 simulation study focused on Southeast Asia demonstrated greater than 25% cooling energy savings when using a liquid-desiccant free cooling system compared to a conventional vapor-compression air conditioning system.[56] Desiccant systems could play a role in adaptive, climate-responsive cooling architectures, enabling data centers in hot and humid climates to reduce PUE and increase resilience while maintaining tight control over air quality and corrosion risk.

- **Thermal batteries** store excess or off-peak cooling energy (e.g., in chilled water/ice, reservoirs or solid media) for use during peak demand, improving load balancing and reducing strain on the grid. They are especially useful in regions with variable renewable energy supply. For example, Exowatt's P3 system uses concentrated solar heat stored in solid thermal storage media to provide around-the-clock power and cooling dispatch.[57] When needed, this thermal energy is converted to electricity using a thermophotovoltaic (TPV) generator, which works by first using the stored heat to warm a specialized material until it emits in the infrared to a photovoltaic system, which does the final conversion. A recent National Renewable Energy Laboratory (NREL) technoeconomic analysis of a 5 MW data center demonstrated a thermal energy storage system that reduced the levelized cost of cooling to $5/MWh compared to $15/MWh for conventional chillers and dry coolers.[58]

- **Phase change materials** (PCMs) passively absorb and release large amounts of thermal energy at specific temperatures. Embedded in server racks or walls, PCMs can buffer temperature spikes during demand surges or cooling system failures. For example, Phase Change Solutions offers BioPCM® and Apollo™ Smart Panels, which are widely used in telecom shelters and edge data center applications to stabilize indoor temperatures and reduce cooling energy demand.[59] A 2021 review highlights that integrating PCMs into data center infrastructure enhances thermal management and provides more precise control of temperature, improving overall energy efficiency and reliability.[60]

- **Refrigerants** are under increasing scrutiny due to their GWP. Next-generation systems are shifting toward low-GWP or natural refrigerants (e.g., R-1234yf, carbon dioxide (CO_2), ammonia), in line with global reduction mandates.[16] For example, Vertiv recently launched its Liebert® Evaporative Free Cooling (EFC) unit for data centers, featuring a low-GWP refrigerant (called "R-454B"), which is quickly becoming the new HVAC industry standard to enable significant efficiency gains and carbon footprint reduction.[61] Oak Ridge National Laboratory's field-scale demonstrations are revealing that replacing conventional high-GWP refrigerants with ultra-low GWP fluids can reduce a data center's annual CO_2 emissions by up to 40%.[16] In the United States, regulatory frameworks like US Environmental Protection Agency (EPA) Rule 23 requires new residential and light-commercial HVAC systems to use refrigerants with a GWP of 700 or less. In the United Kingdom, under the retained EU F-gas Regulation and domestic Net Zero Carbon Buildings Standard, refrigerants must generally have a GWP of 677 or lower to qualify as low-GWP products.[62]

- **Modular cooling architectures** support rapid deployment, localized upgrades and climate-specific customization. Scaling modular cooling reduces vendor lock-in, accelerates innovation through openness and facilitates global alignment on cooling best practices. Modular units can be factory-built and pre-engineered with liquid or hybrid systems, reducing on-site labor, construction time and permitting challenges. Modular cooling architectures emerged in the late 2000s and early 2010s with early adoption by hyperscalers and vendors offering in-row, rear-door and containerized solutions that enabled faster deployment and greater energy efficiency. However, broad adoption was initially limited by higher upfront costs and integration challenges with legacy infrastructure. A 2024 study found that these systems can be deployed in less than 12 weeks (compared to 18-24 months for traditional builds) with a reduction of on-site labor of over 60%.[63] The Open Compute Project (OCP) is promoting interoperability for modular solutions across Original Equipment Manufacturers (OEMs) and hyperscalers and is enabling faster adoption of high-density, high-efficiency cooling solutions.[64]

- **Advanced heat exchangers** are critical to enhancing liquid-cooled and hybrid systems (see case study inset).

Case Study 3: Liquid heat exchangers at scale—Microsoft pushes thermal boundaries

As the race to cool high-density AI workloads intensifies, Microsoft made major strides in 2024 deploying cutting-edge liquid heat exchanger technologies designed to reduce water use, shrink energy footprints and future-proof data centers against rising thermal loads.

Microsoft's 2024 launch of next-generation liquid-to-air heat exchanger units improves heat rejection by circulating warm coolant from its closed-loop, DTC cooling system through compact ambient-air heat exchangers. This allows servers' heat to be expelled directly to outside air without evaporative cooling, thereby eliminating water use while still efficiently removing thermal energy from the data hall. The system features advanced surface engineering to enhance thermal transfer efficiency.[65] While Microsoft does not provide details, surface engineering may include features such enhanced microchannel designs and/or specialized coatings. This innovation supports Microsoft's broader sustainability targets.

C. Barriers

While innovation in data center cooling is accelerating, widespread adoption of advanced systems faces a range of technical, economic and regulatory challenges. These barriers affect decisions across facility design, operations and long-term investment strategy and must be addressed to scale energy- and water-efficient cooling globally.

Table 2.3-2. Barriers.

Barrier	Emergine solution or mitigation path
Lack of technology maturity leading to inconsistent standards	Open standards (e.g., OCP, ASHRAE Advanced Cooling Solutions); industry-led interoperability efforts
Supply chain constraints, such as long lead times	Localized sourcing; multi-vendor compatibility; simplified system design; modularity
Increased system complexity/reliability requiring monitoring, training, maintenance	AI/ML-powered smart controls; predictive maintenance tools
Capital and operational cost uncertainty	Total cost of ownership (TCO) modeling; case studies; policy-backed green financing tools
Water scarcity	Shift to closed-loop, zero-water systems; adopt advanced dry or air-side cooling
Retrofit constraints, such as space limitations, airflow paths	Use of modular, edge-integrated cooling units; retrofit incentives
Regional regulatory considerations (e.g., building codes, state-level energy codes, water and environmental regulations)	Early engagement with permitting authorities; harmonized environmental and building codes; regional tech pilots
Market/policy levers undervalue energy/water savings	Carbon pricing; water-use transparency mandates; performance-based energy incentives or tax credits

D. Recommendations

Achieving scalable, sustainable cooling across the data center industry will require coordinated engagement from local and national governments, industry, academia and other stakeholders. The recommendations below are organized by stakeholder group and grounded in ongoing industry best practices, emerging research and field-tested case studies.

1. ***Local governments should:***
   a. ***Work with utilities to expand access to non-potable or recycled water sources.***
   b. ***Offer data centers incentive structures****—such as tax abatements or expedited permitting—****tied to clear sustainability performance benchmarks****, such as targets for PUE, WUE and heat recovery ratios.*

2. ***Local governments in regions with naturally favorable climates for cooling should:***
   a. ***Actively promote this advantage to data center operators*** *and develop targeted incentives to attract new facilities, positioning their areas as energy- and cost-efficient locations for sustainable data center development.*
   b. *Accelerate deployment of advanced cooling by* ***streamlining permitting*** *for projects that integrate sustainable thermal management strategies, such as free cooling, heat reuse or closed-loop systems.*
   c. ***Update zoning regulations*** *to enable co-location of data centers with facilities that can use waste heat, such as greenhouses or municipal buildings.*

3. ***Local governments in regions with favorable conditions for thermal integration into district energy systems should work directly with data center operators*** *who understand regional opportunities on such projects.*[52]

4. ***National policymakers should:***
   a. ***Establish the market conditions and regulatory frameworks*** *necessary for the broad adoption of energy- and water-efficient cooling technologies.*
   b. ***Create and enforce minimum energy performance standards for data centers,*** *along with voluntary or mandatory reporting requirements for PUE and WUE; expand programs, such as the US EPA's ENERGY STAR for Data Centers*[66] *or the EU Code of Conduct for Data Centres,*[67] *which can help set performance baselines and identify leaders.*
   c. ***Use government funding to support research and pilot programs*** *for promising but commercially immature technologies, such as two-phase immersion cooling, modular systems, thermal batteries and zero-water liquid cooling.*
   d. ***Provide tax credits, green bonds and procurement incentives*** *to help de-risk early adoption and support widespread deployment of sustainable cooling systems.*

5. ***Universities and research institutions should:***
   a. ***Prioritize studies of novel cooling techniques,*** *including many of the innovations described above, such as desiccant-based systems, thermal batteries, AI-optimized thermal control platforms and climate-specific hybrid systems.*
   b. ***Host experimental testbeds*** *or collaborate with industry to evaluate the performance of emerging solutions in field conditions.*
   c. ***Create open access datasets, simulation tools and digital twins*** *to allow broader communities to model, compare and benchmark advanced cooling approaches; standardizing such tools will improve planning accuracy and reduce design risk.*[68]
   d. ***Create or expand curricula on thermal systems, green data infrastructure and resilient design*** *to train the next generation of engineers and planners.*

6. ***Standards organizations*** *such as ASHRAE, International Standards Organization (ISO) and OCP should:*
    a. ***Facilitate innovation and interoperability*** *to evolve their guidance.*
    b. ***Establish uniform testing protocols and certification pathways*** *to validate performance of new technologies—especially liquid cooling, rear-door heat exchangers and high-efficiency refrigerants.*
    c. ***Push for global alignment on definitions and performance thresholds*** *to lower costs, reduce vendor lock-in and allow data center operators to deploy advanced cooling with greater confidence across international markets.*

7. ***Cooling equipment manufacturers,*** *bridging the gap between research and widespread implementation, should:*
    a. ***Invest in research and development (R&D)*** *focused on compact cold plates, advanced heat exchanger surfaces and system-integrated controls with predictive maintenance and AI optimization capabilities.*
    b. ***Offer comprehensive, modular solutions*** *that include sensors, telemetry and leak detection to reduce operational complexity.*
    c. ***Prioritize low-GWP and natural refrigerant alternatives,*** *consistent with the climate goals outlined in the Montreal Protocol Kigali Amendment, particularly ss regulations around refrigerants evolve.*[16]

8. ***Data center developers and operators should:***
    a. ***Integrate design early in the project development process,*** *especially in siting decisions.*
    b. ***Select locations*** *that enable the use of free cooling, heat reuse or access to non-potable water and renewable energy.*
    c. ***Install climate-appropriate cooling systems****—such as evaporative cooling in dry regions or air-side economization in temperate zones—in tandem with IT deployment strategies.*
    d. ***Establish facility-level energy and water performance targets and publish sustainability metrics annually.***
    e. ***Evaluate and design for heat reuse opportunities,*** *either through district heating connections or local use cases like agricultural greenhouses, building heating or industrial preheating.*
    f. ***Set aside dedicated infrastructure for pilot deployments*** *of emerging cooling systems, allowing testing without disrupting core operations.*

9. ***Utilities should:***
    a. ***Partner with data centers to support load shifting*** *(to align significant workload periods with cooler times of day).*
    b. ***Integrate waste heat into community heating systems.***

   c. ***Offer incentive structures for grid-responsive cooling*** *in which data center cooling systems adjust their operation in response to signals from the electric grid.*

10. ***Environmental organizations should advocate for:***
   a. ***Low-carbon and water use***
   b. ***Transparent reporting***
   c. ***Waste heat reuse***
   d. ***Responsible siting*** *to align with climate and sustainability goals*

11. Investors and financiers should require disclosures *on WUE, PUE and refrigerant use in project finance deals because these metrics directly impact a data center's operational efficiency, climate risk exposure, regulatory compliance and long-term sustainability performance—all of which influence financial returns, reputational risk and alignment with Environmental, Social, and Governance (ESG) commitments.*

12. Insurance providers should reduce risk premiums for data center operators that adopt redundant and fault-tolerant cooling systems*—especially those with active leak detection and real-time monitoring—because these technologies significantly lower the likelihood of costly outages, equipment damage, and water or refrigerant leaks, thereby reducing the insurer's exposure to operational and environmental claims. For most companies, this is a shift from current practice, where premiums often do not fully account for the added risk mitigation these systems provide.*

13. End-use customers*, such as large cloud clients,* ***should shape demand by requiring data centers to meet high-efficiency and low-emissions cooling benchmarks*** *in service agreements, such as maintaining a low PUE (ideally below 1.3) and minimizing greenhouse gas emissions by using low-GWP refrigerants, carbon-free electricity for cooling, and water-efficient or closed-loop systems.*

Realizing the full potential of sustainable data center cooling will require a collective shift—from isolated innovation to coordinated implementation. Each stakeholder group has a distinct and essential role to play, from setting performance standards and funding next-generation technologies to deploying climate-responsive designs and sharing operational data. By aligning incentives, accelerating open collaboration and embedding sustainability into the planning and operation of digital infrastructure, we can ensure that data centers not only meet the demands of the digital age but also contribute to a more resource-efficient and climate-resilient future.

E. References

1. Refrigeration American Society for Heating, and Air Conditioning Engineers (ASHRAE). Data Center Resource Page; ASHRAE, Peachtree Corners, Georgia, https://www.ashrae.org/technical-resources/bookstore/datacom-series (Accessed August 2025).

2. Vertiv. Understanding direct-to-chip cooling in HPC infrastructure: A deep dive into liquid cooling; Vertiv, Thane, Maharashtra, India, https://www.vertiv.com/en-in/about/news-and-insights/articles/educational-articles/understanding-direct-to-chip-cooling-in-hpc-infrastructure-a-deep-dive-into-liquid-cooling/ (Accessed August 2025).

3. Arman Shehabi, Alex Newkirk, Sarah J Smith, Alex Hubbard, Nuoa Lei, Md Abu Bakar Siddik, Billie Holecek, Jonathan Koomey, Eric Masanet & Dale Sartor. 2024 United States Data Center Energy Usage Report. Report No. LBNL-2001637, (2024), https://doi.org/10.71468/P1WC7Q.

4. Vertiv. Figure 1 in "Understanding direct-to-chip cooling in HPC infrastructure: A deep dive into liquid cooling"; Vertiv, Thane, Maharashtra, India, https://www.vertiv.com/49ffc2/globalassets/images/on-page-image/800x450/800x450-liquid-cooling-chart_383570_0.jpg (Accessed August 2025).

5. Nuoa Lei, Jun Lu, Arman Shehabi & Eric Masanet. The water use of data center workloads: A review and assessment of key determinants. Resources, Conservation and Recycling 219, 108310 (2025). https://doi.org/10.1016/j.resconrec.2025.108310.

6. Eric Masanet, Arman Shehabi, Nuoa Lei, Sarah Smith & Jonathan Koomey. Recalibrating global data center energy-use estimates. Science 367, 984-986 (2020). https://datacenters.lbl.gov/sites/default/files/Masanet_et_al_Science_2020.full_.pdf.

7. Ali Badiei, Eric Jadowski, Saba Sadati, Arash Beizaee, Jing Li, Leila Khajenoori, Hamid Reza Nasriani, Guiqiang Li & Xin Xiao. The Energy-Saving Potential of Air-Side Economisers in Modular Data Centres: Analysis of Opportunities and Risks in Different Climates. Sustainability 15, 10777 (2023). https://doi.org/10.3390/su151410777.

8. Tamzid Ahmed. World's 5 Bigest Underwater Data Center Projects; Brightlio, Culver City, California, https://brightlio.com/underwater-data-centers/ (2025).

9. Microsoft. Project Natick (Homepage); Redmond, Washington, https://natick.research.microsoft.com/ (Accessed August 2025).

10. Subsea Cloud. Subsea Cloud Homepage; Houston, Texas, https://www.subseacloud.com (Accessed August 2025).

11. You Xiaoying. China Is Putting Data Centers in the Ocean to Keep Them Cool; Scientific American, New York, New York, https://www.scientificamerican.com/article/china-powers-ai-boom-with-undersea-data-centers/ (2025).

12. Hewlett Packard Enterprise. HPE Spaceborne Computer-2 returns to the International Space Station (Press Release); Hewlett Packard Enterprise Spring, Texas, https://www.hpe.com/us/en/newsroom/press-release/2024/01/hpe-spaceborne-computer-2-returns-to-the-international-space-station.html (2024).

13. Douglas Donnellan, Daniel Bizo, Jacqueline Davis, Andy Lawrence, Dr. Owen Rogers, Lenny Simon & Max Smolaks. Uptime Institute Global Data Center Survey 2023: Executive Summary; Uptime Intelligence, New York, New York, https://uptimeinstitute.com/resources/asset/executive-summary-uptime-institute-global-data-center-survey-2023 (2023).

14. Miguel Yañez-Barnuevo. Data Centers and Water Consumption; Environmental and Energy Study Institute (EESI), Washingon, DC, https://www.eesi.org/articles/view/data-centers-and-water-consumption (2025).

15. Andrew Higgins. What Is Water Usage Effectiveness (WUE) in Data Centers?; Equinix, Readwood City, California, https://blog.equinix.com/blog/2024/11/13/what-is-water-usage-effectiveness-wue-in-data-centers/ (2024).

16. United Nations Environment Programme (UNEP) and International Energy Agency (IEA). Cooling Emissions and Policy Synthesis Report; UNEP, Nairobi and IEA, Paris, https://www.unep.org/resources/report/cooling-emissions-and-policy-synthesis-report (2020).

17. European Commission. F-gas legislation; Bruxelles, Belgium, https://climate.ec.europa.eu/eu-action/fluorinated-greenhouse-gases/f-gas-legislation_en (Accessed August 2025).

18. Husam Alissa, Teresa Nick, Ashish Raniwala, Alberto Arribas Herranz, Kali Frost, Ioannis Manousakis, Kari Lio, Brijesh Warrier, Vaidehi Oruganti, T. J. DiCaprio, Kathryn Oseen-Senda, Bharath Ramakrishnan, Naval Gupta et al. Using life cycle assessment to drive innovation for sustainable cool clouds. Nature 641, 331-338 (2025). https://doi.org/10.1038/s41586-025-08832-3.

19. Uptime Institute. Uptime Institute's 2022 Global Data Center Survey Reveals Strong Industry Growth as Operators Brace for Expanding Sustainability Requirements (Press Release); New York, New York, https://uptimeinstitute.com/about-ui/press-releases/2022-global-data-center-survey-reveals-strong-industry-growth (2022).

20. Google. Google data center PUE performance: 2024 PUE Yearly Report; Mountain View, California, https://datacenters.google/efficiency/#2024 (2024).

21. Qian Wei, Jun Lu, Xiaoping Xia, Bin Zhang, Xiang Ying & Leihong Li. Performance and Applicability Analysis of Indirect Evaporative Cooling Units in Data Centers Across Various Humidity Regions. Buildings 14, 3623 (2024). https://doi.org/10.3390/buildings14113623.

22. Fred Rebarber. Quantifying the Impact on PUE and Energy Consumption When Introducing Liquid Cooling Into an Air-cooled Data Center (Blog Post); Vertiv Group Corp., Westerville, Ohio, https://www.vertiv.com/en-us/about/news-and-insights/articles/blog-posts/quantifying-data-center-pue-when-introducing-liquid-cooling/ (2023).

23. Rui Kong, Hainan Zhang, Mingsheng Tang, Huiming Zou, Changqing Tian & Tao Ding. Enhancing data center cooling efficiency and ability: A comprehensive review of direct liquid cooling technologies. Energy 308, 132846 (2024). https://doi.org/10.1016/j.energy.2024.132846.

24. Nuoa Lei & Eric Masanet. Climate- and technology-specific PUE and WUE estimations for U.S. data centers using a hybrid statistical and thermodynamics-based approach. Resources, Conservation and Recycling 182, 106323 (2022). https://doi.org/10.1016/j.resconrec.2022.106323.

25. Arman Shehabi, Sarah Smith, Dale Sartor, Richard Brown, Magnus Herrlin, Jonathan Koomey, Eric Masanet, Inês Azevedo Nathaniel Horner & William Lintner. United States Data Center Energy Usage Report (LBNL-1005775); Ernest Orlando Lawrence Berkeley National Laboratory, Berkeley, California, https://eta-publications.lbl.gov/sites/default/files/lbnl-1005775_v2.pdf.

26. Anthony Sbarra & Laurie Williams. Uptime Institute 2024 Cooling Systems Survey: Direct Liquid Cooling Results; Uptime Institute, New York, New York, https://datacenter.uptimeinstitute.com/rs/711-RIA-145/images/2024.Cooling.Survey.Report.pdf (2024).

27. Kaifang Shi, Yun Chen, Bailang Yu, Tingbao Xu, Zuoqi Chen, Rui Liu, Linyi Li & Jianping Wu. Modeling spatiotemporal CO_2 (carbon dioxide) emission dynamics in China from DMSP-OLS nighttime stable light data using panel data analysis. Applied Energy 168, 523-533 (2016). https://doi.org/10.1016/j.apenergy.2015.11.055

28. Rich Miller. Microsoft's Chiller-less Data Center; Data Center Knowledge, San Francisco, California, https://www.datacenterknowledge.com/hyperscalers/microsoft-s-chiller-less-data-center (2009).

29. Flann Gao & Luica Mak. Alibaba Cloud launches new energy-efficient Qiandao Lake Data Center; Alibaba Cloud, Hangzhou, China, https://www.alibabacloud.com/en/press-room/alicloud-launches-new-energy-efficient-qiandao-lake-data-center?_p_lc=1 (2015).

30. Nimbus. Adiabatic Cooling Applications: Data Centers; Massillon, Ohio, https://nimbus.cool/resources/articles/adiabatic-cooling-applications-data-centers/ (Accessed August 2025).

31. Open Compute Project. Water Efficiency at Facebook's Prineville Data Center (Blog); Austin, Texas, https://www.opencompute.org/blog/water-efficiency-at-facebooks-prineville-data-center (2013).

32. Chi Zhou, Doris Gao, Lisa Rivalin, Andrew Grier, Gerson Arteaga Ramirez & John Fabian. in Deployable RL: From Research to Practice@ Reinforcement Learning Conference 2024; (Amherst, Massachusetts, 2024; https://openreview.net/pdf?id=3hZL9Vv0Ay).

33. Jerry Luo, Cosmin Paduraru, Octavian Voicu, Yuri Chervonyi, Scott Munns, Jerry Li, Crystal Qian, Praneet Dutta, Jared Quincy Davis, Ningjia Wu, Xingwei Yang, Chu-Ming Chang, Ted Li et al. Controlling Commercial Cooling Systems Using Reinforcement Learning. arXiv:2211.07357 (2022). https://doi.org/10.48550/arXiv.2211.07357.

34. Richard Evans & Jim Gao. DeepMind AI Reduces Google Data Centre Cooling Bill by 40%; Google DeepMind, London, United Kingdom, https://deepmind.google/discover/blog/deepmind-ai-reduces-google-data-centre-cooling-bill-by-40/ (2016).

35. Wayne Williams. Megawatt-class AI server racks may well become the norm before 2030 as Nvidia displays 600kW Kyber rack design; TechRadar, London, United Kingdom, https://www.techradar.com/pro/megawatt-class-ai-server-racks-may-well-become-the-norm-before-2030-as-nvidia-displays-600kw-kyber-rack-design (2025).

36. Christopher Helman & Rashi Shrivastava. Meet The Tiny Startup Building Stargate, OpenAI's $500 Billion Data Center Moonshot; Forbes, Jersey City, New Jersey, https://www.forbes.com/sites/christopherhelman/2025/04/10/meet-the-tiny-startup-building-stargate-openais-500-billion-data-center-moonshot/ (2025).

37. Ali Heydari, Ahmad R. Gharaibeh, Mohammad Tradat, Qusai Soud, Yaman

Manaserh, Vahideh Radmard, Bahareh Eslami, Jeremy Rodriguez & Bahgat Sammakia. Experimental evaluation of direct-to-chip cold plate liquid cooling for high-heat-density data centers. Applied Thermal Engineering 239, 122122 (2024). https://doi.org/10.1016/j.applthermaleng.2023.122122.

38. Nicola Pieretti (Vertiv). Chill factor: Top liquid cooling considerations for high-density environments; Data Center Dynamics, London, United Kingdom, https://www.datacenterdynamics.com/en/opinions/chill-factor-top-liquid-cooling-considerations-for-high-density-environments/ (2025).

39. Vertiv. Liquid cooling options for data centers; Westerville, Ohio, https://www.vertiv.com/en-us/solutions/learn-about/liquid-cooling-options-for-data-centers/ (Accessed August 2025).

40. Mark E Steinke (Committee Chair). Mission Critical Facilities, Data Centers, Technology Spaces and Electronic Equipment: ASHRAE Technical Committee 9.9; American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE), Peachtree Corners, Georgia, https://tpc.ashrae.org/?cmtKey=fd4a4ee6-96a3-4f61-8b85-43418dfa988d (Accessed August 2025).

41. Zutacore. https://zutacore.com/ (Homepage); San Jose, California, https://zutacore.com/ (Accessed August 2025).

42. Steve Solomon. Sustainable by design: Next-generation datacenters consume zero water for cooling; Microsoft, Redmond, Washington, https://www.microsoft.com/en-us/microsoft-cloud/blog/2024/12/09/sustainable-by-design-next-generation-datacenters-consume-zero-water-for-cooling/ (2024).

43. Alex Davies. Next-gen AI demands smarter cooling tech. Here's how AWS delivered it in just 11 months; Amazon, Seattle, Washington, https://www.aboutamazon.com/news/aws/aws-liquid-cooling-data-centers (2025).

44. Mike Wheatley. AWS integrates liquid cooling and simplifies electrical distribution to lower data center power consumption; Silicon Angle Media, Inc., Palo Alto, California, https://siliconangle.com/2024/12/02/aws-integrates-liquid-cooling-simplifies-electrical-distribution-lower-data-center-power-consumption/ (2024).

45. Georgia Butler. AWS global data centers achieved PUE of 1.15 in 2023; Data Center Dynamics, London, United Kingdom, https://www.datacenterdynamics.com/en/news/aws-global-data-centers-achieved-pue-of-115-in-2023/ (2024).

46. Drew Robb. Data Center Cooling Trends for 2025; Upsite Technologies, Albuquerque, New Mexico, https://www.upsite.com/blog/data-center-cooling-trends-for-2025/ (2025).

47. Green Revolution Cooling (GRC). Green Revolution Cooling: Redefining the Efficiency and Sustainability of Data Center Cooling (Homepage); GRC, Austin, Texas, https://www.grcooling.com (Accessed August 2025).

48. Submer.Datacenters that make sense (Homepage); Barcelona, Spain, https://submer.com/ (Accessed August 2025).

49. Pranati Sahoo. Immersion Cooling for Data Center Cooling-Innovation Speaks it All!; Energetica India, Pune, Maharashtra, India, https://www.energetica-india.net/articles/immersion-cooling-for-data-center-cooling—innovation-speaks-it-all (2022).

50. Sumit Kumar Singh, Dibakar Rakshit, K. Ravi Kumar & Anurag Agarwal. Recent advancements and sustainable solutions in adsorption-based cooling systems integrated with renewable energy sources and industrial waste heat: A review. Cleaner Engineering and Technology 23, 100827 (2024). https://doi.org/10.1016/j.clet.2024.100827.

51. Zhaopeng Cui, Shuai Du, Tianhao Zhao, Zhihui Chen & Ruzhu Wang. High-power-density adsorption chiller driven by data center waste heat using encapsulated composite as adsorbent. Energy 311, 133391 (2024). https://doi.org/10.1016/j.energy.2024.133391.

52. Maria Lind Arlaud (Editor in Chief). District energy: The backbone of a flexible, resilient and efficient energy system; State of Green, Copenhagen, Denmark, https://stateofgreen.com/en/publications/district-energy/ (2024).

53. Microsoft. Surplus datacenter heat will be repurposed to heat homes in Denmark (Blog); Redmond, Washington, https://local.microsoft.com/blog/datacenter_heat_repurposed/ (Accessed August 2025).

54. Terje Solsvik. Denmark's European Energy signs long-term power deals with Microsoft; Reuters, London, United Kingdom, https://www.reuters.com/business/energy/denmarks-european-energy-signs-long-term-power-deals-with-microsoft-2024-06-12/ (2024).

55. Tao Ding, Xiaoxuan Chen, Hanwen Cao, Zhiguang He, Jianmin Wang & Zhen Li. Principles of loop thermosyphon and its application in data center cooling systems: A review. Renewable and Sustainable Energy Reviews 150, 111389 (2021). https://doi.org/10.1016/j.rser.2021.111389.

56. Hyo-Lim Park, Soo-Jin Lee, Jae-Hee Lee & Jae-Weon Jeong. A liquid desiccant-assisted free-cooling system for energy-efficient data centers in hot and humid climates. Case Studies in Thermal Engineering 52, 103683 (2023). https://doi.org/10.1016/j.csite.2023.103683.

57. Exowatt. Powering AI with 24-hour Solar (Homepage); Miami, Florida, https://www.exowatt.com/ (Accessed August 2025).

58. National Renewable Energy Laboratory (NREL). High-Performance Computing Data Center; NREL, Applewood, Colorado, https://www.nrel.gov/computational-science/hpc-data-center (Accessed August 2025).

59. Phase Change Solutions (PCS). Data Centers & Telecom; Asheboro, North Carolina, https://phasechange.com/data-center-telecom/ (Accessed August 2025).

60. Xiaolei Yuan, Xuetao Zhou, Yiqun Pan, Risto Kosonen, Hao Cai, Yang Gao & Yu Wang. Phase change cooling in data centers: A review. Energy and Buildings 236, 110764 (2021). https://doi.org/10.1016/j.enbuild.2021.110764.

61. Vertiv. Vertiv Homepage; Thane, Maharashtra, India, https://www.vertiv.com/en-us/ (Accessed August 2025).

62. Vertiv. Evolving Refrigerant Regulations and Applications in Data Center Environments (Vetiv White Paper); Vertiv, Westerville, Ohio, https://www.vertiv.com/48f7e3/contentassets/8eee3437fd734143a36d0395e65e8de7/vertiv-low-gwp-refrigerants-wp-en-na-sl-71228-web.pdf (2025).

63. Jinghan Sun, Zibo Gong, Anup Agarwal, Shadi Noghabi, Ranveer Chandra, Marc Snir & Jian Huang. “Exploring the Efficiency of Renewable Energy-based Modular Data Centers at Scale” in Proceedings of the 2024 ACM Symposium on Cloud Computing, Redmond, Washington. 552-569, https://doi.org/10.1145/3698038.3698544,(2024).

64. Open Compute Project. Scaling Innovation Through Collaboration! (Open Compute Project Homepage); Austin, Texas, https://www.opencompute.org/ (Accessed August 2025).

65. AndyD750 (Azure Infrastructure Blog). Liquid Cooling in Air Cooled Data Centers on Microsoft Azure; Microsoft Community Hub, Redmond, Washington, https://techcommunity.microsoft.com/blog/azureinfrastructureblog/liquid-cooling-in-air-cooled-data-centers-on-microsoft-azure/4268822 (2024).

66. ENERGY STAR. Data Centers (under “Energy Efficient Products”); US Environmental Protection Agency (EPA), Washington, DC, https://www.energystar.gov/products/data_centers (Accessed August 2025).

67. The Joint Research Centre: EU Science Hub. European Code of Conduct for Energy Efficiency in Data Centres; European Commission, Bruxelles, Belgium, https://joint-research-centre.ec.europa.eu/jrc-news-and-updates/eu-code-conduct-data-centres-towards-more-innovative-sustainable-and-secure-data-centre-facilities-2023-09-05_en (Accessed August 2025).

68. Jyotika Athavale, Cullen Bash, Wesley Brewer, Matthias Maiterth, Dejan Milojicic, Harry Petty & Soumyendu Sarkar. Digital Twins for Data Centers. 57, 151-158 (2024). https://doi.ieeecomputersociety.org/10.1109/MC.2024.3436945.).

2.4 Heat Reuse

Roger Aines and David Sandalow

Almost all the electricity used at a data center turns into heat. Very few data centers use this heat, but there is considerable interest in doing so, both to offset energy costs and to reduce environmental impacts. The rise of liquid-cooling systems (see Chapter 2.3 of this Roadmap) and higher rack densities are creating new opportunities for beneficial use of heat from data centers. This chapter explores these topics—discussing challenges, analyzing opportunities, describing current and planned projects, and concluding with recommendations.

A. Challenges

Today's data centers are dominantly air-cooled. The low exit temperature (close to ambient) and low heat capacity of the air in these data centers makes any reuse of heat difficult and often impossible. (The National Renewable Energy Lab (NREL) in Colorado has a functioning example of heat reuse at an air-cooled data center,[1] but this model is not replicable on a commercial scale.)

Heat reuse projects at data centers face additional challenges. Most fundamentally, heat is difficult to transport over long distances, so data centers must be close to facilities that need heat ("heat hosts") for projects to be successful. In addition, coordination between data centers and potential heat hosts can be complicated, requiring alignment on a range of topics, including technologies, schedules and business goals.

The Open Compute Project lists six issues limiting deployment of data center heat reuse projects:

1. Data centers and heat hosts are not always located in close proximity. The cost of new infrastructure to connect them can be significant.

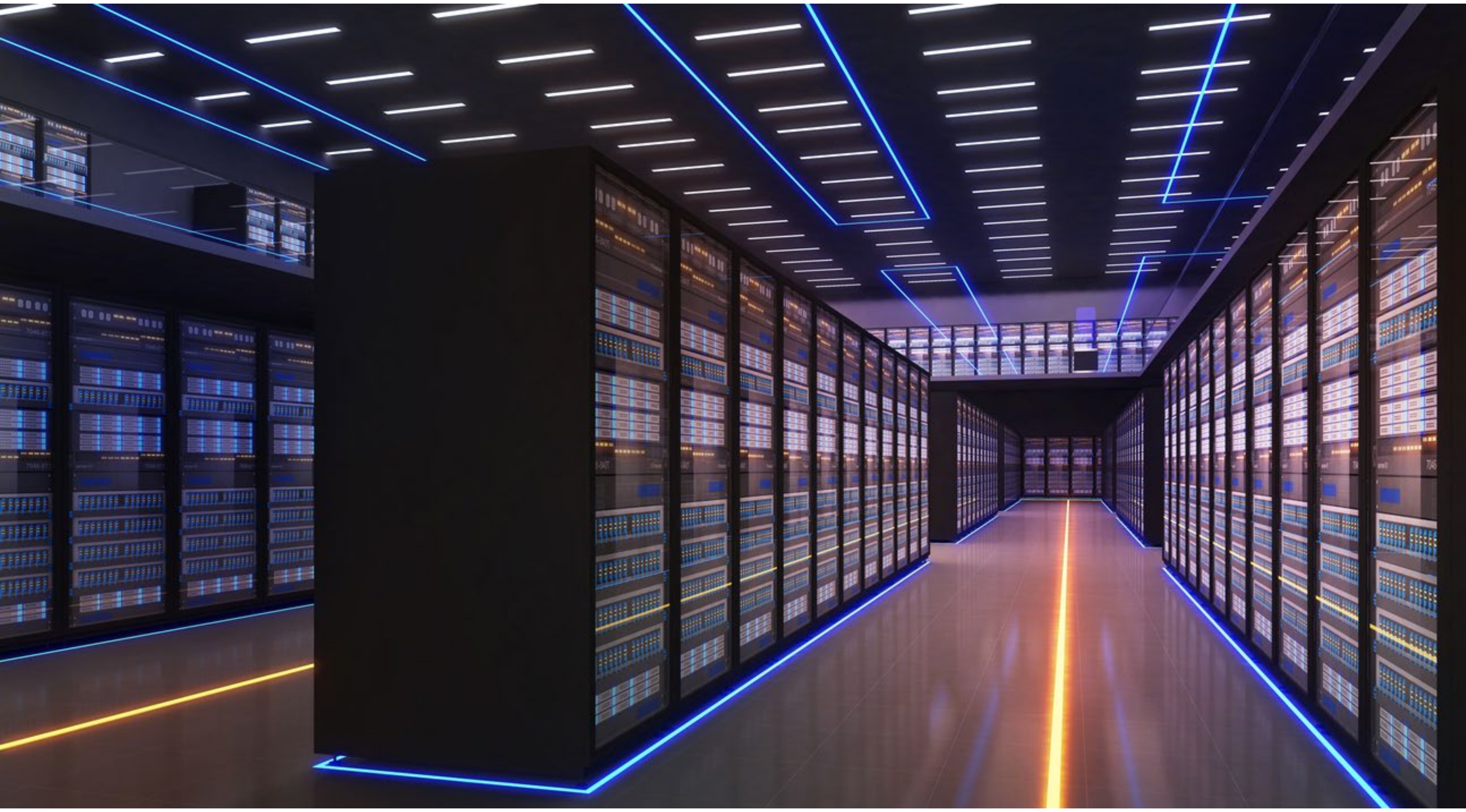
2. Many locations have climates in which heat is not required for human comfort for much of the year.

3. Local and national legislation does not always facilitate these projects.

4. Governments rarely offer subsidies for heat reuse projects.

5. Collaboration between data centers and heat hosts can be difficult.

6. Heat reuse projects require significant investment, often including underground pipes, pumping stations, heat pumps, controls and similar equipment.[2]



B. Opportunities

Liquid Cooling systems can enable reuse of heat that would otherwise be wasted. These systems have exit temperatures as high as 70 °C, limited only by the safe operating range of the chips. In liquid cooling systems, heat can be recycled at 45-70 °C, aligning with the thermal cycles in many district heating systems.

Liquid Cooling systems are rapidly growing in number, driven in part by the rise of graphics processing unit (GPU)-based computing for artificial intelligence (AI). The higher rack densities typical of GPUs result in greater waste heat and, in turn, the need for liquid cooling. This, in turn, creates new opportunities for heat reuse since the systems operate with higher exit temperatures.

Liquid Cooling systems with ultra-high rack densities (>100 kW/cabinet) also lower power usage effectiveness (PUE)[1] significantly (while capturing more than 99% of waste heat at ~55-70 °C).[3] With two-phase immersion cooling technology, data center PUE can reach 1.03-1.05.[4] Although capital costs for such systems are high, return on investment (ROI) can be relatively quick. One study reports ROI ranges of 1-3 years for single-phase immersion cooling systems.[3]

Table 2.4-1. *Types of cooling with estimated market share and heat reuse capability.*

Exit Temps	Air Cooling* 15–30 °C	Rear Door Cooling* 30–40 °C	Liquid Coolng – Direct to Chip or Immersion* 40–70 °C
Current Usage (2023)	95% Market Share[5,6]	2-3%[7]	2-5%[8]
Future Builds Usage (estimate for 2028) [9-11]	60–70% (remaining after rear door and liquid cooling estimates)	10-15%	20-25%
Comments	Dominant in legacy and smaller systems	Easiest to retrofit—maintains existing cooling methods	Can achieve PUE of 1.02; direct to chip dominant in new builds today
Heat Reuse Capability	None	Minimal—local heating	Significant, particularly at higher exit temperatures

*See Chapter 2.3 of this Roadmap for background on air cooling, rear door cooling and Liquid Cooling.

Examples of data centers with liquid cooling that reuse waste heat include the following.

- Intel + Submer (Barcelona): Deploying SmartPod XL immersion units operating at 55-70 °C, with reported >99% heat capture, feeding local district heating.[12]

- Aquasar (ETH Zurich): Hot-water cooling recovers ~80% of heat, utilized in campus heating, saving ~$1.25 million annually.[13]

- iDataCool (Germany): Custom hot-water cooling for HPC, drives an adsorption chiller; target ~65 °C for useful heat.[14]

- QTS, Digital Realty, Equinix (Nordics): Air-cooled centers use heat to warm homes; immersion cooling pilots in Finland and Sweden are underway.[15]

Combining liquid cooling systems with heat pumps can open opportunities for heat reuse at data centers. Heat pumps take a source of low-temperature heat and raise the temperature, making the heat suitable for a wider range of applications.[16]

Two types of facilities offer the best opportunities for data center heat reuse: district heating and direct air capture. Other possibilities include electricity generation, agriculture and aquaculture. These are discussed below.

i. District heating

In colder regions, the most impactful use of waste heat is often through district heating networks—centralized systems that distribute hot water or steam to buildings for space conditioning and water heating. District heating systems in Nordic countries use moderate-temperature water—typically input at 60-70 °C and returned in the vicinity of 50 °C. These are excellent matches to modern liquid-cooling systems, but in order to optimize district heating, data centers need to be carefully planned to integrate with district heating options.

District heating systems have many benefits.[17] They reduce greenhouse gas emissions and energy use, increase energy efficiency, lower building costs (no separate boilers, chillers or other related hardware) and improve reliability (industrial-grade district energy equipment is more robust than commercial equipment installed at the building level). Cities and communities often support district heating systems due to reduced cost for new housing development and the capacity to provide baseload power and heat for microgrids. Local grid infrastructure can benefit through reduced peak demand enabled by aggregating loads and shifting peak demand with thermal energy storage.[18]

The main challenge in using data center heat for district heating is seasonality. In warmer months when district heating is not needed, the data center must have an alternative method of rejecting its waste heat. This requirement raises the overall capital cost of the data center's cooling infrastructure. As a result, the benefits of linking to a district heating system must be substantial in terms of cost reductions, carbon emission reductions, improved social acceptance or otherwise.

However, several data centers have successfully supplied waste heat use for district heating systems. For example, in 2020, Meta began routing low-grade heat from its Odense data center to the local district heating utility (Fjernvarme Fyn).[19] Through heat pumps, the temperature is raised to meet residential needs. The Odense system provides essentially "free" heat to over 7000 homes (and growing), offsetting fossil fuel use and supporting Denmark's national climate targets.

Following Meta's lead, Microsoft is launching a heat recovery system at its new Danish data centers to supply thermal energy to the local district heating network. Scheduled for full operation by 2026, the system will deliver surplus heat equivalent to thousands of households' annual demand.[20] It forms part of Microsoft's broader strategy to align with Denmark's clean energy and circular economy goals.[21]

Matching data centers with district heating systems in the United States is more challenging than in northern Europe. US district heating systems—including the largest in New York and Boston/Cambridge—tend to be based on steam rather than hot water.[22] These systems cannot easily be converted to hot water because the radiators in buildings are much less efficient than those in northern Europe. To create the necessary steam, heat pumps would need to be used to raise the temperature of liquid-cooling systems. These heat pumps are becoming available but need electricity to operate, undercutting the energy reduction goal of heat reuse programs.

Chinese authorities and data center operators are exploring use of data center waste heat for district heating, though projects are at an early stage. A study by Tsinghua University and the United Nations Environment Programme (UNEP)-Copenhagen estimates that northern Chinese data centers generate about 70 petajoules of recoverable winter waste heat annually, with potential to quadruple by 2060 if coupled with heat pumps and storage.[23] In Tianjin, Tencent captures waste heat from servers at its data center and, with heat pumps, provides hot water for municipal heating.[24]

ii. Direct air capture

Since carbon emissions are one of the key issues associated with data center expansion, it is interesting to examine whether direct air capture (DAC) systems could be operated with waste heat from liquid-cooling systems. DAC facilities have the ability to operate year-round—a major advantage for data center heat reuse when compared to the seasonal nature of district heating systems. Further, while costs for DAC systems are still quite high, there are many commercial contracts to purchase carbon dioxide (CO_2) removals from the initial suppliers, suggesting that the high cost is not necessarily a barrier to deployment.[25]

The primary energy demand in a DAC facility is for heat to regenerate sorbents and solvents (the material that captures CO_2 from the air). One challenge in using data center waste heat for DAC is that the sorbents and solvents typically require temperatures in the vicinity of 90-120 °C to desorb the CO_2 in a pure state. At exit temperatures from data center liquid-cooling systems, producing the steam for desorption requires either a heat pump or very substantial vacuum. Both need significant electricity input, as does converting the captured CO_2 to liquid form for transportation.

Under these conditions the value of the captured CO_2 must be very high to justify the additional energy. The data center is acting to subsidize the DAC system, rather than the DAC system subsidizing the data center. This may be a perfectly acceptable outcome given the apparent high value of CO_2 removed from the air today ($500-$1000/ton).[25]

Microsoft—the world's largest purchaser of CO_2 from DAC—is conducting a test of DAC using data center heat.[26,27] Meta and Alphabet are exploring options in this area as well.[28,29]

iii. Organic Rankin Cycle Power (ORC) electricity generation

Using waste heat to generate electricity would have significant benefits. Such a system would be usable 24/7 and reduce the overall electricity demand of the data center. Organic Rankin Cycle Power (ORC) systems use a heat exchanger to transfer the cooling heat from a liquid system to an organic fluid that can be expanded to vapor to pass through a turbine. The fluid is then condensed again, typically by a cooling tower or air cooler. These systems are in routine use for low-temperature geothermal power and are being developed for engine heat recovery. In both cases the heat sources are higher temperature (typically 80-100 °C) than exit heat from data center liquid-cooling systems.[30]

This approach has received significant academic attention but is not currently in use for data centers. One analysis suggests a payback period of 4-7 years.[31] However, a manufacturer of these systems for engine heat reuse, Vertiv, has sponsored a detailed experimental study of its application to data center heat at 58 °C with an atmospheric heat sink between 14 and 35 °C.[32]

The company found that:

> "When operating at TH [hot-side temperature] ~58 °C, TL (low-side temperature) ~14 °C, and near full load, the ORC can convert ~2% of the waste heat into mechanical energy. Although this may appear negligible, the best data centers consume ~20% of the IT load to transport the waste heat to the outdoors. The ORC provides this cooling with a net output of mechanical energy creating a significant improvement in data center PUE. The regenerative turbine liquid pump consumes ~50% of the energy output of the expander. When the parasitic load to transport the residual waste heat into ambient air is considered, the ORC as a WHR [waste-heat recovery] system is an energy consumer.
>
> That being said, in a data center application the WHR system is the cooling system which must operate continuously as load varies from 0% to 100% and ambient temperature varies according to the prevailing weather conditions predicted based on a century of records. *The unstable flow limits at the expander inlet in the subject ORC system fall well within the spring, summer, and fall high temperatures expected in Ashburn, VA. Thus, use of ORC WHR as the sole data center cooling means is precluded."*

In other words, the system tested was unable to operate stably in temperatures expected of temperate zone data centers. Vertiv makes recommendations for how these fundamental problems may be overcome, but the challenge is significant when considering the relatively low amount of power generated and the anticipated high capital cost. This detailed analysis from a commercial supplier suggests why current implementations are not commercial and why the complexity of this approach is probably hindering even a demonstration system in a data center today.

iv. Other possible uses of waste heat

Other possible heat hosts for data centers include aquaculture projects, greenhouses and low temperature industrial processes (such as drying). A large number of possible applications of this kind exist at temperatures of 60-90 °C,[33] many of which do not have the seasonality constraints of district heating. However, current projects are very limited in number and are mostly in Nordic countries. The Green Mountain data center in Norway is sending waste to the world's largest land-based trout farm.[34] Luleå University of Technology in Sweden and the University of Notre Dame in the US have greenhouses that receive heat from data centers.[35] The EcoDataCenter in Falun, Sweden uses waste heat to dry wood for wood pellet production.[36] However, most applications of this kind remain academic exercises only and are probably too small to attract the attention of new large data centers (>100 MW), which will have liquid cooling systems large enough to integrate into more complex industrial applications.

C. Current and Planned Projects

The Open Compute Project (OCP) maintains a list of heat reuse projects at data centers around the world.[2,37,38] The numbers are modest. OCP lists only 12 projects with a capacity of 5 MW or more of heat reuse in operation today and 13 projects of that size in planning. Almost all the projects are in Northern Europe. Eight of the planned projects are in Germany, which has a policy requiring heat reuse in large data centers by 2026.[39] Operating and planned projects with data from OCP are shown in Table 2.4-2.

Some significant investments in heat reuse projects are underway. Google is spending $1 billion to expand its Hamina, Finland data center and will give the heat for free to the local community.[44] It expects to provide 80% of the heating needs of the community. (There are no details on the summer heat management approach at this site.) Nordic data center provider atNorth has announced a hyperscale facility in Finland, which will open with 60 MW and provide heat to the community, with plans to expand to several hundred megawatts.[40] One source suggests that data centers in Europe near existing district heating networks could be capable of supplying 75 TWh/year—enough for 10% of the EU's heating demand by 2030.[41]

Table 2.4-2. Operating and planned heat reuse at data centers (adapted from OCP database[38]).

Site Name	Company/ Operator	City	Country/ State	Nominal reuse capacity (MW)
Operating Sites, Built Stage				
NREL	NREL	Golden	Colorado	5
Westin Building	Amazon Web Services	Seattle	Washington	5
Ericsson	Ericsson	Rosersberg	Sweden	10
Green Hub	GreenHub Data	Stockholm	Sweden	40
Stockholm Data Park	Centers	Stockholm	Sweden	10
LUMI Supercomputer	Stockholm Exergi	Kajaani	Finland	10
Ficolo	CSC	Vantaa	Finland	10
Green Computing	Ficolo Oy	Paris	France	20
Algae Farm	Green IT Solutions	Enge-Sande	Germany	15
nLighten	WindCloud	Hannover	Germany	5
Equinix AM3	NorthC Datacenters	Amsterdam	Netherlands	14
DC2-Telemark	Equinix	Rjukan	Norway	20
Planned Sites, Planning Stage				
DC Val d Europe	OVHCloud	Val d'Europe	France	7.8
DC Heated Night Club	Mainova Webhouse	Frankfurt Seckbach	Germany	30
Equinix FR4, FR6 & FR8	Equinix	Frankfurt Griesheim	Germany	56
Franky	Telehouse	Frankfurt Gallus	Germany	14
Stack 80 MW data Center	Stack Infrastructure	Taunus	Germany	80
Data Center at ICE Train Station	Stack Infrastructure	Limburg	Germany	35
Digital Park	Digital Realty	Frankfurt Fechenheim	Germany	20
heiCOMACS	heiCOMACS Forschungszentrum	Heidelberg	Germany	54
Jupiter Project	Jülich	Jülich	Germany	15
Elementica	Elementica	Stockholm	Sweden	21
EcoDataCenter2	EcoDataCenter	Östersund	Sweden	20
Qscale Q01 Campus	Qscale	Quebec	Canada	96
Wyoming Hyperscale	Wyoming Hyperscale	Evanston	Wyoming	120

In contrast, Meta's Richland Parish (Holly Ridge), Louisiana data center announcement makes no mention of including any heat reuse approaches.[42] While the project is significant (it is projected to be a 4 million ft^2, $10 billion campus with up to 2 GW compute capacity), it is not near a district heating opportunity. The benefits of other heat reuse options are apparently not sufficient to warrant investment.

As of 2024, NREL was developing the Urban Renewable Building and Neighborhood Optimization (URBANopt) platform to analyze the use of waste heat sources within geographically cohesive building districts. This tool will facilitate integration of waste heat into district energy systems, enhancing overall energy efficiency.[43]

Table 2.4-3 summarizes key factors in considering heat reuse projects in countries around the world.

Table 2.4-3. *Heat reuse economic factors by country.*

Region	Low-Hanging Fruit	Net Cost Insights	Future Centers' Economics
USA	Close-coupled retrofits, hybrid cooling	Hybrid cooling yields positive ROI; immersion retrofits costly[43]	New liquid-designed data centers could deliver lower capex/opex
EU	Free cooling, district heat export	12.2% cost savings via heat reuse; 70% chiller reduction via free cooling	Immersion includes long-term gains; retrofit less viable
UK	Free cooling retrofits	Similar cost benefits to EU; policy incentives still needed	Future builds will favor hybrid and immersion systems
Japan/APAC	Hybrid cooling; free cooling in cooler regions	Fewer studies, but expect similar savings when applied	Fastest-growing immersion adoption; cost benefits likely
China	Hybrid cooling pilots	Retrofitting remains expensive without pipeline incentives	Large new high-performance computing builds suitable for immersion from start
India	Hybrid during cooler seasons	Limited data; cooling efficiency still key	Immersion adoption early; local costs variable
South Korea	District cooling pairing	Waste heat flexible in district networks; economies rising	Immersion viable in new developments
Singapore	Free cooling	Tropical climate limits reuse; focus on cooling energy savings	High-density future data centers may offset higher initial cost
Russia	Close-coupled retrofits, free cooling	Subsidized energy reduces ROI action	New builds likely stick with efficient air; liquid niche
Finland	Free cooling, district heat export	Heat export drives clear ROI; colder climate favorable	New builds with immersion + heat export are highly cost-effective

D. Recommendations

1. *Data center operators should* ***adopt high-temperature liquid-cooling systems****—such as direct-chip or immersion cooling—that achieve exit temperatures of 45-70 °C, enabling effective heat reuse in applications, such as district heating.*

2. *National and subnational governments* ***should require feasibility studies for heat reuse in permitting large new data-center projects and offer incentives, such as fast-track permitting and subsidies, to deploy such systems****. National and subnational governments should consider requiring 10-20% heat reuse mandates for new data centers (such as in Germany).*

3. *District heating utilities and municipal planners should* ***proactively partner with data-center developers to map potential synergies and create or extend thermal infrastructure*** *that connects data centers to buildings, industrial users and aquaculture facilities.*

4. *Heat host industries (e.g., hospitals, laundries, greenhouses and industrial processes) should* ***actively engage with data-center operators to explore using waste heat for 24/7 applications,*** *including agriculture, drying, aquaculture and wastewater treatment.*

5. *Technology developers and standards organizations should* ***produce guidelines, matchmaking tools and technoeconomic frameworks that facilitate collaboration between data-center operators and prospective heat hosts****, building on the work of the Open Compute Project (OCP) and others.*

6. *Research institutions, utilities and innovative companies should* ***pilot alternative uses and technologies—such as data-center-powered DAC systems****—evaluating performance and return on investment to increase reuse pathways.*

E. References

1. National Renewable Energy Laboratory (NREL). High-Performance Computing Data Center Waste Heat Reuse; NREL, Golden, Colorado, https://www.nrel.gov/computational-science/waste-heat-energy-reuse (Accessed August 2025).

2. Cosimo Pecchioli, Jaime Comella, David Sickinger & Otto VanGeet. Data Centers Heat Reuse 101; Open Computer Project (OCP), Austin, Texas, https://www.opencompute.org/documents/20230623-data-centers-heatreuse-101-3-2-docx-pdf (2023).

3. Kun Zhou, Xiaolin Yu, Baogang Xie, Hongxi Xie & Wei Fu. Immersion cooling technology development status of data center. Sci. Tech. Energ. Transition 79, 41 (2024). https://doi.org/10.2516/stet/2024022.

4. National Renewable Energy Laboratory (NREL). High-Performance Computing Data Center; NREL, Golden, Colorado, https://www.nrel.gov/computational-science/hpc-data-center (Accessed August 2025).

5. Yang Jie. Novel Ideas to Cool Data Centers: Liquid in Pipes or a Dunking Bath; Wall Street Journal, New York, New York, https://www.wsj.com/tech/novel-ideas-to-cool-data-centers-liquid-in-pipes-or-a-dunking-bath-cdb7a393 (2024).

6. Uptime Institute. Uptime Institute 2024 Cooling Systems Survey: Direct Liquid Cooling Results; Uptime Institute, New York, New York, https://datacenter.uptimeinstitute.com/ (2024).

7. Archive Market Research (AMR). Rear Door Heat Exchangers (RDHx) Unlocking Growth Potential: Analysis and Forecasts 2025-2033; AMR, Pune, Maharashtra, India, https://www.archivemarketresearch.com/reports/rear-door-heat-exchangers-rdhx-486472 (2025).

8. Yulin Wang & James Edmondson. Thermal Management for Data Centers 2023-2033: Technologies, Markets and Opportunities; IDTechEx, Boston, Massachusetts, https://www.idtechex.com/en/research-report/thermal-management-for-data-centers/944 (Accessed August 2025).

9. Tobias Mann. More than a third of enterprise datacenters expect to deploy liquid cooling by 2026; The Register, San Francisco, California, https://www.theregister.com/2024/04/22/register_liquid_cooling_survey/ (2024).

10. BIS Research. Why Data Center Cooling Is the Next Big Tech Battleground; MarketResearch.com, Rockville, Maryland, https://blog.marketresearch.com/why-data-center-cooling-is-the-next-big-tech-battleground (2025).

11. Peter Judge. Air cooling will never go away; Data Center Dynamics, London, United Kingdom, https://www.datacenterdynamics.com/en/analysis/air-cooling-will-never-go-away/ (2023).

12. Kat Henshaw. Intel And Submer Unite To Transform Datacenter Heat Re-use To Power Municipalities; Submer, Barcelona, Spain, https://submer.com/blog/press/intel-and-submer-unite-to-transform-datacenter-heat-re-use-to-power-municipalities/ (2022).

13. Wikipedia. Aquasar; San Francisco, California, https://en.wikipedia.org/wiki/Aquasar (Accessed August 2025).

14. Wikipedia. iDataCool; San Francisco, California, https://en.wikipedia.org/wiki/IDataCool (Accessed August 2025).

15. Sharon Fisher. Data Centers Leverage Cooling to Heat Homes; Data Center Knowledge (Informa TechTarget), Boston, Massachusetts, https://www.datacenterknowledge.com/cooling/data-centers-leverage-cooling-to-heat-homes (2023).

16. European Heat Pump Association (EHPA). Waste into wealth: how heat pumps can recycle heat to save energy - and the EU policy that matters (Position Paper); EHPA, Brussels, Belgium, https://www.ehpa.org/wp-content/uploads/2024/12/

Waste-into-wealth-through-industrial-heat-pumps-report_EHPA_December-2024.pdf (2024).

17. International Energy Agency (IEA). District Heating; IEA, Paris, France, https://www.iea.org/energy-system/buildings/district-heating (Accessed August 2025).

18. US Department of Energy (DOE) Office of Energy Efficiency & Renewable Energy (EERE). Combined Heat and Power Technology Fact Sheet Series: District Energy Systems Overview; US DOE EERE, Washington, DC, https://www.energy.gov/sites/default/files/2021/03/f83/District_Energy_Fact_Sheet.pdf (2021).

19. State of Green. District energy: The backbone of a flexible, resilient and efficient energy system (White Papers for a Green Transition); Copenhagen, Denmark, https://dbdh.org/wp-content/uploads/2024/11/SoG_WhitePaper_DistrictEnergy2024_210x297_V09.pdf (2024).

20. Microsoft. Surplus datacenter heat will be repurposed to heat homes in Denmark; Redmond, Washington, https://local.microsoft.com/blog/datacenter_heat_repurposed/ (Accessed August 2025).

21. Terje Solsvik. Denmark's European Energy signs long-term power deals with Microsoft; Reuters, London, United Kingdom, https://www.reuters.com/business/energy/denmarks-european-energy-signs-long-term-power-deals-with-microsoft-2024-06-12/ (2024).

22. Tom Kerr. CHP/DHC Country Scorecard: United States; International Energy Agency (IEA) International CHP/DHC Collaborative, Paris, France, https://www.energy.gov/sites/prod/files/2013/11/f4/chp_profile_united_states.pdf (2013).

23. Siyue Guo & Shiyan Chang. Waste heat recovery of data centers for urban heating in China; Tsinghua University Institute of Energy, Environment and Economy, Beijing, China, https://c2e2.unepccc.org/wp-content/uploads/sites/3/2024/11/guo-siyu-ses2024-submit-waste-heat-recovery-of-data-centers-for-urban-heating-in-china.pdf (2024).

24. Tencent. Tencent Carbon Neutrality Target and Roadmap Report; Tencent, Shenzhen, China, https://static.www.tencent.com/attachments/TencentCarbonNeutralityTargetandRoadmapReport.pdf (2022).

25. CDR.fyi. Homepage; New York, New York, https://www.cdr.fyi/ (Accessed August 2025).

26. Zachary Skidmore. Microsoft pilots use of data center waste heat to power Direct Air Capture systems; Data Center Dynamics, London, United Kingdom, https://www.datacenterdynamics.com/en/news/microsoft-pilots-use-of-data-center-waste-heat-to-power-direct-air-capture-systems/ (2025).

27. Microsoft. 2025 Environmental Sustainability Report: Accelerating progress to 2030; Redmond, Washington, https://cdn-dynmedia-1.microsoft.com/is/content/microsoftcorp/microsoft/msc/documents/presentations/CSR/2025-Microsoft-Environmental-Sustainability-Report-PDF.pdf (2025).

28. Sebastian Moss. Meta and Alphabet's X developing direct air capture systems using data center waste heat; Data Center Dynamics, London, United Kingdom, https://www.datacenterdynamics.com/en/news/meta-and-alphabets-x-developing-direct-air-capture-systems-using-data-center-waste-heat/ (2024).

29. Reed Albergotti. Amazon AI data centers could double as carbon capture machines; Semafor, New York, New York, https://www.semafor.com/article/12/01/2024/amazon-ai-data-centers-to-double-as-carbon-capture-machines (2024)

30. Ormat. Geothermal power: Ormat is a world class developer with unrivaled expertise Reno, Nevada, https://www.ormat.com/en/renewables/geothermal/main/ (Accessed August 2025).

31. Yu Him (Mike) Au, Olivia Lattanzi, Daniel Seeley & Claire Victor. A Comprehensive Mechanical Engineering Perspective on the Implementation of an Organic Rankine Cycle for Data Center Waste Heat Recovery; Worcester Polytechnic Institute (WPI), Worcester, Massachusetts, https://cpb-us-w2.wpmucdn.com/wp.wpi.edu/dist/9/451/files/2021/04/R_3CQ8vKgrQkLQbWa_DataCenterORC_MQP_Sustainability.pdf (2024).

32. Russell Tipton, Gregory Kremer & Muhammad Ali. Experimental lessons of an organic Rankine cycle applied for waste heat recovery in a data center environment (Vertiv White Paper); Vertiv Group Corp., Westerville, Ohio, https://www.vertiv.com/49d45e/globalassets/documents/white-papers/vertiv-experimental-lessons-of-organic-rankine-wp-en-na-sl-71085-web_383464_0.pdf (2023).

33. Carrie A. Schoeneberger, Colin A. McMillan, Parthiv Kurup, Sertac Akar, Robert Margolis & Eric Masanet. Solar for industrial process heat: A review of technologies, analysis approaches, and potential applications in the United States. Energy 206, 118083 (2020). https://doi.org/10.1016/j.energy.2020.118083.

34. Sten Falkum. The world's largest land-based trout farm will use waste heat from data center; Hima Seafood, Oslo, Norway, https://himaseafood.com/trout-farm-will-use-waste-heat-from-data-center/ (Accessed August 2025).

35. Richard Pallardy. Reusing Waste Heat from Data Centers to Make Things Grow; Information Week (Informa TechTarget), Boston, Massachusetts, https://www.informationweek.com/sustainability/reusing-waste-heat-from-data-centers-to-make-things-grow (2024).

36. Eco Data Center. EcoDataCenter 1: A large-scale data center site for HPC, AI in Falun, Sweden; Falun, Sweden, https://ecodatacenter.tech/data-center/ecodatacenter-1 (Accessed August 2025).

37. Bharath Ramakrishnan & Jaime Comella. Heat Reuse; Open Compute Project (OCP), Austin, Texas, https://www.opencompute.org/projects/heat-reuse (Accessed August 2025).

38. Open Compute Project (OCP). 2024-08_OCP-HR_Data-Centers-with-Heat-Reuse-data base; OCP, Austin, Texas, https://www.opencompute.org/w/index.php?title=Cooling_Environments/Heat_Reuse (Accessed August 2025).

39. Rostyslav Telyatnykov, Thomas Burmeister & Petra Karin Iffert (née Kistner). Data center requirements under the new German Energy Efficiency Act; White and Case LLP, New York, New York https://www.whitecase.com/insight-alert/data-center-requirements-under-new-german-energy-efficiency-act (2023).

40. Peter Judge. atNorth announces 60MW+ data center in Finland - with heat reuse; Data Center Dynamics, London, United Kingdom, https://www.datacenterdynamics.com/en/news/atnorth-announces-60mw-data-center-in-finland-with-heat-reuse/ (2023).

41. Edward Wheatcroft, Henry Wynn, Kristina Lygnerud & Giorgio Bonvicini. The role of low temperature waste heat recovery in achieving 2050 goals: a policy positioning paper. arXiv:1912.06558 (2019). https://doi.org/10.48550/arXiv.1912.06558.

42. Hiller, Jennifer. "The AI Data-Center Boom Is Coming to America's Heartland." Wall Street Journal, March 30, 2025. https://www.wsj.com/business/energy-oil/the-ai-data-center-boom-is-coming-to-americas-heartland-eb060a32?gaa_at=eafs&gaa_n=ASWzDAgtIAgtzImchjcQcHTomvQuXR6UEu9tIfxdxsQEZq2acTtvfiaLKxhgjhsB6Rw%3D&gaa_ts=68e25e81&gaa_sig=EXW2J5PJLVZZRfgaZO2kB1hsqzGOQENcYy7oHYDtmE1RnSTZIcAoKZuw0aS81CksHScdqGtANmVDrgAMSzPDug%3D%3D (2025).

43. Data Center Frontier. Hybrid Cooling: A Cost and ROI Analysis Lawrenceville, New Jersey, https://www.datacenterfrontier.com/sponsored/article/55289185/trane-hybrid-cooling-a-cost-and-roi-analysis (2025).

44. Matthew Goodings. Google launches heat recovery project at data center in Hamina, Finland, https://www.datacenterdynamics.com/en/news/google-launches-heat-recovery-project-at-data-center-in-hamina-finland/ (2024).

TEXT BOX

Data Center Energy Efficiency Metrics

Eric Masanet

Over the past two decades, numerous data center energy efficiency metrics have been proposed, yet arguably only one has been adopted on a widespread basis across data center operators, regions and data center types.

This metric is known as **"power usage effectiveness" or PUE**. First introduced in 2007 by the Green Grid,[1] and later institutionalized in the ISO/IEC 30134-2 standard,[2] PUE is calculated as the ratio of a data center's total energy use to the energy use of its IT equipment alone (see Chapter 2.3 of this Roadmap). As such, it provides a measure of how much energy goes to non-IT end uses that do not provide business value. Its theoretical lower limit is 1.0. Several hyperscale operators (e.g., Google, Meta and AWS) regularly achieve fleet averages approaching PUEs of 1.1.

However, PUE does not capture the efficiency of IT equipment nor the efficiency of IT workloads in the data center. Indeed, two data centers with the same PUE can have substantially different IT equipment and workload efficiencies, meaning that PUE cannot be used as a stand-alone comparison between data centers.[3] (See Chapter 2.1 of this Roadmap).

The **"IT power usage effectiveness" (ITUE)** metric aims to better capture how much rack power results in useful computations. It is calculated as the ratio of total energy into IT equipment to the energy use of IT compute components alone. Multiplying ITUE by PUE results in a metric known as **"total power usage effectiveness" (TUE)**, which is the ratio of total data center energy use to the energy use of IT compute components alone. While ITUE and TUE were proposed and demonstrated around 15 years ago, reporting of either is difficult to find.[4,5]

ITUE and TUE can capture the efficiency of providing power to compute components; however, they fall short of measuring the useful work actually done by these components. To address this gap, several metrics have been proposed to quantify workload-level energy efficiencies, including the Green Grid's **"data center energy productivity" (DCeP)**,[6] which is calculated as useful work produced divided by total energy consumed by the data center, and **"server energy productivity" (SEP)**,[7] which compares the energy consumption of a server in relation to the share of compute work the server is performing. Like ITUE and TUE, however, reporting of these metrics by data center operators is difficult to find.

Most recently, the Green Grid has provided additional guidance for standardizing the calculation of server work capacity for use in "work per unit energy" metrics and reporting moving forward. This metric is known as "**IT Work Capacity (ITWC)**."[8]

Importantly, these metrics and indexes do not accurately capture key environmental parameters, including water consumption or energy-related greenhouse gas emissions. Other metrics have been proposed to quantify, for example, cooling efficiency, waste heat reuse, and energy-related factors like carbon utilization effectiveness and shares of renewable energy. For a detailed review of metrics, see Sustainable Digital Infrastructure Alliance (SDIA) (2025).[7]

References

1. The Green Grid. PUE™: A Comprehensive Examination of the Metric (White Paper #49); San Ramon, California, https://datacenters.lbl.gov/sites/default/files/WP49-PUE%20A%20Comprehensive%20Examination%20of%20the%20Metric_v6.pdf (2012).

2. International Organization for Standardization (ISO). ISO/IEC 30134-2:2016 Information technology — Data centres — Key performance indicators—Part 2: Power usage effectiveness (PUE); ISO, Geneva, Switzerland, https://www.iso.org/standard/63451.html (2016).

3. Eric Masanet, Arman Shehabi & Jonathan Koomey. Characteristics of low-carbon data centres. Nature Climate Change 3, 627-630 (2013). https://doi.org/10.1038/nclimate1786.

4. Dan Swinhoe. Is PUE too long in the tooth?; Data Center Dynamics (DCD), London, United Kingdom, https://www.datacenterdynamics.com/en/analysis/is-pue-too-long-in-the-tooth/ (2022).

5. Michael K. Patterson, Stephen W. Poole, Chung-Hsing Hsu, Don Maxwell, William Tschudi, Henry Coles, David J. Martinez & Natalie Bates. “TUE, a New Energy-Efficiency Metric Applied at ORNL’s Jaguar” in International Supercomputing Conference (ISC) 2013, Berlin, Heidelberg. 372-382, https://doi.org/10.1007/978-3-642-38750-0_28 (2013).

6. Jason Verge. The Green Grid Unveils Energy Productivity Metric for Data Centers; Data Center Knowledge (Informa TechTarget), Newton, Massachusetts, https://www.datacenterknowledge.com/sustainability/the-green-grid-unveils-energy-productivity-metric-for-data-centers (2014).

7. Sustainable Digital Infrastructure Alliance (SDIA). Data Center Metrics; SDIA, Hamburg, Germany, https://knowledge.sdialliance.org/archive/data-center-metrics (Accessed August 2025).

8. The Green Grid's ITE Metrics Work Group. IT Work Capacity metric V1 - A Methodology (White Paper #94); The Green Grid Inc., San Ramon, California, https://www.thegreengrid.org/resources/library-and-tools/wp-94-itwc-methodology-calculate-server-work-capacity-cserv-data (2025).

3.1 On-Site Greenhouse Gas Emissions (Scope 1)

Colin McCormick

Some equipment at data centers emits greenhouse gases. Diesel generators for backup power emit carbon dioxide (CO_2), and cooling equipment and fire suppression systems can leak hydrofluorocarbons (HFCs). Data concerning these emissions are very limited. This chapter discusses greenhouse gas emissions from equipment at data centers (referred to as direct, on-site or Scope 1 emissions) and mitigation strategies.

(Some data centers have "behind-the-meter" generators for primary, non-backup power, but this is rare. Most data centers get their primary power from the grid. Emissions from primary power for data centers are discussed in Chapter 3.2 of this Roadmap.)

A. Backup Power Systems

i. Emissions

Diesel generators are the most common backup power system at data centers.[1] In combination with uninterruptible power supply (UPS) systems, diesel generators are capable of starting within approximately 10 seconds of a grid power supply outage and can continue to supply power at data centers indefinitely, depending on on-site fuel reserves.[2] While operating, backup generators burn diesel fuel and emit CO_2, as well as air pollutants such as nitrogen oxides (NOx) and particulate matter (PM).[3]

These generators are primarily intended to be used for emergency situations, and many jurisdictions limit the total number of hours they can be run annually. For example, the US state of Virginia allows backup generators to operate for up to 500 hours per year for all purposes.[4] However, the US Environmental Protection Agency (EPA) allows backup diesel generators at data centers to be operated for up to 100 hours per year for testing and maintenance, up to 50 hours of which can be part of a non-emergency demand response program to support grid reliability.[5]

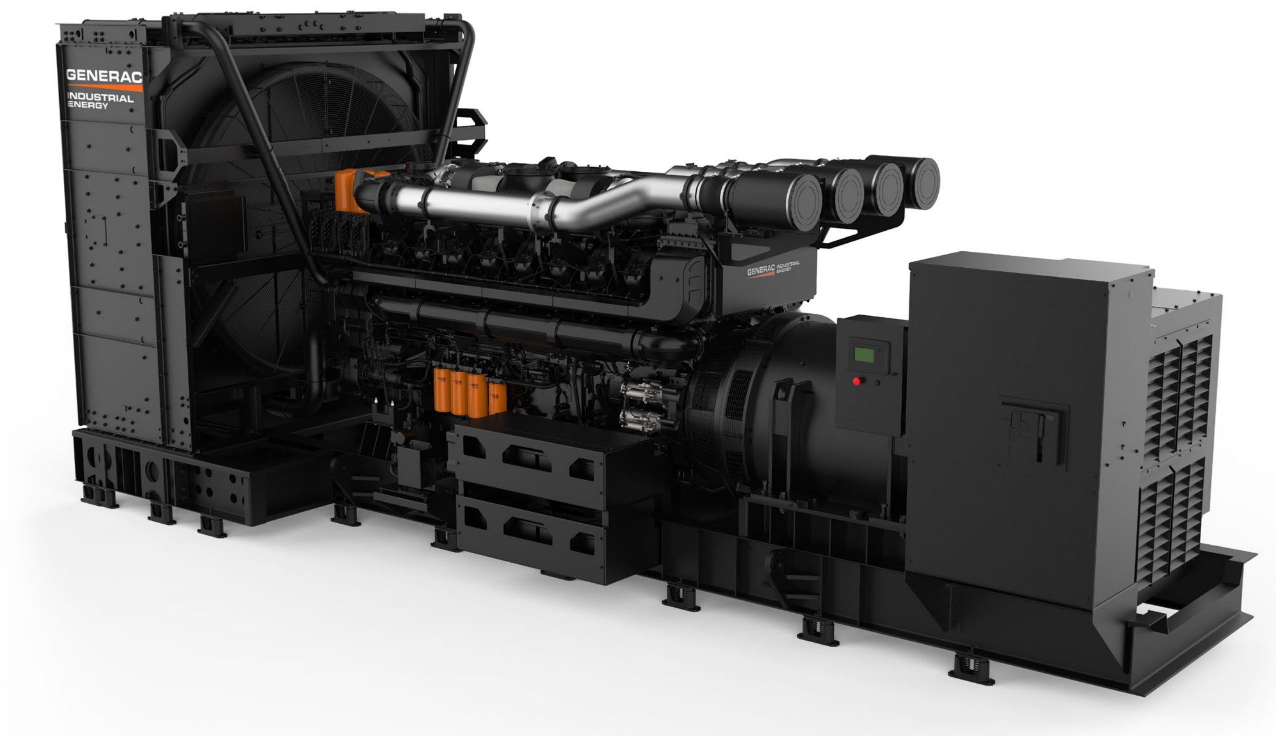


Figure 3.1-1. *A 3 MW diesel generator.*

Modern backup diesel generators emit approximately 0.79 kgCO_2 per kWh of power,[6] far higher than the average carbon intensity of electric grids globally (0.48 kgCO_2/kWh).[7] A typical 3 MW diesel generator operating for the maximum 500 hours in a year would therefore emit approximately 1000 tCO_2 annually. During grid emergencies, backup generators are designed to serve the full critical information technology (IT) load, meaning that a hypothetical 100 MW data center running on backup generators for 500 hours per year would lead to approximately 39.5 ktCO_2 of emissions.

However, this is typically far lower than the total emissions that result from the grid-supplied electricity for this data center during the rest of the year. In this example, if the remaining 8260 hours in the year were served by average US grid power (rather than diesel backup generators) that would result in 323 ktCO_2 of emissions. (These would be considered Scope 2 under the Greenhouse Gas Protocol.)

Some data centers use natural gas turbines and reciprocating engines[8] or combined heat and power (CHP) systems for backup power (and to reduce peak demand charges). Natural gas-fired generation, while less CO_2-intensive than diesel generators, still contributes Scope 1 CO_2 and potentially methane emissions during sustained or intermittent operation.[9] CHP systems at data centers also produce Scope 1 emissions if powered by fossil fuels.

ii. Mitigation strategies

Options to reduce greenhouse gas emissions from on-site diesel generators at data centers include using natural gas generators, renewable fuels, batteries and hydrogen fuel cells. All these strategies have challenges or limitations.

Some data center operators have begun to adopt natural-gas-fired backup generators in place of diesel, which can moderately reduce CO_2 and NOx emissions. Natural-gas-fired generators are increasingly available with fast start-up times that can approach or match the performance of diesel. In addition to the fact that emissions reductions from this strategy are limited, one of the challenges for adopting it is a perception that natural gas supply may not be as reliable during emergency situations as diesel fuel, which is typically stored on site and re-supplied via truck rather than delivered via pipeline.[10]

Renewable fuels, such as hydrotreated vegetable oil (HVO), biodiesel (FAME) and synthetic paraffinic fuels, can be used as drop-in replacements for conventional diesel (with some caveats) and have been adopted by some data center operators.[11] They offer significant reductions in lifecycle CO_2 emissions while maintaining compatibility with existing diesel generator systems.[12] However, reliable supplies of these fuels are not universally available, and biodiesel is typically subject to a maximum blending limit with conventional diesel for drop-in use.

Battery energy storage systems can reduce life-cycle CO_2 emissions compared with diesel generators, depending principally on the type of power used to charge the batteries. If charged from low-carbon power sources, batteries can help reduce emissions; however in general, data center energy storage systems charge from grid-supplied power and/or on-site generators, in which case there may be no emissions benefits. The most relevant energy storage technologies include lithium-ion batteries and sodium-sulfur batteries.[13] However, their ability to serve load is generally limited in duration, and longer-duration storage technologies are attracting increasing attention.[14]

Hydrogen fuel cells provide another alternative backup power solution, with the advantage of integrating on-site hydrogen storage[15] to extend the duration over which they can serve load. While there are no on-site CO_2 or NOx emissions from fuel cell operation, life-cycle CO_2 emissions depend on the hydrogen production method, meaning that sources of low-carbon hydrogen are required to make this strategy result in significant emissions reductions compared to diesel. Similar to renewable fuels, supply of low-carbon hydrogen is more constrained than diesel and natural gas, which some operators may see as a barrier to adoption.

The use of energy storage and on-site generation must be carefully coordinated, which can be accomplished using a microgrid.[16] This approach enables the data center operator to have increased control and flexibility over the power supply to critical IT loads, improving resilience during grid outages and introducing the capability of providing load flexibility as a service to the grid (see Chapter 4 of this Roadmap). Microgrids are receiving increasing attention from data center operators and are particularly important for high-capacity data centers in regions with constrained grids.[17,18]

Carbon capture is not a good candidate for mitigating emissions from diesel generators at data centers, mainly due to the intermittent nature of backup generators and their relatively small capacity. Carbon capture may be a good mitigation strategy for primary power generation for data centers (see Chapter 3.2 of this Roadmap).

Fig. 3.1-2. *Multiple diesel generator sets, of the kind used to provide reliable backup power at data centers.*

B. Cooling and Fire Suppression Systems

Cooling systems in data centers primarily use HFC refrigerants such as R-134a (see Chapter 2.3 of this Roadmap). These F-gases have global warming potentials (GWPs) well over 1000, meaning that even minor leaks can lead to high CO_2-equivalent emissions. Leakage can occur during standard operation, maintenance or equipment decommissioning.[19] Leakage rates of heating, ventilation and air condition (HVAC) refrigerants vary significantly, depending on equipment type and age, ranging from roughly 1% to 15% per year.[20,21] For a hypothetical 100 MW data center using the current generation of refrigerants, high GWP refrigerant leaks could be equivalent to roughly 230 (low leakage) to 3500 (high leakage) tons of CO_2 equivalent (tCO_2e)[a] per year. With low GWP refrigerants, the equivalent CO_2e emissions from leaks are dramatically reduced, helping to significantly lower the overall climate impact.

Certain fire suppression agents used in data centers, like HFC-227ea (FM-200), also possess high GWPs. Though infrequent, discharges—whether accidental, during testing, or at system end-of-life—can also contribute to greenhouse gas emissions.[22] While the average leakage rates are unknown, they are likely to be significantly smaller than those for HVAC systems given the much lower total charge of refrigerant used in these systems.

Regulatory frameworks, such as the US EPA Section 608 program and the EU F-Gas Regulation, now mandate tighter leak detection, maintenance and phasedown of high-GWP refrigerants. In the United States, data center cooling equipment will be restricted to refrigerants with a maximum GWP of 700, with a compliance date of 2027.[23] Similar European requirements are also phasing in. For fire suppression systems, commonly used high-GWP F-gases (notably FM-200) are being phased down under the Montreal Protocol.[24] The adoption of next-generation refrigerants and fire suppression substances, combined with practicing robust leak detection and repair, can enable data centers to significantly reduce greenhouse gas emissions from these systems. A complementary approach for reducing refrigerant emissions is to adopt alternative cooling strategies (see Chapter 2.3 of this Roadmap).

a This is based on the following: A hypothetical 100 MW data center would require approximately 35,000 tons of cooling capacity. The refrigerant charge in this system would likely be approximately 18 tons (typical charge values are roughly 0.5 kg of refrigerant per ton of cooling). Using the GWP of R-134a (1300), a commonly used refrigerant in large HVAC systems, a 1% annual leak rate would be equivalent to approximately 230 tCO_2e/year, while a 15% leak rate would be equivalent to approximately 3500 tCO_2e/year.

C. Recommendations

1. *Data center operators should* ***examine alternatives to the continued use of diesel for on-site backup generation****. This could include alternative drop-in fuels with low-carbon intensity where available or the adoption of low-carbon backup generation, such as hydrogen fuel cells and the use of on-site energy storage.*

2. *Data center operators should* ***ensure that HVAC and fire suppression equipment leak detection protocols are modernized and carefully implemented to reduce F-gas leakage.*** *They should also closely follow regulatory developments around adopting advanced, low-GWP refrigerants and fire-suppression equipment.*

3. *Governments should* ***review current limitations on maximum operating limits for diesel backup generators*** *to ensure that air quality impacts and greenhouse gas emissions are minimized.*

4. *Utilities should* ***continue to meet high grid reliability performance targets****, reducing the need for on-site backup generation at data centers.*

D. References

1. Brags & Hayes Generators. How to choose the right power generator for data center infrastructure; Pembroke Park, Florida, https://bnhgenerators.com/how-to-choose-the-right-power-generator-for-data-center-infrastructure/ (Accessed August 2025).

2. Brian O'Connor. An Overview of NFPA 110; National Fire Protection Association (NFPA), Quincy, Massachusetts, https://www.nfpa.org/news-blogs-and-articles/blogs/2023/01/23/an-overview-of-nfpa-110 (2023).

3. Nicole Dierksheide. Powering tomorrow: What role do diesel generators play in the future of clean energy solutions?; Data Center Dynamics (DCD), London, United Kingdom, https://www.datacenterdynamics.com/en/opinions/powering-tomorrow-what-role-do-diesel-generators-play-in-the-future-of-clean-energy-solutions/ (2024).

4. Virginia Department of Environmental Quality (DEQ). Emergency Generator Air Permit Guidelines (Clarification #2025-01); Virginia DEQ, Richmond, Virginia, https://www.deq.virginia.gov/home/showpublisheddocument/27422 (2025).

5. US Environmental Protection Agency (EPA). Fact Sheet and Frequently Asked Questions: Use of Backup Generators to Maintain the Reliability of the Electric Grid; US EPA, Washington, DC, https://www.epa.gov/system/files/documents/2025-05/rice-memo-on-duke-energy-regulatory-interpretation-04_17_25.pdf (2025).

6. PC (FEA) Facilities Engineering Associates. The Carbon-Footprint of Diesel Generators; FEA, Fairfax, Virginia, https://www.feace.com/single-post/the-carbon-footprint-of-diesel-generators (2022).

7. Malgorzata Wiatros-Motyka, Dave Jones & Nicolas Fulghum. Chapter 5: Major Countries and Regions in Global Electricity Review 2024 (Ember, London, United Kingdom, 2024, https://ember-energy.org/latest-insights/global-electricity-review-2024/major-countries-and-regions/).

8. Capstone Green Energy. Microturbines vs. Reciprocating Engines: A Deep Dive into Power Generation Technologies (Smart Energy Blog); Capstone Green Energy Holdings, Inc., Los Angeles, California, https://www.capstonegreenenergy.com/turbine-talk/post/10070/microturbines-vs-reciprocating-engines-a-deep-dive-into (2023).

9. International Energy Agency (IEA). Data Centres and Data Transmission Networks; IEA, Paris, France, https://www.iea.org/energy-system/buildings/data-centres-and-data-transmission-networks (Accessed August 2025)

10. Drew Thompson. Powering Data Centers with Natural Gas: A Report on the Benefits of Natural Gas for Data Center Backup Power; Black & Veatch Corporation, Overland Park, Kansas, https://www.bv.com/en-US/perspectives/powering-data-centers-with-natural-gas-a-report-on-the-benefits-of-natural (2020).

11. Heidi Herzog. Renewable Diesel: A Long-Term Solution for Decarbonization; Biodiesel Magazine, Grand Forks, North Dakota, https://biodieselmagazine.com/articles/renewable-diesel-a-long-term-solution-for-decarbonization (2025).

12. Caterpillar Inc. Power from Renewable Liquid Fuels: Biodiesel and HVO; Irving, Texas, https://www.cat.com/en_US/by-industry/electric-power/electric-power-industries/renewable-liquid-fuels.html (Accessed August 2025).

13. Viswambher Kambhampati, Andy van den Dobbelsteen & Joep Schild. Moving Beyond Diesel Generators: Exploring Renewable Backup Alternatives for Data Centers. Journal of Physics: Conference Series 2929, 012008 (2024). https://doi.org/10.1088/1742-6596/2929/1/012008.

14. Emma Penrod. Data centers could bring alternative battery types into the mainstream, developers say; Utility Dive (Informa TechTarget), Newton, Massachusetts, https://www.utilitydive.com/news/data-center-flow-zinc-battery-xl-eos-prometheus/751144/ (2025).

15. David Chernicoff. Data Center Hydrogen Power Is Making Strides, But Can It Ramp Fast Enough?; Data Center Frontier, Lawrenceville, New Jersey, https://www.datacenterfrontier.com/energy/article/33039386/data-center-hydrogen-power-is-making-strides-but-can-it-ramp-fast-enough (2024).

16. Carsten Baumann. How Microgrids for Data Centers Increase Resilience, Optimize Costs, and Improve Sustainability (White Paper 289, Version 1); Schneider Electric, Rueil-Malmaison, France, https://www.se.com/us/en/download/document/Microgrids_for_Data_Centers/ (2020).

17. Brian Nelson. Microgrids for Data Centers: Enhancing Uptime While Reducing Costs; Data Center Knowledge (Informa TechTarget), Newton, Massachusetts, https://www.datacenterknowledge.com/uptime/microgrids-for-data-centers-enhancing-uptime-while-reducing-costs (2025).

18. Lisa Cohn. Why do data center operators choose diesel backup over cleaner microgrids?; Microgrid Knowledge, Westborough, Massachusetts, https://www.microgridknowledge.com/distributed-energy/article/11427459/why-do-data-center-operators-choose-diesel-backup-over-cleaner-microgrids (2022).

19. US Environmental Protection Agency (EPA). Stationary Refrigeration and Air Conditioning; EPA, Washington, DC, https://www.epa.gov/section608 (Accessed August 2025).

20. Victor Heinerud & André Sahlsten. Natural Refrigerants in Data Center Cooling with Thermosiphon Application; KTH Diva Portal, Stockholm, Sweden, https://kth.diva-portal.org/smash/record.jsf?pid=diva2%3A972685&dswid=9235.(2016).

21. Alexandre F. Santos, Pedro D. Gaspar & Heraldo J. L. de Souza. Ecoenergetic Comparison of HVAC Systems in Data Centers. Climate 9, 42 (2021). https://doi.org/10.3390/cli9030042

22. US Environmental Protection Agency (EPA) Center for Corporate Climate Leadership. Greenhouse Gas Inventory Guidance: Direct Fugitive Emissions from Refrigeration, Air Conditioning, Fire Suppression, and Industrial Gases; US EPA, Washington, DC, https://www.epa.gov/sites/default/files/2015-07/documents/fugitiveemissions.pdf (2014).

23. Vertiv. Evolving Refrigerant Regulations and Applications in Data Center Environments (Vertiv White Paper); Vertiv Group Corp, Westerville, Ohio, https://www.vertiv.com/48f7e3/contentassets/8eee3437fd734143a36d0395e65e8de7/vertiv-low-gwp-refrigerants-wp-en-na-sl-71228-web.pdf

24. Marta Pizano Bella Maranion, Ashley Woodcock. Montreal Protocol on Substances that Deplete the Ozone Layer: Report of the Technology and Economic Assessment Panel (Volume 1: Progress Report); United Nations Environment Programme (UNEP), Nairobi, Kenya, https://ozone.unep.org/system/files/documents/TEAP-May2025-Progress-Report-vol1.pdf (2025).

3.2 Power Supply Greenhouse Gas Emissions (Scope 2)

Colin McCormick

A. Overview

Modern data centers consume large amounts of electricity. The International Energy Agency (IEA) estimates that electricity consumption by data centers produced 180 million tons of carbon dioxide ($MtCO_2$) emissions in 2024 (roughly 0.5% of global CO_2 emissions). In the IEA Base Case Scenario, these CO_2 emissions roughly double by 2030. (See Chapter 1 of this Roadmap.)

Most electricity for data centers is purchased from the electric grid ("in front of the meter" or "FTM"). The resulting greenhouse gas emissions are included under Scope 2 of the Greenhouse Gas Protocol Corporate Standard.[1] This chapter discusses the greenhouse gas emissions implications of data centers' reliance on the grid, current debates on revising Scope 2 methodology, approaches to mitigating Scope 2 emissions for data centers, and related issues that arise with respect to greenhouse gas emissions from on-site generation ("behind the meter" or "BTM"). The chapter concludes with Recommendations.

B. Scope 2 Location-Based Emissions for Data Centers

The Greenhouse Gas Protocol provides two separate methods for calculating emissions from the use of grid-supplied electricity. These methods are known as "location-based" and "market-based," and they can produce significantly different results.[2] Thus, the best practice for corporate emissions disclosure is to report both.

The location-based method considers electricity emissions from an engineering perspective. Because of the nature of electric power flow, the grid's electricity comes from many different power plants (the "grid mix"). Because the power plants on a grid typically use different types of technologies (e.g., gas, coal, wind, solar, etc.), they have very different CO_2 emissions per unit of electricity generated. Determining location-based emissions of grid-supplied electric power therefore requires averaging the emissions of all power plants on the relevant grid.

Typically, the "relevant" grid is considered to be the balancing area (or control area) over which a single grid operator is responsible for ensuring that electricity generation and supply are continually matched. This can be at the national, sub-national or super-national level. Grids with large amounts of low-carbon power generation, such as renewables or nuclear, have very low average greenhouse gas emissions per unit of energy generated ("carbon intensity"). For example, hydro-rich Sweden averages 8 g of CO_2 equivalent (gCO_2e)/kWh and nuclear-rich France averages 48 gCO_2e/kWh.[3] In contrast, grids with large amounts of coal have very high values (e.g., 623 gCO_2/kWh for Indonesia[4]). The global average grid carbon intensity in 2024 was 473 gCO_2/kWh[5]; IEA expects this figure to fall by approximately 4% per year through 2026[6] as grids add more low-emitting power generation and as high-emitting power plants are retired.

Because of the large range in grid carbon intensities in different regions, Scope 2 location-based emissions for data centers are heavily influenced by their location. Identical data centers located in different regions could have Scope 2 location-based emissions that differ by almost 100x. For example, a hypothetical 100 MW data center with a PUE of 1.2 that operates at 80% capacity would have Scope 2 location-based emissions of 6727 tCO_2e/year in Sweden and 523,918 tCO_2e/year in Indonesia[a].

As noted above, IEA estimates that data centers accounted for 180 Mt of CO_2 emissions in 2024.[7] Under the IEA Base Case Scenario, data center energy consumption will grow to 945 TWh/year by 2030. If this power were supplied by electricity with the 2024 global average emissions intensity, it would result in 447 $MtCO_2$ of Scope 2 location-based emissions. However, if average grid carbon intensity falls as projected by IEA, the 2030 global average grid carbon intensity will be 370 gCO_2/kWh, and emissions from data centers in the IEA Base Case would be 350 $MtCO_2$. Data center growth will likely be concentrated in particular countries and regions, so a more detailed analysis of electricity emissions for these specific grids (rather than the global average) would be needed to improve these estimates.

While calculation of Scope 2 location-based emissions currently uses annual average grid emissions intensity, the hour-to-hour emissions intensity of some grids can vary by large amounts. This is particularly true for grids with high solar and wind penetration. They can experience some hours with essentially zero emissions (during sunny and/or windy conditions when power is mostly supplied by solar and wind) followed by other hours with high emissions (when power is mostly supplied by gas or coal). Emissions intensity on some grids can also change substantially from season to season, for similar reasons. While this variability can pose challenges for maintaining grid reliability (e.g., the solar "duck curve" in the US,[8] Australia,[9] India[10] and elsewhere), it also means that shifting electricity consumption to low-emissions hours can significantly reduce actual emissions from electricity use.[11] This has led to proposals for reforming the Scope 2 location-based methodology (see below). It has also helped motivate some data centers to use load flexibility (see Chapter 4 of this Roadmap).

---

a This is calculated as follows. By convention, the 100 MW data center with PUE of 1.2 would draw 120 MW of power from the electric grid. Operating at 80% capacity, it would consume 840,960 MWh of grid power during a single year. Multiplying by the average carbon intensity of the grids in Sweden and Indonesia gives 6,727 tCO_2e and 523,918 tCO_2e, respectively. However, the amounts are usually negligible.

C. Scope 2 Market-Based Emissions for Data Centers

The Scope 2 “market-based” method considers electricity emissions from a market perspective. This method begins with the location-based emissions amount and then replaces grid-supplied energy with any renewable energy that was procured through market-based instruments, such as power purchase agreements (PPAs), on an energy basis (i.e., on a MWh-for-MWh basis, see Box 3.2-1). Under this method, the greenhouse gas emissions from procured renewable energy are effectively zero,[b] meaning that Scope 2 market-based emissions can be significantly reduced or even brought all the way to zero by procuring a sufficient amount of renewable energy. Many data center operators procure large amounts of renewable energy through PPAs for this purpose. The four large hyperscalers (Amazon, Google, Meta and Microsoft) are the largest corporate buyers of renewable energy globally, having collectively procured more than 84 GW via PPAs as of early 2025.[12] Thus, they report far lower Scope 2 market-based emissions than location-based emissions. (Not all hyperscalers report both location-based and market-based emissions, making this comparison difficult in some cases.)

The Greenhouse Gas Protocol Scope 2 guidance is under revision in 2025, with a final version expected in 2027. While the general concepts of separate location- and market-based emissions reporting will likely remain, the methods used to determine

b In some cases, procured renewable energy may be deemed to have a small amount of greenhouse gas emissions for the purposes of the Scope 2 market-based method. However, this is not typical, and the amounts are usually negligible.

the actual amount of emissions offsets will change. Several alternative approaches have been proposed, with the intention of more closely matching the physical nature of renewable power supply (such as time-variation in the grid mix) and/or more closely addressing overall impacts of power generation[13] on CO_2 emissions. The revisions will likely focus on more regional- and time-based matching requirements, limiting the ability to “replace” grid-supplied electricity with procured renewable energy on grids that are not where a company consumes power (e.g., through virtual PPAs (VPPAs), see Box 3.2-1) or with procured renewable energy that does not match the time profile of a company’s power consumption.[14]

Box 3.2-1

Market-based instruments for renewable power procurement

Corporate procurement of renewable electricity using market-based instruments is widespread, but the different types of instruments can lead to confusion. Power purchase agreements (PPAs) are the primary type of contract used in corporate procurement of renewable energy. PPAs can be either “physical” or “virtual” (also known as VPPAs).[15] Under physical PPAs, the buyer is located on the same grid as the supplier and takes ownership of the actual generated energy, “bundled” with its renewable attributes in the form of energy attribute certificates (EACs, see below). In a small number of cases, the buyer and seller are physically co-located, and the generated energy can be directly delivered behind the meter; this is known as a “direct PPA.” However, in most cases the buyer is not physically co-located with the seller’s power plant (despite being on the same grid). In these cases, the PPA is “sleeved,” with the generated energy supplied to the relevant utility, which then provides the equivalent amount of energy to the PPA buyer’s facilities. This may be supplied at different times of day from when the energy is physically generated, meaning that the utility must provide balancing services (for a “sleeving fee”).[16,17]

Because direct and physical PPAs require a buyer to find a renewable generator on the same grid, the possible supply is limited. To address this, many companies use VPPAs, which are purely financial transactions. Under a VPPA, the buyer guarantees a long-term fixed price for renewable energy generated by a power plant, and the seller sells the energy into its local grid (essentially a contract-for-differences). In return, the buyer takes title to the EACs. VPPAs allow buyers to secure EACs from renewable energy generation on grids that are far away from their physical location, greatly increasing the available supply of PPAs.[18] Direct and physical PPAs are usually regulated under different legal frameworks than VPPAs.[19] While VPPAs have dominated corporate renewable energy purchasing in the United States,[20] physical PPAs are more prevalent in Europe.[21]

Electricity EACs that are "bundled" with PPAs and VPPAs are denominated in MWh of generation. Specific examples of electricity EACs include renewable energy certificates (RECs)[22] in North America, Guarantees of Origin (GOs)[23] in Europe and Green Electricity Certificates (GECs) in Japan[24] and China.[25] Guidelines for EAC quality published by RE100 are used by many corporate procurers of renewable energy.[25] Many electricity EACs can also be procured "unbundled," which is essentially a direct purchase of the certificates themselves (from a renewable power generator or an intermediary broker) without any consideration of the underlying electricity. Typically, this is the lowest-cost way to obtain EACs, but it has the least direct and demonstrable linkage to accelerating deployment of renewable energy generation.

PPA buyers "retire" EACs to reduce their Scope 2 market-based emissions. This is done on a MWh-for-MWh basis, meaning that, in calculating emissions, each MWh of retired EACs replaces a MWh of electricity purchased from a grid, with the EAC typically carrying a zero (or sometimes near-zero) greenhouse gas emissions value. By procuring and retiring enough MWh of EACs to match or exceed the actual total procured MWh of electricity, companies can reduce their Scope 2 market-based emissions to zero.

While the use of PPAs has provided large amounts of capital to the renewable energy industry and has supported its rapid growth in many markets, it has also attracted criticism for enabling corporate emitters to "deem" emissions to be lower under the Scope 2 market-based method (by retiring the bundled EACs).[26,27] The use of unbundled EACs has attracted even more criticism.[28]

D. Mitigation Strategies for Scope 2 Location-Based Emissions

A variety of strategies can be used to reduce Scope 2 location-based emissions for data centers. In general, the strategies that will have the largest impact depend on the nature of the electric grid where the data center is located.

I. **Energy efficiency.** Ensuring maximum energy efficiency at both existing and new data centers by adopting the use of advanced cooling methods, increasing algorithmic efficiency, and implementing related techniques is highly impactful for reducing emissions on high-emissions grids. It also has numerous additional benefits even on relatively low-emissions grids (see Chapter 2 of this Roadmap). Energy efficiency reduces the total amount of high-emissions power generation required to serve a data center and minimizes additional impacts like land and water use for power production.

II. **Siting.** Physically siting new data centers in regions with low-emissions grids can have a large impact on ensuring Scope 2 emissions are low, given the large difference in average grid emissions intensity. However, this strategy does not apply to existing data centers that are located in high-emissions grid regions, and many planned data centers have other siting constraints that outweigh emissions considerations. Large new data centers may also require more additional power than low-emissions grids can supply.

III. **Load flexibility.** Implementing data center load flexibility by using on-site clean generation and storage and actively managing the timing of power consumption from the grid is highly impactful for reducing emissions on grids with large daily variation in their emissions intensity, such as those experiencing the "duck curve." Load flexibility can help ensure that data centers reduce their electricity demand during periods of peak grid demand, reducing the need to run the highest-emitting generators (typically gas-fired peaker plants). It can also be used to charge on-site storage during periods of high renewable generation, potentially avoiding the need to curtail solar and wind.[29,30] However, for grids that do not experience significant emissions variation throughout the day, load flexibility does not have a significant short-term impact on reducing emissions. In the longer term, load flexibility can reduce the amount of new generation capacity that a utility will build since utilities typically build new capacity to meet the projected future peak load. Particularly in cases when new gas-fired power plants are being planned for capacity addition, reducing the forecasted peak load can limit the construction of these plants and minimize potential "lock-in" of future emissions. (See Chapter 4 of this Roadmap.)

In addition to these strategies for individual data centers, utilities and grid operators can add low-carbon power generation to existing grids to serve new data center load. Adding power generation to existing grids ("expanding capacity") is primarily under the control of utilities and grid operators, who respond to forecasted power demand from a wide range of power users. However, in a growing number of high-profile cases, the projected additional power demand from a data center is so large that the data center operator (typically a hyperscaler) has become directly involved with the utility planning process and tariff design.[31,32]

Data center operators can also seek to secure dedicated (BTM) on-site/co-located power to supplement or entirely replace grid-supplied power, often for the purpose of moving more quickly than utility planning processes allow. This can be challenging for several reasons, including limited physical space available for on-site generation, backlogged supply chains for key equipment, and operational reliability challenges (see below).

Several factors influence which types of generation technology are optimal for a given grid region and data center, including the following. These factors guide which types of generation technologies utilities plan to add to serve new load from data centers and other end users and also influence potential selection of on-site/BTM generation options.

i. **Dispatchability.** Generators that can be fully controlled, or "dispatched," are relatively simple to integrate into an existing grid. Grid operators can directly ramp this generation up and down to match changes in power demand and to balance variable power supply from other generators. They can also incorporate this predictable generation into near-term power markets (e.g., day-ahead) to ensure adequate supply in advance. Dispatchable generation technologies include gas, coal, hydroelectric, geothermal and nuclear.

   Notably, while all of these generators can be dispatched, some can only adjust their generation output ("ramp") relatively slowly (particularly nuclear and coal), while others can ramp more quickly, such as hydroelectric and gas. Open-cycle gas turbines (OCGTs) are typically the fastest-ramping generators and are dispatched by grid operators to respond to rapid changes in load or generation. A related but separate issue is the "cold start" time for different generators, which is the time it takes to turn back on from a full shutdown. This can be many hours or longer in the case of coal and nuclear.[33]

Solar and wind ("intermittent" renewables) are non-dispatchable generation technologies because grid operators cannot always turn them on at a specified time. However, their generation is predictable, meaning that grid operators can use weather models to reliably forecast how much power these technologies will generate over the next several hours to days.[34] Increasingly, AI-based weather forecasting is improving these predictions and extending them farther into the future.[35] Using these forecasting methods, in combination with load forecasting methods, allows grid operators to predict how much additional generation will be needed in future hours and to plan accordingly. A growing number of grids are installing large-scale energy storage (typically batteries) to shift when the energy generated by intermittent renewables is delivered to load. This improves reliability and allows the grid to better match time profiles of energy usage.[36] Batteries can also be integrated with solar and wind generation facilities, "firming" the combined facility and making it dispatchable, within some constraints.[37,38] Unfortunately, some grid operators are facing challenges in approving variable renewables for interconnection because of the need to implement accelerated processes to study the impact of intermittent generation.[39,40]

ii. **Technology readiness.** While some generation technologies are fully mature and well-proven commercially, others are in the early stages of development and lack established performance records at scale. In general, only mature technologies can be integrated into commercial power grids at scale, while emerging ones require phased scale-up and testing (with some exceptions).[41] Notably, data center hyperscalers have been willing to employ early PPAs to support development of some emerging power generation technologies for connection to the grid to the grid.[42,43] In addition to providing small amounts of additional low-emissions power, this practice can have important benefits for accelerating these technologies into the market. (See Chapter 4 of this Roadmap.)

iii. **Location flexibility.** All power generation technologies have constraints on where they can be built and operated. Some of these constraints are driven by regulatory or commercial factors, while others depend on technology and infrastructure. Solar and wind rely on natural resources that are unevenly distributed globally, hydroelectric relies on water resources and appropriate terrain, and geothermal depends on appropriate subsurface characteristics. Natural gas relies on gas pipeline infrastructure (including availability of additional transmission capacity), and adding carbon capture and storage (CCS) to natural gas requires appropriate geological storage and/or CO_2 transport. Coal relies on coal availability and transport infrastructure, and nuclear relies on fuel supply chains and waste disposal. The physical footprint of each technology also influences location flexibility, particularly in the context of on-site/co-location strategies for which space may be limited. In general, gas, coal and nuclear have small footprints that are more compatible with constrained existing data center locations, while solar and wind require large land areas that may not be available at these sites.

iv. **Costs.** Cost is central to decisions about which type of power to procure for data centers. Many data center developers prioritize least-cost power options, but this is by no means universal. Some hyperscalers have shown a significant willingness to pay for low-carbon power, even if cheaper high–carbon emitting options are available (including willingness to support emerging and developing technologies as noted above). Cost is influenced by many factors, including the scale and maturity of relevant supply chains, labor expertise and availability, commodity prices and financing.[44]

V. **Low-carbon on-site power generation strategies.** On-site, BTM power generation can in theory provide some or all of the electricity requirements for data centers during normal (non-emergency) operations. This approach has been rare, due to a combination of high-power requirements, high costs and limited onsite space, but some data center developers are now seriously exploring this option at some sites.[45,46] On-site generation can reduce or potentially eliminate the need for grid-supplied power, although completely "off grid" data centers must fully manage power supply reliability without help from the primary electric grid, a major technical challenge. A key consideration for selecting on-site power generation technologies is the need for a generator (usually in combination with some form of energy storage) to provide continuous power over long durations (i.e., a high-capacity factor) to match high data center uptime rates.

In some locations, variable renewable energy (solar and wind) backed by storage systems (e.g., lithium-ion batteries) can provide this power, resulting in very low life-cycle greenhouse gas emissions.[47] Dispatchable renewable generation, particularly geothermal and hydroelectric (both of which require no or minimal additional energy storage) can also provide decarbonized power if sufficient resources are available at or near the data center site.[48] Enhanced geothermal systems (EGS) may be particularly suited for decarbonized data center power due to their greatly expanded geographic reach and their dispatchability,[49] recent examples include Google's agreement with Fervo[50] and Meta's agreements with Sage Geosystems[51] and XGS Energy.[52]

Gas-fired generation with CCS can provide low-carbon power at data center sites, with a relatively small physical footprint and high-capacity factor. The location constraints of this technology are driven primarily by access to CO_2 transport and storage rather than hydrological resources, subsurface heat or weather patterns.[53] Prominent examples of gas-fired generation with CCS include the announced[54] Crusoe/Tallgrass data center and the announced Frontier/Baker Hughes Sweetwater project,[55] both in Wyoming, United States.

Solid oxide fuel cells (SOFCs) are emerging as a source of on-site power generation.[56] These systems will mostly run on natural gas initially but are intended to switch to low-emissions hydrogen at a future date. SOFCs can be highly energy efficient in converting natural gas to electricity. They are particularly energy efficient when they are paired with district heat or other uses for their high-temperature waste heat, further reducing the intensity of greenhouse gas emissions compared to gas-fired generation.[57] However, the commercially deployed scale of SOFCs remains far lower than power generation systems based on gas-fired turbines, suggesting that scaling to multi-hundred-megawatt deployments may be a challenge.[58] SOFCs can be integrated with CCS to reduce emissions, although this approach has not been deployed commercially at scale.

Conventional nuclear power—both full-scale reactors and small modular reactors (SMRs)—is a potential source of decarbonized power for data centers with a compact footprint.[59] It has no inherent locational constraints, but it has substantial cooling requirements and relies on a nuclear fuel cycle supply chain that is highly limited in many relevant jurisdictions.[60] Recent high-profile examples, all of which are in the United States, include the planned restart of the 835-MW Unit 1 at the Three Mile Island nuclear power plant (scheduled for 2027)[61]; the planned construction of several SMRs, ranging from 50 MW to 320 MW in several locations (scheduled for 2030[62] and later[63]); and the planned construction of up to four GW-scale new AP1000 nuclear reactors (scheduled for 2032[64]).

There has also been significant interest in emerging nuclear fusion power technology for data centers, although this technology remains relatively early-stage and lacks an established supply chain for fuel and components.[65] A recent high-profile example is the planned construction of a 400-MW fusion power plant in the United States (scheduled for the early 2030s).[66]

Regardless of generation technology, BTM power requires careful integration with on-site backup power systems and energy storage, as well as potential coordination with grid supplied power, if the site maintains a grid tie. This places substantial reliability and coordination requirements on the BTM power generation.

E. Mitigation Strategies for Scope 2 Market-Based Emissions

All strategies for reducing Scope 2 location-based emissions are also relevant for reducing market-based emissions because the calculation of market-based emissions begins with the location-based total. Data center operators can further reduce market-based emissions by purchasing and retiring electricity EACs (see Box 3.2-1). The highest impact for this approach is by procuring renewable energy through PPAs bundled with EACs. As noted, the major hyperscalers have used this approach for well over a decade,[67] providing a revenue stream that supports many renewable energy projects around the world. The largest market remains North America,[12] but renewable PPAs are growing in Europe[68] and Asia.[69] This strategy will continue to be highly relevant, with hyperscalers expanding their renewable energy procurement in 2025.[70]

However, the growing use of electricity EACs to reduce market-based emissions has been controversial, with criticisms falling into two major categories. The first is based on the observation that the renewable energy represented by an EAC may not match the time profile of the electricity consumed by the buyer, potentially exacerbating grid balancing challenges. Similarly, if EACs come from renewable generation that is not on the same grid as the buyer (or is from a portion of the grid that is transmission-constrained), it may be unreasonable to claim this electricity is “consumed” by the buyer’s facilities. These concerns have led to proposals such as the “three pillars” of time-matching, additionality[71] and location/regional constraints and “24/7 Carbon Free Energy.”[72] In general, these principles would decrease the ability of buyers to use VPPAs and time-agnostic EACs in reducing Scope 2 market-based emissions.

The second broad category of criticism for using electricity EACs to reduce market-based emissions is based on the observation that renewable energy generation has dramatically different impacts on the marginal greenhouse gas emissions of grids in different locations. For grids with fossil fuel (such as open-cycle gas turbines) on the margin, additional renewable generation avoids a large amount of greenhouse gas emissions. However, for grids with renewables, such as hydroelectric, curtailed solar or wind, or even battery storage on the margin, additional renewable generation does not avoid any greenhouse gas emissions. This has led to proposals focused on marginal emissions, which would decrease the ability of buyers to use EACs from low-marginal-emissions grids to reduce Scope 2 market-based emissions.[73,74]

As the Greenhouse Gas Protocol Scope 2 guidance undergoes revision during 2025, the final direction of the Scope 2 market-based emissions method remains unclear. However, the revision will likely have a significant impact on the strategies used by data center operators for renewable energy procurement to reduce market-based emissions.

F. Recommendations

- *Data center operators should **maximize energy efficiency**, including using advanced cooling and other highly efficient equipment and implementing algorithmic efficiency whenever possible.*

- *Data center operators should consider implementing **load flexibility and on-site storage**, particularly for grids with highly variable emissions intensity.*

- *Data center operators should **include grid carbon intensity as a key siting consideration** and seek to site data centers in the lowest-emitting grid regions as much as possible.*

- *Data center operators and utilities should work together to **identify the optimal mix of new low-carbon power generation technologies to add to the grid to meet rising data center load**. This should include consideration of data center load flexibility when determining the amount of new generation required.*

- *Data center operators considering on-site/BTM power generation solutions should seek to **minimize emissions when selecting generation technologies**.*

- *Data center operators should continue to **support emerging/developing low-carbon power generation technology**.*

- *In addition to the above strategies, data center operators should continue to **procure renewable energy through PPAs**. They should also anticipate potential changes to the Greenhouse Gas Protocol Scope 2 guidance and plan accordingly when determining the necessary amount and type of procurement.*

- *Grid operators should work closely with data centers to **understand the appropriate amount of new capacity to add to meet rising load and should seek to maximize low-carbon generation technologies for new capacity additions as much as possible**.*

- *Grid operators should continue to **reform and accelerate the interconnection process for intermittent renewable generation** in order to provide new low-carbon capacity to meet data center and other demand.*

G. References

1. Mary Sotos. GHG Protocol Scope 2 Guidance: An amendment to the GHG Protocol Corporate Standard; World Resources Institute (WRI) Greenhouse Gas Protocol, Washington, DC, https://ghgprotocol.org/scope-2-guidance (2023).

2. Unique Karki, Vedant Sinha, Muhammad Ali Qamar, Ovais Khan, John Kissock & Prakash Rao. Understanding Scope 2 Emissions – Tipsheet 8; Lawrence Berkeley National Laboratory (LBNL), US Department of Energy (DOE) Office of Manufacturing and Energy Supply Chains (MESC), DOE Industrial Assessment Center, and University of California (UC) Davis, Berkeley, California, https://industrialapplications.lbl.gov/tip-sheet-understanding-scope-2 (Accessed August 2025).

3. European Union (EU) European Environment Agency (EEA). Greenhouse gas emission intensity of electricity generation in Europe; EEA, Copenhagen, Denmark, https://www.eea.europa.eu/en/analysis/indicators/greenhouse-gas-emission-intensity-of-1 (2025).

4. Malgorzata Wiatros-Motyka, Dave Jones, Hannah Broadbent, Nicolas Fulghum, Chelsea Bruce-Lockhart, Reynaldo Dizon, Phil MacDonald, Charles Moore, Alison Candlin, Uni Lee, Libby Copsey, Sam Hawkins, Matt Ewen et al. Chapter 6: Country and Region Deep Dives in Global Electricity Review 2023 (Ember, London, United Kingdom, 2023, https://ember-energy.org/latest-insights/global-electricity-review-2023/country-and-region-deep-dives/).

5. Euan Graham, Nicolas Fulghum & Katye Altieri. Global Electricity Review 2025; Ember, London, United Kingdom, https://ember-energy.org/app/uploads/2025/04/Report-Global-Electricity-Review-2025.pdf (2025).

6. International Energy Agency (IEA). Executive summary in Electricity 2024 (IEA, Paris, France, 2024, https://www.iea.org/reports/electricity-2024/executive-summary).

7. International Energy Agency (IEA). Energy demand from AI in Energy and AI (IEA, Paris, France, 2025, https://www.iea.org/reports/energy-and-ai/energy-demand-from-ai).

8. Sean Wolfe. With rapid solar additions, the ‘duck curve’ begins to emerge in Texas; Factor This Power Engineering, Shelton, Connecticut, https://www.power-eng.com/renewables/solar-energy/with-rapid-solar-additions-the-duck-curve-begins-to-emerge-in-texas/ (2024).

9. Asma Aziz. Big batteries are solving a longstanding problem with solar power in California. Can they do the same for Australia?; The Conversation UN, Inc., Waltham, Massachusetts, https://theconversation.com/big-batteries-are-solving-a-longstanding-problem-with-solar-power-in-california-can-they-do-the-same-for-australia-231063 (2024).

10. Neshwin Rodrigues & Aditya Lolla. Solar adoption in India entering “accelerating growth” phase; Ember, London, United Kingdom, https://ember-energy.org/app/uploads/2024/09/India-solar-uptake.pdf (2024).

11. Akshaya Jagannadharao, Nicole Beckage, Dawn Nafus & Scott Chamberlin. Timeshifting strategies for carbon-efficient long-running large language model training. Innovations in Systems and Software Engineering 21, 517-531 (2025). https://doi.org/10.1007/s11334-023-00546-x.

12. Tony Lenoir. Nuclear bolsters top US hyperscalers' clean energy portfolio, now over 84 GW; S&P Global, New York, New York, https://www.spglobal.com/market-intelligence/en/news-insights/research/nuclear-bolsters-top-us-hyperscalers-clean-energy-portfolio-now-over-84-gw (2025).

13. World Resources Institute (WRI) and World Business Council for Sustainable Development (WBCSD). Standards Development and Governance Repository; WRI and WBCSD, Washington, DC and Geneva, Switzerland, https://ghgprotocol.org/standards-development-and-governance-repository (Accessed August 2025).

14. Sarah Huckins. Scope 2 Standard Advances: ISB Approves Consultation on Market- and Location-Based Revisions; Signals Cross-Sectoral Work Ahead on Avoided Emissions; World Resources Institute (WRI) and World Business Council for Sustainable Development (WBCSD), Washington, DC and Geneva, Switzerland, https://ghgprotocol.org/blog/scope-2-standard-advances-isb-approves-consultation-market-and-location-based-revisions (2025).

15. Rachit Kansal. Introduction to the Virtual Power Purchase Agreement; Rocky Mountain Institute (RMI) Business Renewables Center (BRC), Basalt, Colorado, https://rmi.org/insight/virtual-power-purchase-agreement/ (2018).

16. Chris Bowden. Understanding the evolution of the PPA market; Squeaky Clean Energy Group Ltd., London, United Kingdom, https://www.squeaky.energy/blog/understanding-the-evolution-of-the-ppa-market (2023).

17. Plàcido Ostos. The Role of Balancing in Virtual and Physical PPAs; LevelTen Energy, Seattle, Washington, https://www.leveltenenergy.com/post/balancing-in-vppas (2024).

18. Rob Collier. 4 Questions to Ask Before Choosing a Physical or Virtual Power Purchase Agreement; LevelTen Energy, Seattle, Washington, https://www.leveltenenergy.com/post/physical-power-purchase-agreement-or-virtual-ppa (2019).

19. Jordan Farrell. ESG, VPPAs and Dodd-Frank; Husch Blackwell LLP, Milwaukee, Wisconsin, https://www.climatesolutionslaw.com/2023/06/esg-and-renewable-energy/ (2023).

20. Benjamin Grayson. Corporate VPPAs: Risks and sensitivities; Norton Rose Fulbright US LLP, Houston, Texas, https://www.projectfinance.law/publications/2020/june/corporate-vppas-risks-and-sensitivities/ (2020).

21. Nicolas Briet. Virtual PPAs: State of Play in 2025; Pexapark, Zurich, Switzerland, https://pexapark.com/blog/prmc-virtual-ppas-state-of-play-in-2025/ (2025).

22. US Environmental Protection Agency (EPA). Renewable Energy Certificates (RECs); EPA, Washington, DC, https://www.epa.gov/green-power-markets/renewable-energy-certificates-recs (Accessed August 2025).

23. Association of Issuing Bodies (AIB). Renewable Energy Guarantees of Origin; AIB, Brussels, Belgium, https://www.aib-net.org/certification/certificates-supported/renewable-energy-guarantees-origin (Accessed August 2025).

24. Masaya Ishida. Electricity Certificate for Renewables: Comparison of Japanese and International Systems (English Edition); Renewable Energy Institute, Tokyo, Japan, https://www.renewable-ei.org/pdfdownload/activities/REI_RE-Certificates_EN.pdf (2022).

25. Daniel Kramb. China's energy market 'open for business' after support from RE100 on certificates; The Climate Group, London, United Kingdom, https://www.theclimategroup.org/our-work/press/chinas-energy-market-open-business-after-support-re100-certificates (2025).

26. Tina Freese. Virtual PPA Framework a Good First Step, but Regulatory Backing is Essential; Mercom Capital Group, LLC, Austin, Texas, https://www.mercomindia.com/vppas-in-india-not-gained-momentum-cerc-guidelines (2025).

27. Simone Accornero. Corporate renewable procurement: greenwashing or key driver of the EU's clean industrial deal?; European Climate, Infrastructure and Environment Executive Agency (CINEA), Brussels, Belgium, https://sustainable-energy-week.ec.europa.eu/news/corporate-renewable-procurement-greenwashing-or-key-driver-eus-clean-industrial-deal-2025-02-06_en (2025).

28. Anders Bjørn, Shannon M. Lloyd, Matthew Brander & H. Damon Matthews. Renewable energy certificates threaten the integrity of corporate science-based targets. Nature Climate Change 12, 539-546 (2022). https://doi.org/10.1038/s41558-022-01379-5.

29. John Fitzgerald Weaver. California solar curtailment down 12% on back of batteries; PV Magazine USA, Providence, Rhode Island, https://pv-magazine-usa.com/2025/07/22/california-solar-curtailment-down-12-on-back-of-batteries/ (2025).

30. Lucia Colaluce. Solar curtailment surges by 97% in Germany in 2024 despite lower operational costs; Strategic Energy Europe (Strategic Energy Corporation Business Unit), Bogota, Colombia, https://strategicenergy.eu/solar-curtailment-germany/ (2025).

31. Matthew Gooding. Google buys 115MW of geothermal energy to power Nevada data centers; Data Center Dynamics (DCD), London, United Kingdom, https://www.datacenterdynamics.com/en/news/google-buys-115mw-of-geothermal-energy-to-power-nevada-data-centers/ (2024).

32. Briana Kobor. Our first-of-its-kind partnership for clean energy has been approved in Nevada; Google Blog, Mountain View, California, https://blog.google/feed/nevada-clean-energy/ (2025).

33. Owen Comstock. Today in Energy: About 25% of U.S. power plants can start up within an hour; US Energy Information Administration (EIA), Washington, DC, https://www.eia.gov/todayinenergy/detail.php?id=45956 (2020).

34. Rui Yang. Solar and Wind Forecasting; National Renewable Energy Laboratory (NREL), Golden, Colorado, https://www.nrel.gov/grid/solar-wind-forecasting (Accessed August 2025).

35. Alessio Verdone, Massimo Panella, Enrico De Santis & Antonello Rizzi. A review of solar and wind energy forecasting: From single-site to multi-site paradigm. Applied Energy 392, 126016 (2025). https://doi.org/10.1016/j.apenergy.2025.126016.

36. Arcelia Martin. More Solar and Battery Storage Were Added to Texas' Grid Than Any Other Power Source Last Year; Inside Climate News, New York, New York, https://insideclimatenews.org/news/10022025/solar-battery-storage-texas-grid/ (2025).

37. Rob Clow. Australian firmed solar overtakes wind projects; Ion Analytics, New York, New York, https://ionanalytics.com/insights/infralogic/australian-firmed-solar-overtakes-wind-projects/ (2025).

38. Kostantsa Rangelova & Dave Jones. Solar electricity every hour of every day is here and it changes everything; Ember, London, United Kingdom, https://ember-energy.org/latest-insights/solar-electricity-every-hour-of-every-day-is-here-and-it-changes-everything/ (2025).

39. Joseph Rand, Nick Manderlink, Will Gorman, Ryan Wiser, Joachim Seel, Julie Mulvaney Kemp, Seongeun Jeong & Fritz Kahrl. Queued Up: Characteristics of Power Plants Seeking Transmission Interconnection; Lawrence Berkeley National Laboratory (LBNL) Energy Technologies Area: Energy Markets & Policy, Berkeley, California, https://emp.lbl.gov/queues (2024).

40. Kaitlin Fung. 5 key questions about US grid interconnection answered; Wood Mackenzie, Edinburgh, United Kingdom, https://www.woodmac.com/news/opinion/5-key-questions-about-us-grid-interconnection-answered/ (2025).

41. Ally Copple. Fervo Energy Announces 320 MW Power Purchase Agreements with Southern California Edison; Fervo Energy, Houston, Texas, https://fervoenergy.com/fervo-energy-announces-320-mw-power-purchase-agreements-with-southern-california-edison/ (2024).

42. Helion Energy. Helion announces world's first fusion energy purchase agreement with Microsoft; Everett, Washington, https://www.helionenergy.com/articles/helion-announces-worlds-first-fusion-ppa-with-microsoft/ (2023).

43. Nuclear Newswire. Google announces power purchase agreement with Commonwealth Fusion; American Nuclear Society, Westmont, Illinois, https://www.ans.org/news/2025-07-02/article-7164/google-announces-power-purchase-agreement-with-commonwealth-fusion/ (2025).

44. Sascha Samadi. A Review of Factors Influencing the Cost Development of Electricity Generation Technologies. Energies 9, 970 (2016). https://doi.org/10.3390/en9110970.

45. Crusoe and Tallgrass. Crusoe and Tallgrass announce AI data center in Wyoming; Crusoe Energy Systems LLC, Denver, Colorado, https://www.crusoe.ai/resources/newsroom/crusoe-and-tallgrass-announce-ai-data-center-in-wyoming (2025).

46. Dylan Baddour & Arcelia Martin. Data centers are building their own gas power plants in Texas; The Texas Tribune, Austin, Texas, https://www.texastribune.org/2025/06/05/texas-data-centers-gas-power-plants-ai/ (2025).

47. DataBank Blog. Powering Data Centers With Renewable Energy For A Sustainable Future; DataBank Holding, Ltd., Dallas, Texas, https://www.databank.com/resources/blogs/powering-data-centers-with-renewable-energy-for-a-sustainable-future/ (2024).

48. Andrew Webber. How Hydro and Data Centers Are Pairing to Create Unique Value; National Hydropower Association, Washington, DC, https://www.hydro.org/powerhouse/article/how-hydro-and-data-centers-are-pairing-to-create-unique-value/ (2025).

49. Ben King, Wilson Ricks, Nathan Pastorek & John Larsen. The Potential for Geothermal Energy to Meet Growing Data Center Electricity Demand; Rhodium Group, LLC, New York, New York, https://rhg.com/research/geothermal-data-center-electricity-demand/ (2025).

50. Emma Penrod. NV Energy seeks new tariff to supply Google with 24/7 power from Fervo geothermal plant; Utility Dive (Informa TechTarget), Newton, Massachusetts, https://www.utilitydive.com/news/google-fervo-nv-energy-nevada-puc-clean-energy-tariff/719472/ (2024).

51. Dan McCarthy. Sage Geosystems and Meta sign 150MW geothermal power agreement; Canary Media, Asheville, North Carolina, https://www.canarymedia.com/articles/geothermal/sage-geosystems-and-meta-sign-150mw-geothermal-power-agreement (2024).

52. Lamar Johnson. Meta signs geothermal power deal for New Mexico data centers; Utility Dive (Informa TechTarget), Newton, Massachusetts, https://www.utilitydive.com/news/meta-xgs-energy-announce-geothermal-deal-new-mexico-data-centers-ai/750913/ (2025).

53. Nicki Stuckert, Patti Smith, Daniel Garcia & Jordan Alford. Meeting Data Center Electricity Demand: Mapping carbon capture potential for natural gas-fired generators in the US and Canada; Carbon Direct Inc., New York, New York, https://www.carbon-direct.com/research-and-reports/meeting-data-center-electricity-demand (2025).

54. Violet George. Wyoming Lands 1.8-Gigawatt AI Data Center Powered By Gas And Carbon Capture; Carbon Herald Inc., New York, New York, https://carbonherald.com/wyoming-lands-1-8-gigawatt-ai-data-center-powered-by-gas-and-carbon-capture/ (2025).

55. Baker Hughes. Baker Hughes, Frontier Infrastructure Announce Partnership to Accelerate Development of Carbon Capture and Storage, Data Center Projects in the U.S. (Press Release); Houston, Texas, https://investors.bakerhughes.com/news-releases/news-release-details/baker-hughes-frontier-infrastructure-announce-partnership

56. Bloom Energy. Bloom Energy Announces World's Largest Fuel Cell Installation in History; San Jose, California, https://www.bloomenergy.com/news/bloom-energy-announces-worlds-largest-fuel-cell-installation-in-history/ (2024).

57. Xiaotian Zhang, Zuwen Liu, Peng Qiu, Peixiang Wang, Kangwei Wei & Zhiguo Guo. Progress in methane-based solid oxide fuel cells: Challenges for clean and efficient utilization from natural gas. Fuel 395, 135211 (2025). https://doi.org/10.1016/j.fuel.2025.135211.

58. Erhan Eren. Fuel Cell Installations in Data Centers: Top 10 Projects & Companies; EnkiAI, San Francisco, California, https://enkiai.com/fuel-cell-installations-in-data-centers-top-10-projects-companies (2025).

59. US Department of Energy (DOE) Office of Nuclear Energy. Advantages and Challenges of Nuclear-Powered Data Centers (Blog); Washington, DC, https://www.energy.gov/ne/articles/advantages-and-challenges-nuclear-powered-data-centers (2025).

60. Jane Accomando. Nuclear-Powered Data Centers—What U.S. Developers Need to Know; Power Magazine, Rockville, Maryland, https://www.powermag.com/nuclear-powered-data-centers-what-u-s-developers-need-to-know/ (2025).

61. Amber Jackson. Microsoft & Constellation's Bid to Restart Three Mile Island; Data Centre Magazine, Norwich, Norfolk, United Kingdom, https://datacentremagazine.com/critical-environments/microsoft-constellation-restarting-a-nuclear-reactor (2025).

62. Amanda Peterson Corio. Our first advanced nuclear reactor project with Kairos Power and Tennessee Valley Authority (Blog); Google, Mountain View, California, https://blog.google/outreach-initiatives/sustainability/google-first-advanced-nuclear-reactor-project-with-kairos-power-and-tennessee-valley-authority/ (2025).

63. Amazon Staff. Amazon signs agreements for innovative nuclear energy projects to address growing energy demands; Amazon.com, Inc., Seattle, Washington, https://www.aboutamazon.com/news/sustainability/amazon-nuclear-small-modular-reactor-net-carbon-zero (2024).

64. World Nuclear News. Fermi America, Hyundai E&C team up for Texan reactors; World Nuclear Association, London, United Kingdom, https://www.world-nuclear-news.org/articles/fermi-america-hyundai-ec-team-up-for-texan-reactors (2025).

65. Digitalisation World. Fusion-powered data centres: Google purchases 200MW of fusion power from US$2 billion startup; Digitalisation World, Coventry, United Kingdom, https://digitalisationworld.com/news/70232/fusion-powered-data-centres-google-purchases-200mw-of-fusion-power-from-us-2-billion-startup (2025).

66. Sebastian Moss. Commonwealth Fusion Systems raises more than $1bn, backed by unnamed hyperscale data center developer - report; Data Center Dynamics (DCD), London, United Kingdom, https://www.datacenterdynamics.com/en/news/commonwealth-fusion-systems-raises-more-than-1bn-backed-by-unnamed-hyperscale-data-center-developer-report/ (2025).

67. Julia Komitova. Google Energy sings first wind farm PPA; Renewables Now, Sofia, Bulgaria, https://renewablesnow.com/news/google-energy-sings-first-wind-farm-ppa-84105/ (2010).

68. Mark Thomton, Hla Myat Mon, Chris Boba & Angélica Juárez. European renewable PPA market sees 19GW of new capacity contracted in 2024; Wood Mackenzie, Edinburgh, United Kingdom, https://www.woodmac.com/press-releases/european-renewable-ppa-market-sees-19gw-of-new-capacity-contracted-in-2024/ (2025).

69. You-Jie Cai. A Deep Dive into PPAs and VPPAs in East Asia; Apala Group, San Francisco, California, https://www.apalagroup.com/angles/wudgp1hcm054ccvow9mhu3waen8yl1 (2025).

70. Tony Lenoir. Corporate PPA leaderboard – Microsoft leap cuts into Amazon lead; S&P Global, New York, New York, https://www.spglobal.com/market-intelligence/en/news-insights/research/corporate-ppa-leaderboard-microsoft-leap-cuts-into-amazon-lead (2025).

71. American Clean Power. APC Framework: 3 Pillars for Building a Green Hydrogen Industry for Decarbonization; American Clean Power, Washington, DC, https://cleanpower.org/gateway.php?file=2023/06/ACP_GreenHydrogenFramework_OnePager.pdf (2023).

72. Google. 24/7 Carbon-Free Energy: Methodologies and Metrics; Mountain View, California, https://sustainability.google/reports/24x7-carbon-free-energy-methodologies-metrics/ (2021).

73. Emissions First Partnership (EFP). First Things First: Accelerating Grid Decarbonization (Homepage); https://www.emissionsfirst.com/ (Accessed August 2025).

74. Sam Koebrich, Joel Cofield, Gavin McCormick, Ishan Saraswat, Nat Steinsultz & Pierre Christian. Towards objective evaluation of the accuracy of marginal emissions factors. Renewable and Sustainable Energy Reviews 215, 115508 (2025). https://doi.org/10.1016/j.rser.2025.115508.

3.3 Embodied Greenhouse Gas Emissions (Scope 3)

Julio Friedmann and Colin McCormick

The embodied emissions of data centers can be significant, chiefly due to upstream emissions from manufacturing steel, cement, concrete and computer hardware. (These are often referred to as Scope 3 emissions.[1]) In cases where a data center consumes only very low-carbon power (e.g., from renewables or nuclear), embodied emissions can exceed 40% of a data center's total greenhouse gas emissions and may dominate the lifetime greenhouse gas emissions footprint.[2,3] For example, despite significant investments in renewable power and carbon dioxide (CO_2) removal since 2021,[4] Microsoft's corporate net greenhouse gas footprint increased in 2023 and 2024. This increase was due chiefly to Scope 3 (embodied) emissions from data center production and building, not Scope 2 emissions from power supplies.[5]

The growing size and complexity of data centers adds to the total carbon footprint of new sites. Individual sites can be many hectares and even more than 1 square mile (2.6 km^2) in some recent examples. In addition, chip production can contribute significant greenhouse gases to a facility's footprint due to fossil electricity and use of fluorinated gases (F-gases), such as sulfur hexafluoride (SF_6) and carbon tetrafluoride (CF_4), which are very potent greenhouse gases.

Unfortunately, there are few options to purchase low-carbon goods today. Supplies of low-carbon cement, concrete, steel and chips cost significantly more than conventional supplies. More importantly, total volumes of these goods are very small, making purchasing difficult. Production sites are often geographically distant from data center construction, adding cost and carbon to direct use.

Decision makers must better understand the scale of the embodied emissions from data centers and the challenges to abatement. *They should have as much familiarity with embodied and Scope 3 contributions as with power usage effectiveness (PUE).* Practices such as extending server life cycles can reduce total embodied carbon and should be considered by operators. Opportunities for innovation in technology, policy and commercial models exist but will require sustained investment and commitment to achieve important climate and economic outcomes.

A. Primary Sources of Embodied Emissions (Scope 3)

The main infrastructure of a data center includes the core and shell,[6] pipes, wiring, cooling systems, IT hardware and ancillary systems, such as back-up power, batteries and low-voltage switchgear. (See Chapter 1 of this Roadmap.) All these contribute to Scope 3 emissions.

- The core is the space that houses pipes, wiring, cooling systems and related equipment. The core comprises load bearing walls, elevator shafts, pilings and interior foundations, including columns, beams, slabs and walls.
- The shell includes the exterior elements of the facility, including structural foundation, roofing, exterior walls, waterproofing and parking.
- Information technology (IT) hardware includes servers, power supplies, networking equipment and data storage/memory equipment.

Cement and concrete, steel, chips and other IT equipment release the largest greenhouse gas emissions associated with data center construction. Concrete and steel constitute the largest Scope 3 component of the core and shell, with different amounts and kinds of concrete and steel selected to match the facility design needs.[7]

Figure 3.3-1. Embodied emissions of a modern data center (year-1 build)
Selected categories only (per MW basis)

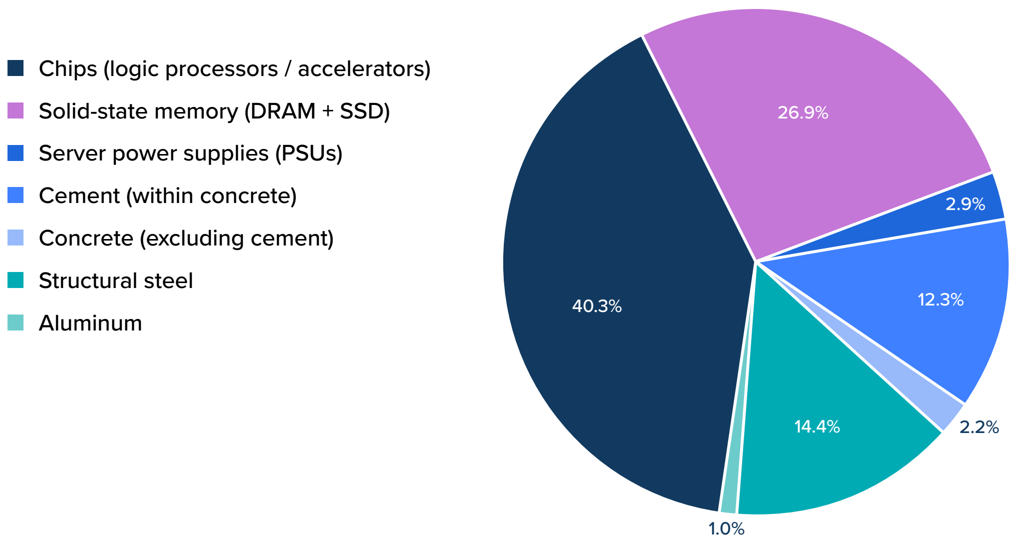


Table 3.3-1. Embodied CO_2e by Category (Per MW and Scaled to Annual New Capacity)

Category	tCO_2e per MW - year 1 build	$MtCO_2e$ per year at 20 GW (mid-case)	$MtCO_2e$ per year at 18 GW (low-case)	$MtCO_2e$ per year at 27 GW (high-case)
Chips (logic processors/ accelerators)	768.2	15.36	13.83	20.74
Solid-state memory (DRAM + SSD)	512.1	10.24	9.22	13.83
Server power supplies (PSUs)	54.9	1.1	0.99	1.48
Cement (within concrete)	235.0	4.7	4.23	6.34
Concrete (excluding cement)	41.5	0.83	0.75	1.12
Structural steel	274.9	5.5	4.95	7.42
Aluminum	19.2	0.38	0.35	0.52

Figure 3.3-1 and Table 3.3-1. Source: Schneider Electric (2023)[2], Schneider et al. (2025)[8]

Box 3.3-1

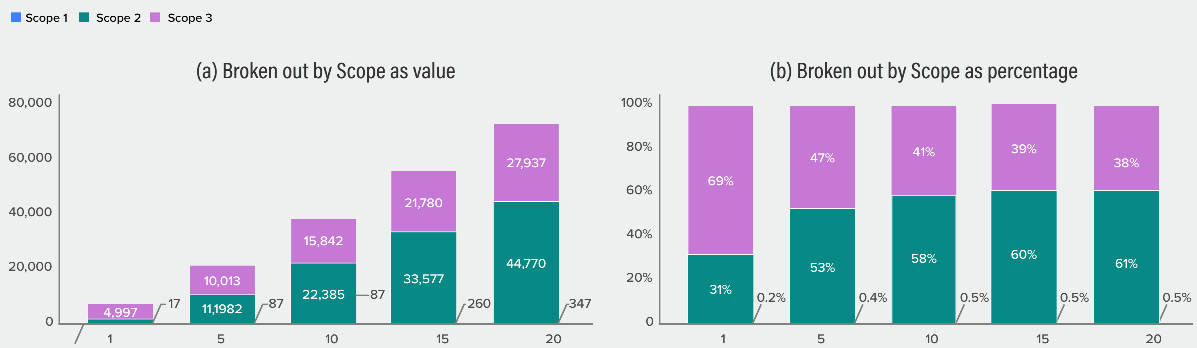
A 200 MW reference data center

To anchor the analysis of this chapter, we considered a 200 MW data center and the materials required to make it. The physical footprints of 200 MW data centers vary between roughly 60,000 m^2 (~650,000 ft^2) and 120,000 m^2 (~1.3 million ft^2), depending on energy density, geography, design, computational function and other factors.[9,10]

- We assume 93,000 m^2 (1 million ft^2) for a 200 MW data center.
- Some components of a data center are replaced or upgraded during its life, especially racks and computer hardware, while other components, such as core and shell, do not change. We use a 15 year life-cycle estimate, a typical number for a data center's life,[2] which will include some IT hardware replacement.
- A 200 MW data center can require a wide range of concrete volumes as a function of building code, seismic requirements and design. Estimates range between 55,000 and 500,000 m^3. We assume 300,000 m^3 of concrete for a facility of this size.
- Robust, validated estimates for use of structural steel and rebar in data centers are scarce. Assuming 50-75 kg/ft^2 of floor (535-802 kg/m^2), total steel would be 50-75 million kg for a 200 MW data center.
- The broad category of IT equipment comprises a wide variety of components, including servers, networking equipment, storage, and racks and enclosures. Some hyperscalers have developed detailed models for the embodied emissions of these systems, but the models are not comprehensively available. Rough estimates suggest that these systems collectively have embodied carbon[11] of 750-1500 tons of CO_2 equivalent (tCO_2e)/MW.

Accurate data on physical materials going into modern data centers are scarce. Decision makers in industry and governments should prioritize making these data available to the public and ensuring their quality.

Figure 3.3-2. Scope 1, 2 and 3 footprints for a representative data center at different years in its operation. Source: Schneider Electric (2023).[2]



i. Cement and concrete

Cement production generates 1.6 Gt/year of CO_2 emissions, or roughly 6% of global annual greenhouse gas flux.[12] Approximately 50% is from fuel used to generate high-temperature heat and 50% is due to by-product emissions from production chemistry.[13] Concrete is a mixture of cement, aggregate (sand and gravel) and water. Although cement represents roughly 15% of concrete by weight, it contributes the overwhelming majority (roughly 88%) of concrete's greenhouse gas footprint.[14] Typical concrete emissions for 1 m^3 of concrete are 410 kg/m^3 but can range between 290 and 610 kg/m^3 depending on cement content (e.g., substitution of pozzolanic materials), energy input mix (coal, gas, biomass, used tires), water footprint, transportation and similar factors.[15]

A data center requires significant volumes of cement and concrete, which contributes significantly to its embodied emissions. Depending on setting and design, the embodied emissions from concrete for a 200 MW data center would equal 123,000 tCO_2e if built using 300,000 m^3 concrete (87,000-183,000 tons).

Estimates for total concrete used in data center construction globally in 2025 vary from 1.3 million to 4.5 million m^3. This equates to roughly 0.5-1.8 million tCO_2, with a median estimate of 1.15 million tCO_2 each year. For comparison, a single 200 MW natural gas plant would emit nearly 0.5 million tCO_2 each year, not including potential upstream emissions, which could significantly add to the total footprint. Said differently, the total annual emissions from concrete added by all new data centers globally could equal the annual emissions from 1-3 natural gas power plants.

ii. Steel

Steel production worldwide releases 3.6 Gt CO_2/year,[16] or roughly 8% of annual CO_2 emissions and 6.5% of greenhouse gas emissions. Each production pathway has different emissions profiles and different pathways to abatement.[17] On average, iron and steel production generates ~1.9 kg of CO_2 per kg of steel.[16] Given that, the embodied emissions of a 200 MW data center could range between 95,000 and 140,000 tonnes of CO_2.

- Roughly 75% of steel globally is produced through blast-furnace/basic oxygen furnace (BF/BOF) facilities. Like with cement, significant portions of these emissions are chemical production by-products from these operations.

- Electric arc furnaces (EAF) recycle scrap steel but cannot produce primary iron or steel. Roughly 20% of global supply comes from EACs. Their footprint is significantly lower than for BF/BOF systems, although the total embodied emissions will vary significantly based on the source of electricity (coal-fired, gas-fired, nuclear or renewable source).

- Roughly 5% comes from direct reduction of iron (DRI) facilities that use coal or natural gas for both heat and chemical reduction in tandem with an EAF to process the sponge iron made by the DRI. A very small volume (<1%) comes from DRI-EAF systems that run on low-carbon hydrogen.

Public data on total steel used in data center construction are not available, making estimation difficult. One estimate[18] asserts that data centers require 150-200 kg/m^2 (30-40 lbs/ft^2) but without additional documentation. Even if correct, the estimated square meterage of data centers built or projected is poorly known, preventing easy estimation. Using a value of 275 tons of steel/MW (from Schneider Electric[2]; see Table 3.3-1), a 2.4 carbon intensity for average steel (tonnes CO_2/tonne steel), and estimates of recent and projected builds, Table 3.3-2 estimates steel use in data center builds and associated Scope 3 emissions. This may be an under-estimate since it does not include emerging markets.

Given the anticipated growth of data centers discussed in Chapter 1 of this Roadmap, annual Scope 3 emissions from steel alone should grow to more than 10 Mt/year soon, most likely before 2030.

Ultimately, this estimate underscores the need for better data and greater attention on embodied emissions. It also suggests that gathering and sharing these data may prove complicated for governments, requiring greater transparency by builders and owners of data centers.

Table 3.3-2. Estimated data center construction and associated Scope 3 emissions in 2024.

	Estimated MW of construction in 2024	Estimated Scope 3 emissions from steel (tonnes CO_2) in 2024
United States	6900 MW[19]	4,500,000
Asia Pacific	1600 MW[20]	1,100,000
Europe-ME	750 MW[21,22]	500,000
Total	9250 MW	6,100,000

iii. Semiconductors, chips and other information technology (IT) hardware

Production of servers, power supplies, networking equipment and data storage/memory equipment generates greenhouse gas emissions in two ways. First, the electricity generated to power manufacturing equipment leads to CO_2 emissions, depending on the emissions intensity of the generation source. The electricity (Scope 2) emissions of the global semiconductor industry were approximately 44 million tCO_2e in 2021.[23] (Other industries, such as electronics-component manufacturing, also consume electricity to produce IT equipment for data centers.) Second, semiconductor fabrication involves use of F-gases, such as SF_6 and CF_4, which can be vented to the atmosphere under some circumstances. While F-gas emissions intensity from the semiconductor industry has fallen, total emissions have continued to rise, exceeding 15 million tCO_2e in 2020.[24] Only a portion of these electricity-related and F-gas-related emissions are directly attributable to semiconductors produced for data centers because semiconductors are also used in a wide variety of other sectors. Comprehensive, recent data on these emissions are not readily available.

Additional equipment at data centers, such as gas-insulated electrical switchgear, can contain F-gases (specifically SF_6), which can potentially escape to the atmosphere during maintenance, fault events and end-of-life disposal. Estimates of current emissions rates from this type of equipment at data centers are not available, although the use of gas-insulated switchgear at data centers is growing.[25]

B. Technology Options for Low-Carbon Data Center Supply Chains and Construction

Builders and operators of data centers could significantly reduce their emissions with existing technologies. Although it is not possible to achieve zero Scope 3 emissions today, advanced technologies could potentially reduce these emissions dramatically. Innovative approaches, both in design and within supply chains, have the potential to reduce embodied emissions if developed, applied and purchased.

i. Material substitution

If approximately 30% of Scope 3 greenhouse gas emissions come from steel, cement and concrete, reducing the total amount of these materials could reduce embodied emissions in data centers. This strategy chiefly involves substituting materials that use less of the emitting constituents.

When making concrete, it is possible to substitute a portion of clinker with other materials that perform the same function (called pozzolinic materials or supplementary cementitious materials (SCMs)). Fly ash, steel slag and silica fume are examples of man-made SCMs, and volcanic ash, metakaolin and diatomaceous earth are natural examples.[26] In particular, fly ash, a byproduct of coal combustion, is used to some degree in 60% of US Portland Cement and can replace 15-40% of the clinker based on type of fly ash and concrete performance requirements.[27] In addition, new concrete formulations can use advanced cements, such as pozzolins, binding additional CO_2 into the concrete matrix.

Another strategy is to completely substitute one material for another. For example, some structural reinforced concrete can be replaced with novel wood products, including cross-laminated timber (CLT) or timber-concrete composites (TCC). This can represent an effective greenhouse gas reduction of 75% for the portions replaced.[28] This approach can be done at scale: Microsoft is building a new data center using CLT as a structural building material, representing an overall greenhouse gas reduction of 35% for the core and shell.[29]

Although this strategy can and should be adopted more widely, material substitution reductions are real but limited. Obtaining and using these materials can add cost and complexity to commercial projects. Moreover, concrete foundations and structural steel elements are very difficult and/or very expensive to replace. Additional steps and strategies are required to achieve deep Scope 3 reductions.

ii. Low-C manufacturing of building materials

Because many of the materials used in constructing data centers have high embodied carbon intensities, producers must change feedstocks or add technologies and practices to achieve significantly lower carbon production. Consequently, supply of low-carbon products is globally limited. To achieve economic viability, production facilities will require buyers willing to pay a green premium or additional market-aligning policies that support low-carbon production. However, producers could significantly reduce Scope 3 emissions of products by applying existing or novel technology, changing feedstocks, and innovating supply chains and business models as options develop and enter the market.

Box 3.3-2

Environmental attribute certificates (EACs)

It is often impractical, costly or carbon intensive to transport low-carbon goods from their point of manufacturing to the point of use. Increasingly, environmental attribute certificates (EACs) allow companies, in effect, to sever environmental attributes from low-carbon goods, pay for them, and claim them without direct physical transport or use. This technique has served buyers and producers of sustainable aviation fuel,[30] clean electricity[31] and low-carbon cement.[32]

EACs can help speed investment and development of clean goods, materials and services.[33] They also present risks, where carbon accounting practices are not strictly followed, potentially leading to false claims and misrepresentation of greenhouse gas enterprise emissions.

Companies and regulators should acquaint themselves with the use of EACs to speed technology entry into markets. They should also use quality criteria to ensure that use of EACs avoids double-counting estimated carbon benefits or misrepresenting them in corporate claims. Quality criteria are beginning to emerge based on both procurement practices and science-based technical assessment (e.g., Carbon Direct and Microsoft (2025)[33]).

a: CCUS

Carbon capture, use and storage (CCUS) is a set of established technologies to directly control carbon emissions. CCUS includes the following:

- Carbon capture, which separates and concentrates CO_2 from air or industrial facilities, such as power or steel production. This process most commonly involves chemical or physical separation with solvents.[34-36]

- CO_2 transportation, which brings CO_2 from the capture facility to the site of use or storage.[37,38] This process most commonly involves dedicated CO_2 pipelines, but it can also involve transport by ship, barge, truck or rail.[39]

- Use of CO_2, either through direct use or conversion into other products like fuel, chemicals or building materials.[40,41]

- Geological storage of CO_2 in dedicated storage facilities. This process most commonly involves storage in deep saline formations or depleted oil and gas fields, but it can also involve storage in basaltic formations or direct mineralization.[42-44]

More than 50 CCUS facilities operate today, capturing and storing over 60 million tons of CO_2 per year.[45] However, while many workers have identified CCUS as a promising technology to reduce the greenhouse gas footprint of heavy industrial production,[46,47] very few plants operate on these facilities.

Because of cement's intrinsic chemistry, CCUS remains one of the few ways to deeply decarbonize cement production.[13,48] Similarly, CCUS remains an important opportunity for decarbonizing steel production that uses blast furnaces,[17] both due to blast-furnace chemistry and high temperature heat requirements. CCUS can also provide low-carbon "blue" hydrogen for DRI production, as it does today in Abu Dhabi.[42] However, the timeline for decarbonized steel appears to be much longer than that for cement.

Box 3.3-3

Brevik low-carbon cement facility (Norway)

Recently, Heidelberg Materials commissioned the carbon capture, use and storage (CCUS) facility at their Brevik plant in Norway, which will capture 400,000 tonnes of CO_2 each year, ship it to the North Sea and store it over 1 km below the sea floor.[49] This will reduce the carbon intensity of the cement they produce by 50%.[50]

The Brevik facility, with integrated cement production and CCUS. Source: Heidelberg Materials[51]

Although CCUS technologies are well established, deployment today is limited. Commercial deployment is capital and energy intensive, adding significant cost. CCUS for heavy industrial applications would enter markets with small margins and lack of global standards for product carbon intensity. The EU carbon border adjustment mechanism (CBAM) covers both steel and cement but does not yet impact production.

Ultimately, CCUS is promising for emissions reductions. Additional policy measures are required to cover the green premium or additional costs.

b: Biomass

In producing low-carbon building materials, biomass is both prominent and promising. Industrial-grade biomass comprises forestry residues, agricultural wastes, municipal wastes and biomethane, which could provide a low-carbon, high energy-density fuel for many industrial applications, including steel and concrete production.[13,52,53] In the case of primary iron production, biocoke can also provide the chemical energy and physical properties necessary for blast-furnace operation,[54,55] as demonstrated in commercial operations in both Brazil[56] and Japan.[57] When combined with CCUS, use of biomass creates the potential for both profound emissions reductions and CO_2 removal in production—carbon-negative manufacturing.[17,58]

Several factors and conditions must be met for biomass use to significantly reduce the carbon intensity of manufactured goods:

- **Sustainability**: First and foremost, biomass must be sustainably sourced.[33,59] If biomass is not harvested and sourced sustainably, it could lead to increased deforestation, loss of soil carbon, and other direct and indirect greenhouse gas emissions.[60] Multiple international standards exist, but nations and operators have not yet agreed to adopt one standard or set.

- **Energy density**: Different forms of biomass have different energy content and energy density. Industrial operations commonly require high energy density fuels for heat, and biocoke in particular requires both high energy density and specific physical properties (e.g., high yield strength).

- **Continuous supply**: Commercial manufacturing operations require biomass delivery that is consistent in volume, energy content, moisture and schedule. Many biomass conversion facilities face challenges in maintaining consistent and regular biomass supply.

To avoid poor or counterproductive purchases, data center buyers and builders must acquaint themselves with the potential risks of biomass-based reduction strategies, chiefly from compromised feedstock supplies.

c: Clean direct electrification

Electrification of heavy manufacturing is challenging. This is particularly true for processes that involve chemical reduction or dissociation or require F-gas use.[48] In some cases, key geographies (e.g., Taiwan, Korea, Japan) lack low-carbon electricity supplies, limiting what emissions reduction is possible simply due to the high carbon intensity of the grid. Finally, cases where high-quality heat is essential[13,61] often cannot be electrified at all, meaning significant greenhouse gas reductions through direct electrification (i.e., >10%) are extremely difficult.[17]

However, the situation is improving. Cost reductions, technology development and performance improvements have created decarbonization opportunities within specific assets and regions. Many companies have emerged since 2020 to produce low-carbon goods through direct or indirect electrification, with some technologies and companies reaching pilot or early demonstration stages.

EAFs are a mature technology to process scrap metal or sponge iron into usable steel or steel feedstock. Use of EAFs represents 30% of global steel production today.[62] While EAFs are incapable of new iron and steel production, they are essential for recycling steel and DRI production.[17,63] The industry has increased the fraction of steel produced with EAFs, with this trend expected to continue.[64] To reduce their Scope 1 and 2 emissions, some operators of scrap-EAFs have transitioned their electricity supply to renewable power,[65-67] which in turn reduces the Scope 3 emissions of their buyers, including data center builders.

For primary iron production, molten oxide electrolysis (MOE) and electrowinning are novel electrochemical technology approaches that convert iron ore into iron using electricity directly. MOE immerses iron ore in a molten oxide bath, electrically heated to ~1600 °C, at which point electric currents break down the ore into molten iron. In electrowinning, iron ore is suspended in a low-temperature alkaline solution (~110 °C), where a current reduces the iron ore to iron in a process similar to electroplating. Unlike the DRI process described below, MOE and electrowinning eliminate the need for an EAF to process sponge iron. Today, the technical readiness is low (TRL 5) with only small pilot plants in operation or design.[68]

Compared to the steel industry, direct electrification in the cement industry is not close to commercial deployment. There are no commercial-scale pilots in operation or construction. Conventional rotary kilns cannot be electrified, and few are near retirement today in China, India or the United States.[69] Advanced technologies—TRL 5 or lower today—are promising but still at the pilot scale. Promising technologies include electric calciners like LEILAC[70] and electrifying the kiln burning zone with resistive heating, mechanical heating, plasmas, or a combination of approaches.[71] Even then, these technologies would still require that by-product CO_2 be managed with CCUS for deep abatement—roughly 50% of the total footprint. Unlike with MOE or electrowinning with steel, there is no cement production pathway without byproduct CO_2.

LIMITS: While technical readiness is a principal challenge to direct electrification, supply of low-carbon electricity presents major challenges in most geographies:

- Overall, the lack of capacity in nuclear, wind or solar presents a challenge—indeed, data centers themselves are challenged finding such power to operate with minimal Scope 2 emissions. (See Chapter 3.2 of this Roadmap.)
- To make efficient use of the capital invested, high-capacity factors for facilities must be patched to high-capacity production of green electricity. This greatly limits geographies of operation.
- For plants seeking to develop renewable power, land access for wind and solar is increasingly challenging. This is particularly true in densely populated areas where industrial production is concentrated.
- All together, these three elements can add significant cost. Since these are commodity industries with small margins, small increases in manufacturing cost can end investment without advanced market commitments, pre-purchase or other kinds of guaranteed offtake.
- Finally, addition of large loads onto the grid to operate electrified industrial production could lead to addition of fossil generation, either new generation or increased capacity factors of existing plants. These could affect the system-wide footprint of electricity and prove counterproductive to climate goals.

Wide deployment of these technologies will have to overcome these challenges in many contexts, markets and geographies.

d: Clean indirect electrification - green hydrogen

Electrolytic hydrogen production using low-carbon electricity, also called green hydrogen production, provides both thermal energy and chemical reduction for manufacturing low-carbon materials in data centers.[17,72,73] The most prominent pathway involves DRI, using hydrogen instead of natural gas, in tandem with an EAF to convert the sponge iron to pig iron.[74] Interest in this pathway has grown significantly, with new projects announced and under development.[75] Several small commercial facilities are being built or are in operation today, including SSAB's (Svenskt Stål AB's) Hybrit facility.

Sweden has pioneered production of "green steel" using green hydrogen and direct reduction of iron (DRI). In particular, two Swedish companies have built and are building green steel facilities.

SSAB built and operates the Hybrit facility,[76] which it describes as "fossil free steel." Begun in 2016 in partnership with Vattenfall, Hybrit received research and development (R&D) support from the Swedish government from 2018 to 2024.[77] The plant began construction in 2018 and produced the first sponge iron in 2020. In 2025 the pilot plant became fully operational. As of 2024, the facility has produced 5000 tons of steel, with expectations to expand to 1.2 million tonnes per year before 2030.

Volvo, Mercedes-Benz, Ruukii Construction and several other companies have purchased offtake from Hybrit, but to date, no data center builders or operators have purchased this steel.

The Hybrit demonstration plant. Source: High North News.[78]

Stegra is deploying the same technology as SSAB at their Bowen Plant.[79] This facility will be Europe's first greenfield steel plant in 50 years and is set to commission in 2030. It uses 700 MW of renewable power (mostly hydropower) to make enough hydrogen for 5 million tonnes of steel per year.

LIMITS: The limits of direct electrification (above) with electricity supply also apply to indirect electrification and green hydrogen production—cost, duty cycle, grid impacts, etc. Moreover, electrolytic hydrogen faces challenges in the capital costs of electrolyzers and balance of plan. In addition, DRI-EAF production requires a special iron ore (magnetite) that is higher cost and has limited supply.

iii. Low-C manufacturing of IT equipment

The production of IT and related equipment for data centers involves many distinct manufacturing steps for hundreds to thousands of individual components and associated assembly, testing, packaging and shipment. Unlike structural materials, such as concrete and steel, IT equipment and the related systems that support them are highly heterogeneous. Thus, the methods to reduce emissions from their production vary widely. However, some general guidelines for emissions reductions include the following:

- Maximize the use of low-carbon electricity at all stages of production, including initial materials extraction and processing, electronic and electrical component production, semiconductor fabrication, final assembly and testing.

- Maximize the use of recycled materials (e.g., copper, steel, printed circuit board resins).[80]

- Extend the lifespan of IT equipment, including servers and networking equipment.[81,82] Notably, while refresh intervals for most IT equipment are generally lengthening, graphics processing unit (GPU) lifetime may be an exception to this, driven by factors such as rapid technology development.[83]

- Follow low-embodied-emissions procurement standards for electronics developed by industry consortia, such as the Global Electronics Council.[84]

- For semiconductor fabrication, ensure that F-gas emissions are minimized through exhaust gas destruction and related methods.[85]

- For gas-insulated switchgear, explore use of alternatives to SF_6 for electrical insulation.[86]

While some IT equipment providers have released estimates of the embodied emissions of specific hardware components,[8,87] comprehensive studies on optimizing overall embodied emissions reductions across data center IT and related equipment are lacking. Thus, there is a clear need for these assessments in a transparent and accessible form.[88,89]

Box 3-3.4

CO_2 removal and superpollutant reductions.

Even with significant action and investment, all pathways find that a significant fraction of emissions cannot be reduced by existing technology. This is particularly true for the embodied emissions of data centers, which lack cost-effective solutions and clean manufacturing capacity today and in the near term (before 2035).

Acknowledging these facts, many technology firms and data center builders have made significant commitments to purchasing valid, durable CO_2 removals to complement their commitments to rapid and profound greenhouse gas reductions, including their Scope 3 burden. Microsoft,[90,91] Google,[92] Meta,[93] Apple[94] and Amazon[95,96] have made significant corporate commitments and have large programs in CO_2 removal. These include the LEAF (Lowering Emissions by Accelerating Forest finance) program[97] and significant CO_2 removal purchases, such as nature-based,[98] engineered,[99] hybrid[100] and novel approaches.[101] All these companies acknowledge that emissions reductions—including Scope 3 emissions—are their priority but that they must be complemented by high quality CO_2 removal to manage irreducible emissions.[102]

Reducing short-lived, very strong greenhouse gases (sometimes called “superpollutants”) is another complementary approach to direct reductions. This approach includes investments in the destruction of potent non-CO_2 greenhouse gases like methane, nitrous oxides and F-gases outside of their direct value chain.[103] Recently, Google announced new commitments in superpollutant destruction,[103,104] again to complement their existing reduction targets and to strengthen their overall corporate commitments to achieving net-zero.

C. Innovation Agenda

The high cost and difficulty of reducing Scope 3 emissions demands investing in innovation that leads to solutions in global markets. Many governments, including Japan, the United States, the United Kingdom and European Union nations, have substantial research, development and demonstration (RD&D) programs focused on reducing embodied emissions through novel manufacturing processes, carbon management, electrification and dematerialization. In addition, many technology company buyers and manufacturers have ambitious sustainability targets and have made significant and sustained investment in RD&D to reduce Scope 3 greenhouse gas emissions in their value chains.[104]

Ultimately, rapid and profound reductions in embodied emissions requires more effort and investment. Specifically, the range of pathways to lowering embodied emissions and the focus on data center-related products and practices must increase. In some cases, promising pathways must receive additional support and scale to manifest solutions that can scale commercially. Examples include the following:

- **Targeted research**
    - Minimizing F-gas leakage and use in IT hardware manufacturing.
    - Developing novel pozzolinic materials that can reduce total clinker use in concrete.
    - Improving capital costs for green hydrogen, MOE and electrowinning pathways, with a focus on reducing balance of system cost.
    - Integrating carbon capture systems into cement and BF-BOF iron production.
    - Increasing energy density and mechanical strength in biocoke.
- **Cross-cutting research**
    - Identifying and removing adoption barriers for dematerialization strategies.
    - Identifying new options to generate low-carbon electricity in critical markets for producing IT hardware (e.g., Taiwan).
    - Generating novel design options for data centers that are inherently low embodied emissions.
    - Ranking opportunities based on speed of implementation, levelized cost and readiness (technical, infrastructure, workforce).

Given the rapid changes in data center markets, both technology evolution and emergence of needs, governments should not develop programs in isolation. Rather, they should tune existing programs in partnership with industry to maximize impact and avoid waste. Doing so will require a new level of trust and transparency, well beyond existing circumstances. The proprietary nature of many industrial innovations makes this difficult. Industrial, academic and governmental actors should come together to prioritize pre-commercial and shared RD&D agendas that can demonstrate progress against the rate and scale of embodied emissions growth.

D. Recommendations

1. *Governments should* ***assemble and share data related to direct, indirect and embodied greenhouse gas emissions from data center construction and operation****. Data center owners and operators should* ***volunteer to share site-specific estimated Scope 3 emissions data proactively and invite third-party review****. If necessary, governments should* ***require disclosure of this information****.*

2. *All stakeholders should* ***gain familiarity with the embodied emissions of data centers. They should recognize that abatement options today are real but limited and potentially expensive.***

3. *Before designing and siting data centers, data center owners and operators should* ***identify and assess potential options to reduce Scope 3 emissions through material reduction and substitution****. Companies should use existing scientific criteria for high-quality, low-carbon goods and should consider developing their own criteria.*

4. *During procurement and construction phases, data center owners and operators should* ***assess the availability of low-carbon strategies and materials, including IT materials and building materials*** *and use those low-carbon strategies and materials wherever possible. They should consider EACs to speed emissions reduction and support low-carbon manufacturing facilities, such as biocoke in blast furnaces, carbon-free steel production and cement with CCUS. They should also consider adhering to low embodied-carbon procurement standards for electronics developed by industry consortia.*

5. *Governments should support* ***comprehensive, transparent studies on optimizing overall embodied emissions reductions*** *across the full spectrum of data center IT equipment. These studies should be conducted by independent, third-party researchers, with relevant data shared voluntarily by data center operators.*

6. *During the operational phase, data center operators should* ***minimize IT equipment refresh rates*** *and seek to procure low embodied-emissions servers, networking equipment, memory and related equipment.*

7. *Governments should* ***assess the current supplies of low-carbon building materials*** *and* ***consider adding production capacity through policy measures****, including direct grants, government-backed procurement, contracts for differences, etc. They should also consider regulating production of IT hardware to reduce emissions, in particular focusing on F-gas use, leakage and destruction.*

8. *Governments should* ***support development of advanced technologies that limit the greenhouse gas footprint*** *associated with data center construction. They should explore and support applied research into alternative production approaches to chip-making that use less F-gases and manage their leakage better. They should explore alternative pathways to manufacturing cement, concrete and steel.*

E. References

1. World Resources Institute (WRI) and World Business Council for Sustainable Development (WBCSD). Greenhouse Gas Protocol: FAQ; WRI and WBCSD, Washington, DC, https://ghgprotocol.org/sites/default/files/2022-12/FAQ.pdf (Accessed August 2025).

2. Paul Lin, Robert Bunger & Victor Avelar. Quantifying Data Center Scope 3 GHG Emissions to Prioritize Reduction Efforts (White Paper 99; Version 1); Schneider Electric - Energy Management Research Center, Andover, Massachusetts, https://media.datacenterdynamics.com/media/documents/SE-Quantifying_Data_Center_Scope_3_GHG_Emissions_to_Prioritize_Reduction_Efforts.pdf (2023).

3. Tuğana Aslan, Peter Holzapfel, Lutz Stobbe, Andreas Grimm, Nils F. Nissen & Matthias Finkbeiner. Toward climate neutral data centers: Greenhouse gas inventory, scenarios, and strategies. iScience 28 (2025). https://doi.org/10.1016/j.isci.2024.111637.

4. Microsoft. Environmental Sustainability Report 2025; Redmond, Washington, https://www.microsoft.com/en-us/corporate-responsibility/sustainability/report/?icid=environmental2025_sustainablereport_MSSustWebsite_Homepage (2025).

5. Steven Downes. Microsoft's New Strategy to Cut Soaring Scope 3 Emissions; Sustainability Magazine, Norwich, Norfolk, Great Britain, https://sustainabilitymag.com/articles/microsofts-new-strategy-to-cut-soaring-scope-3-emissions (2024).

6. JRM Construction Management. Core and Shell Construction Guide; New York, New York, https://www.jrmcm.com/core-shell/core-and-shell-construction-guide/#:~:text=The%20core%20is%20a%20building's,exterior%20features%20and%20protective%20envelope (2024).

7. Mary Zhang. Building a Data Center: An In-Depth Guide to Construction; Dgtl Infra LLC, New York, New York, https://dgtlinfra.com/building-data-center-construction/#:~:text=The%20shell%20is%20the%20external,%2C%20exterior%20walls%2C%20and%20waterproofing. (2023).

8. Ian Schneider, Hui Xu, Stephan Benecke, David Patterson, Keguo Huang, Parthasarathy Ranganathan & Cooper Elsworth. Life-Cycle Emissions of AI Hardware: A Cradle-To-Grave Approach and Generational Trends. arXiv:2502.01671 (2025). https://ui.adsabs.harvard.edu/abs/2025arXiv250201671S.

9. Data Center Knowledge. Data Center Power: Fueling the Digital Revolution; Data Center Knowledge (Informa TechTarget), Newton, Massachusetts, https://www.datacenterknowledge.com/energy-power-supply/data-center-power-fueling-the-digital-revolution (2024).

10. Josh Mahan. Understanding Data Center Energy Consumption; C&C Technology Group, Mahwah, New Jersey, https://cc-techgroup.com/data-center-energy-consumption/ (Accessed August 2025 (Last updated June 2023)).

11. Anthony Waterman. Technical Information Paper: Embodied Carbon in Enterprise Data Centre IT Equipment; ADW Developments, Bushey, United Kingdom, https://adwdevelopments.com/sustainability/technical-information-paper-embodied-carbon-in-enterprise-data-centre-it-equipment/ (2025).

12. Michael Purton. Cement is a big problem for the environment. Here's how to make it more sustainable; World Economic Forum, Cologny/Geneva, Switzerland, https://www.weforum.org/stories/2024/09/cement-production-sustainable-concrete-co2-emissions/ (2024).

13. David Sandalow, Julio Friedmann, Roger Aines, Colin McCormick, Sean McCoy & Joshuah Stolaroff. ICEF Industrial Heat Decarbonization Roadmap; ICEF Innovation Roadmap Project, https://www.icef.go.jp/wp-content/themes/icef_new/pdf/roadmap/icef2019_roadmap.pdf (2019).

14. Natesan Mahasenan, Steve Smith & Kenneth Humphreys. - The Cement Industry and Global Climate Change: Current and Potential Future Cement Industry CO_2 Emissions in Greenhouse Gas Control Technologies - 6th International Conference (eds J. Gale & Y. Kaya) 995-1000 (Pergamon, Oxford, 2003, https://www.sciencedirect.com/science/article/pii/B9780080442761501574).

15. Johanna Lehne & Felix Preston. Making Concrete Change: Innovation in Low-carbon Cement and Concrete (Chatham House Report); The Royal Institute of International Affairs, Chatham House, London, United Kingdom, https://www.chathamhouse.org/sites/default/files/publications/research/2018-06-13-making-concrete-change-cement-lehne-preston.pdf (2018).

16. World Steel Association. Sustainability Indicators 2024 report: Sustainability performance of the steel industry 2003-2023; Brussels, Belgium, https://worldsteel.org/wider-sustainability/sustainability-indicators/#co2-emissions-and-energy-intensity (2024).

17. Zhiyuan Fan & S. Julio Friedmann. Low-carbon production of iron and steel: Technology options, economic assessment, and policy. Joule 5, 829-862 (2021). https://doi.org/10.1016/j.joule.2021.02.018.

18. Amber Jackson. The Role of Steel in Today's Data Centre Industry; Data Centre Magazine, Norwich, Norfolk, United Kingdom, https://datacentremagazine.com/articles/the-role-of-steel-in-todays-data-centre-industry (2025).

19. Pat Lynch, Gordon Dolven & Josh Ruttner. North America Data Center Trends H2 2024; CBRE Group Inc., Dallas, Texas, https://www.cbre.com/insights/reports/north-america-data-center-trends-h2-2024 (2025).

20. John McWilliams, Jessica Howe, Andrew Fray, Vivek Dahiya & Pritesh Swamy. Global Data Center Market Comparison; Cushman & Wakefield, Chicago, Illinois, https://www.cushmanwakefield.com/en/insights/global-data-center-market-comparison (2025).

21. Charlotte Kenna. Data Centre Take-up in Europe to Reach New Peak in 2025 (Press Release); CBRE Group, Inc., Dallas, Texas, https://www.cbre.co.uk/press-releases/data-centre-take-up-in-europe-to-reach-new-peak-in-2025 (2025).

22. Laura Wood (ResearchAndMarkets.com). Middle East & Africa Data Center Market Landscape Report 2025-2030 | Investments to Propel MEA Data Center Market to USD 19.89 Billion by 2030 - ResearchAndMarkets.com Business Wire, Inc., San Francisco, California, https://www.businesswire.com/news/home/20250523886842/en/Middle-East-Africa-Data-Center-Market-Landscape-Report-2025-2030-Investments-to-Propel-MEA-Data-Center-Market-to-USD-19.89-Billion-by-2030---ResearchAndMarkets.com (2025).

23. Qi Wang, Nan Huang, Zhuo Chen, Xiaowen Chen, Hanying Cai & Yunpeng Wu. Environmental data and facts in the semiconductor manufacturing industry: An unexpected high water and energy consumption situation. Water Cycle 4, 47-54 (2023). https://doi.org/10.1016/j.watcyc.2023.01.004.

24. Sébastien Raoux. Fluorinated greenhouse gas and net-zero emissions from the electronics industry: the proof is in the pudding. Carbon Management 14, 2179941 (2023). https://doi.org/10.1080/17583004.2023.2179941.

25. Muhammad Usman. Switching Gears: The U.S. Move Toward Gas-Insulated MV Switchgear; Microgrid Knowledge (Endeavor Business Media, EndeavorB2B), Nashville, Tennessee, https://www.microgridknowledge.com/engineering/article/55140304/switching-gears-the-us-move-toward-gas-insulated-mv-switchgear (2024).

26. Bill Flederbach, Scott Subler, Lauren Mechak & Kayla Carey. Low-Carbon Cement Productions (Issue Paper); ClimeCo Corporation, Boyertown, Pennsylvania, https://climateactionreserve.org/wp-content/uploads/2022/10/Low-Carbon-Cement-Issue-Paper-05-20-2022_final.pdf (2022).

27. National Concrete Pavement Technology Center. National Concrete Pavement Technology Center; Iowa State University of Science and Technology, Ames, Iowa, https://www.cptechcenter.org/cementitious-materials/ (Accessed August 2025).

28. Jae-Won Oh, Keum-Sung Park, Hyeon Soo Kim, Ik Kim, Sung-Jun Pang, Kyung-Sun Ahn & Jung-Kwon Oh. Comparative CO_2 emissions of concrete and timber slabs with equivalent structural performance. Energy and Buildings 281, 112768 (2023). https://doi.org/10.1016/j.enbuild.2022.112768.

29. Sally Beatty. Microsoft builds first datacenters with wood to slash carbon emissions; Microsoft, Redmond, Washington, https://news.microsoft.com/source/features/sustainability/microsoft-builds-first-datacenters-with-wood-to-slash-carbon-emissions/ (2024).

30. Elsa Wenzel. Corporations buy into sustainable fuel certificates to address air travel emissions; Trellis Group Inc., Oakland, California, https://trellis.net/article/corporations-buy-sustainable-fuel-certificates-address-air-travel-emissions/ (2024).

31. Shannon Hughes & Samuel Huestis. Clean Energy 101: The REC Market; Rocky Mountain Institute (RMI), Basalt, Colorado, https://rmi.org/clean-energy-101-the-rec-market/ (2022).

32. First Movers Coalition. Sublime Systems and Microsoft conclude significant multi-year offtake deal to supply FMC-aligned cement; Centre for Nature and Climate (World Economic Forum), Cologny/Geneva, Switzerland, https://initiatives.weforum.org/first-movers-coalition/sublimesystems-microsoft (2025).

33. Carbon Direct & Microsoft. Criteria for High-Quality Environmental Attribute Certificates (2025 Edition); New York, New York and Redmond, Washington, https://www.globalccsinstitute.com/publications/ (2025).

34. Hugh Barlow, Shahrzad S. M. Shahi & David T. Kearns. Advancements in CCS Technologies and Costs; Global CCS Institute, Melbourne, Australia, https://www.globalccsinstitute.com/wp-content/uploads/2025/01/Advancements-in-CCS-Technologies-and-Costs-Report-2025.pdf (2025).

35. Cristiano Veloso. The Science of Carbon Capture: A Deep Dive into the Chemistry (Blog); Verde Agritech, Belo Horizonte, Brazil, https://blog.verde.ag/en/the-science-of-carbon-capture/ (2023).

36. Hugh Barlow, Shahrzad S. M. Shahi, Javad Jeddizahed & David T. Kearns. State of the Art: CCS Technologies 2025 (Technical Report); Global CCS Institute, Melbourne, Australia, https://www.globalccsinstitute.com/wp-content/uploads/2025/08/State-of-the-Art-CCS-Technologies-2025-Global-CCS-Institute.pdf (2025).

37. Joey Minervini, Chris Consoli & David Kearns. CCS Networks in the Circular Carbon Economy: Linking Emissions Sources to Geologic Storage Sinks; Global CCS Institute, Melbourne, Australia, https://www.globalccsinstitute.com/resources/publications-reports-research/ccs-networks-in-the-circular-carbon-economy-linking-emissions-sources-to-geologic-storage-sinks/ (2021).

38. Peter Psarras & Hélène Pilorgé. The CO_2 Transportation Challenge; Kleinman Center for Energy Policy (University of Pennsylvania), Philadelphia, Pennsylvania, https://kleinmanenergy.upenn.edu/commentary/podcast/the-co2-transportation-challenge/ (2024).

39. Jennifer Pett-Ridge, Hamed Ziad Ammar, Alvina Aui, Mark Ashton, Sarah E. Baker, Bruno Basso, Mark Bradford, Alexander P. Bump, Ingrid Busch, Edna Rodriguez Calzado, Jackson W. Chirigotis, Nicolas Clauser, Sinéad Crotty et al. Chapter 5:

CO_2 and Biomass Transportation in Roads to Removal: Options for Carbon Dioxide Removal in the United States (LLNL-TR-852901) (Lawrence Livermore National Laboratory, Livermore, California, 2023, https://roads2removal.org/wp-content/uploads/05_RtR_CO2-Biomass-Transport.pdf).

40. National Academies of Sciences Engineering and Medicine (NASEM). Carbon Utilization Infrastructure, Markets, and Research and Development: A Final Report; The National Academies Press, Washington, DC, https://doi.org/10.17226/27732 (2024).

41. Amar Bhardwaj, Colin McCormick & Julio Friedmann. Opportunities and Limits of CO_2 Recycling in a Circular Carbon Economy: Techno-economics, Critical Infrastructure Needs, and Policy Priorities; Center on Global Energy Policy at Columbia University, School of International and Public Affairs, New York, New York, https://www.energypolicy.columbia.edu/publications/opportunities-and-limits-co2-recycling-circular-carbon-economy-techno-economics-critical/ (2021).

42. Carbon Sequestration Leadership Forum (CSLF). Carbon Storage Atlas - Fifth Edition (2015); Office of Fossil Energy, US Department of Energy (DOE), Washington, DC, https://fossil.energy.gov/archives/cslf/Publication/CO2StorageAtlasV.html (2015).

43. Peter D. Warwick, Madalyn S. Blondes, Sean T. Brennan, Margo D. Corum & Matthew D. Merrill. U.S. Geological Survey Geologic Carbon Dioxide Storage Resource Assessment of the United States. Energy Procedia 37, 5275-5279 (2013). https://doi.org/10.1016/j.egypro.2013.06.444.

44. Susan Hovorka & Peter Kelemen. Chapter 2.9: Geological sequestration in Carbon Dioxide Removal (CDR) Primer (2021, https://cdrprimer.org/read/chapter-2#sec-2-9).

45. Global CCS Institute. The Global Status of CCS 2024; Melbourne, Australia, https://www.globalccsinstitute.com/resources/global-status-report/ (2024).

46. National Energy Technology Laboratory (NETL). Point Source Carbon Capture from Industrial Sources; NETL, US Department of Energy (DOE), Albany, Oregon, https://netl.doe.gov/carbon-capture/industrial (Accessed August 2025).

47. European Commission. Industrial Carbon Management; Brussels, Belgium, https://climate.ec.europa.eu/eu-action/industrial-carbon-management_en (Accessed August 2025).

48. Adair Turner, Ajay Mathur, Faustine Delasalle, Tugce Balik, Laetitia de Villepin, Saira George, Thea Jung, Sachin Kapila, Isabel Lewren, Jeremy Oppenheim, Ricardo Santana, Per Klevnäs, Anders Ahlen et al. Mission Possible: Reaching net-zero carbon emissions from harder-to-abate sectors; Energy Transitions Commission, London, United Kingdom, https://www.energy-transitions.org/publications/mission-possible/ (2018).

49. Vetle Houg. Pioneering carbon capture in the cement industry: The Brevik CCS project; Heidelberg Materials, Heidelberg, Germany, https://www.heidelbergmaterials.com/sites/default/files/2024-07/Vetle%20Houg_Pioneering%20carbon%20capture%20in%20the%20cement%20industry.pdf (2024).

50. Christoph Beumelburg. World premiere at Heidelberg Materials: Opening of CCS facility in Norway marks new era of sustainable construction; Heidelberg Materials, Heidelberg, Germany, https://www.heidelbergmaterials.com/en/pr-2025-06-18 (2025).

51. Heidelberg Materials. About the project; Heidelberg, Germany, https://www.brevikccs.com/en/about-the-project (Accessed August 2025).

52. US Department of Energy (M. H. Langholtz (Lead)). BETO: Billion-Ton 2023 (ORNL/SPR-2024/310); Oak Ridge National Laboratory (ORNL), Oak Ridge, Tennessee, https://

www.energy.gov/eere/bioenergy/2023-billion-ton-report-assessment-us-renewable-carbon-resources (https://doi.org/10.23720/BT2023/2316165) (2024).

53. US Energy Information Administration (EIA). Biomass explained: Waste-to-energy (Municipal Solid Waste); EIA, Washington, DC, https://www.eia.gov/energyexplained/biomass/waste-to-energy-in-depth.php (Accessed August 2025 (Last updated 2024)).

54. Sahar Safarian. To what extent could biochar replace coal and coke in steel industries? Fuel 339, 127401 (2023). https://doi.org/10.1016/j.fuel.2023.127401.

55. Janusz Krupanek, Grzegorz Gałko, Marcin Sajdak & Marta Pogrzeba. Comparison of Bio-Coke and Traditional Coke Production with Regard to the Technological Aspects and Carbon Footprint Considerations. Energies 17, 2978 (2024). https://doi.org/10.3390/en17122978.

56. Brasil Pelo Meio Ambiente (BPMA). Biocoke; Amcham Brasil, São Paulo, Brazil, https://brasilpelomeioambiente.com.br/en/project/biocoke/ (Accessed August 2025).

57. Quantum Commodity Intelligence. Japanese auto parts supplier starts bio-coke trial at production lines; Quantum Commodity Intelligence Ltd., London, United Kingdom, https://www.qcintel.com/biofuels/article/japanese-auto-parts-supplier-starts-bio-coke-trial-at-production-lines-40688.html (2025).

58. Ilman Nuran Zaini, Anissa Nurdiawati, Joel Gustavsson, Wenjing Wei, Henrik Thunman, Rutger Gyllenram, Peter Samuelsson & Weihong Yang. Decarbonising the iron and steel industries: Production of carbon-negative direct reduced iron by using biosyngas. Energy Conversion and Management 281, 116806 (2023). https://doi.org/10.1016/j.enconman.2023.116806.

59. Asri Gani, Erdiwansyah, Edi Munawar, Mahidin, Rizalman Mamat & S. M. Rosdi. Investigation of the potential biomass waste source for biocoke production in Indonesia: A review. Energy Reports 10, 2417-2438 (2023). https://doi.org/10.1016/j.egyr.2023.09.065.

60. Sustainability Directory. Why Is Biomass Sourcing so Important?; https://energy.sustainability-directory.com/question/why-is-biomass-sourcing-so-important/ (2025).

61. Julio Friedmann, Zhiyuan Fan & Ke Tang. Low-Carbon Heat Solutions for Heavy Industry: Sources, Options, and Costs Today; Center on Global Energy Policy at Columbia University, School of International and Public Affairs, New York, New York, https://www.energypolicy.columbia.edu/publications/low-carbon-heat-solutions-heavy-industry-sources-options-and-costs-today/ (2019).

62. World Steel Association. What is the EAF-BOF split; Brussels, Belgium, https://worldsteel.org/about-steel/facts/steelfacts/what-is-steel/what-is-the-eaf-bof-split/ (Accessed August 2025).

63. World Steel Association. The steelmaking process; Brussels, Belgium, https://worldsteel.org/about-steel/steelmaking-process/ (Accessed August 2025).

64. Yale Environment 360. Steel Industry Pivoting to Electric Furnaces, Analysis Shows; Yale School of the Environment, New Haven, Connecticut, https://e360.yale.edu/digest/steel-industry-carbon-coal-electric-arc-furnaces (2023).

65. World Steel Association. Solar energy is fuelling more sustainable steel production; Brussels, Belgium, https://worldsteel.org/media/steel-stories/infrastructure/solar-energy-fuels-sustainable-production-of-rails/ (Accessed August 2025).

66. Freedom Solar. Colorado's Now Home to the World's Largest Solar-Powered Steel Plant; Freedom Solar LLC, Austin, Texas, https://freedomsolarpower.com/blog/worlds-largest-solar-powered-steel-plant-in-colorado (2023).

67. Nucor & WattTime. Nucor, Emissionality, and the Pursuit of Green Steel; Charlotte, North Carolina and Oakland, California, https://watttime.org/wp-content/uploads/2023/10/WattTime-Nucor-Case-Study-202012-vFinal.pdf (2023).

68. World Steel Association. Fact sheet: Electrolysis in ironmaking; Brussels, Belgium, https://worldsteel.org/wp-content/uploads/Fact-sheet-Electrolysis-in-ironmaking.pdf (2021).

69. Clean Air Task Force (CATF) & Inc. Synapse Energy Economics. Recasting the Future: Policy Approaches to Drive Cement Decarbonization; Boston, Massachusetts and Cambridge, Massachusetts, https://www.catf.us/resource/recasting-future-policy-approaches-drive-cement-decarbonization/ (2025).

70. LEILAC. LEILAC Technology Roadmap to 2050; Pymble, Australia, https://www.leilac.com/wp-content/uploads/2022/09/LEILAC-Roadmap.pdf (2021).

71. Kanthal. ELECTRA project: Heidelberg Materials and Kanthal unite to electrify cement; Kanthal AB, Hallstahammar, Sweden, https://www.kanthal.com/en/knowledge-hub/inspiring-stories/heidelberg-kanthal-electrify-cement-industry/ (2025).

72. Zhiyuan Fan, Emeka Ochu, Sarah Braverman, Yushan Lou, Griffin Smith, Amar Bhardwaj, Jack Brouwer, Colin McCormick & Julio Friedmann. Green Hydrogen in a Circular Carbon Economy: Opportunities and Limits; Center on Global Energy Policy at Columbia University, School of International and Public Affairs, New York, New York, https://www.energypolicy.columbia.edu/publications/green-hydrogen-circular-carbon-economy-opportunities-and-limits/ (2021).

73. International Energy Agency (IEA). Hydrogen; IEA, Paris, France, https://www.iea.org/energy-system/low-emission-fuels/hydrogen (Accessed August 2025).

74. Soroush Basirat. Hydrogen unleashed: Opportunities and challenges in the evolving H2-DRI-EAF pathway beyond 2024; Institute for Energy Economics & Financial Analysis, Valley City, Ohio, https://ieefa.org/resources/hydrogen-unleashed-opportunities-and-challenges-evolving-h2-dri-eaf-pathway-beyond-2024 (2024).

75. International Energy Agency (IEA). Steel; IEA, Paris, France, https://www.iea.org/energy-system/industry/steel (Accessed August 2025).

76. Svenskt Stål AB (SSAB). HYBRIT®. A new revolutionary steelmaking technology; SSAB, Stockholm, Sweden, https://www.ssab.com/en-us/fossil-free-steel/insights/hybrit-a-new-revolutionary-steelmaking-technology (Accessed August 2025).

77. Hybrit. HYBRIT: Facts and milestones; Stockholm, Sweden, https://www.hybritdevelopment.se/en/hybrit-facts-and-milestones/ (Accessed August 2025).

78. Hilde-Gunn Bye & Birgitte Annie Hansen. Fossil-Free Steel Production in Northern Sweden: The Hybrit Project Goes Industrial; High North News, Bodø, Norway, https://www.highnorthnews.com/en/fossil-free-steel-production-northern-sweden-hybrit-project-goes-industrial (2024).

79. Stegra. Our Boden plant; Stockholm, Sweeden, https://stegra.com/the-boden-plant (Accessed August 2025).

80. Xiao-xuan Wei, Chun-chen Nie, Yue-xian Yu, Jun-xiang Wang, Xian-jun Lyu, Peng Wu & Xiang-nan Zhu. Environment-friendly recycling of resin in waste printed circuit boards. Process Safety and Environmental Protection 146, 694-701 (2021). https://doi.org/10.1016/j.psep.2020.12.008.

81. Data Center Knowledge. Data Center Hardware Refresh Cutback by Microsoft — What's Next?; Data Center Knowledge (Informa TechTarget), Newton, Massachusetts, https://www.datacenterknowledge.com/hyperscalers/data-center-hardware-refresh-cutback-by-microsoft-what-s-next- (2022).

82. Horizon Editorial. Navigating Hardware Refresh Cycles in the Data Center; Horizon Technology, Lake Forest, California, https://horizontechnology.com/news/data-center-hardware-refresh-cycles/ (2024).

83. Luke Emberson & David Owen. The stock of computing power from NVIDIA chips is doubling every 10 months; Epoch AI, San Francisco, California, https://epoch.ai/data-insights/nvidia-chip-production (2025).

84. Green Electronics Council. EPEAT-CCM-2023: Climate Change Mitigation Criteria; Green Electronics Council dba Global Electronics Council (GEC), Portland, Oregon, https://globalelectronicscouncil.org/wp-content/uploads/EPEAT_CCM_2023.pdf (2023).

85. Shikai Zhu, Haoqian Hu, Haoyi Yang, Yunzhuo Qu & Yuanzhe Li. Mini-Review of Best Practices for Greenhouse Gas Reduction in Singapore's Semiconductor Industry. Processes 11, 2120 (2023). https://doi.org/10.3390/pr11072120

86. Karsten Burges, Kristina Warncke & Barbara Gschrey. Briefing paper: SF6 and alternatives in electrical switchgear and related equipment; RE-xpertise and Öko-Recherche, Berlin, Germany and Frankfurt am Main, Germany, https://climate.ec.europa.eu/system/files/2020-04/2020_03_25_sf6_and_alternatives_en.pdf (2020).

87. NVIDIA. Product Carbon Footprint (PCF) Summary for NVIDIA HGX H100; Santa Clara, California, https://images.nvidia.com/aem-dam/Solutions/documents/HGX-H100-PCF-Summary.pdf (2025).

88. Christiane Plociennik, Ponnapat Watjanatepin, Karel Van Acker & Martin Ruskowski. Life Cycle Assessment of Artificial Intelligence Applications: Research Gaps and Opportunities. Procedia CIRP 135, 924-929 (2025). https://doi.org/10.1016/j.procir.2025.01.079.

89. Adrien Berthelot, Eddy Caron, Mathilde Jay & Laurent Lefèvre. Estimating the environmental impact of Generative-AI services using an LCA-based methodology. Procedia CIRP 122, 707-712 (2024). https://doi.org/10.1016/j.procir.2024.01.098.

90. Microsoft. Carbon dioxide removal: Removing our historical carbon emissions by 2050 Redmond, Washington, https://www.microsoft.com/en-us/corporate-responsibility/sustainability/carbon-removal-program (Accessed August 2025).

91. Mark Segal. Microsoft Signs Deal to Remove 1.1 Million Tons of CO_2 Through Waste-to-Energy Carbon Capture; ESG Today, New York, New York, https://www.esgtoday.com/microsoft-signs-deal-to-remove-1-1-million-tons-of-co2-through-waste-to-energy-carbon-capture/ (2025).

92. Randy Spock. Our progress to accelerate carbon removal solutions (Blog); Google, Mountain View, California, https://blog.google/outreach-initiatives/sustainability/our-progress-to-accelerate-carbon-removal-solutions/ (2025).

93. Meta. Growing Our Commitment to Carbon Removal with the U.S. Department of Energy; Menlo Park, California, https://sustainability.atmeta.com/blog/2024/10/11/growing-our-commitment-to-carbon-removal-with-the-u-s-department-of-energy/ (2024).

94. Apple. Apple's Carbon Removal Strategy; Apple, Inc., Cupertino, California, https://www.apple.com/environment/pdf/Apples_Carbon_Removal_Strategy_White_Paper.pdf (2024).

95. Amazon. How Amazon approaches carbon credits, a key tool in the fight against climate change; Amazon.com, Inc., Seattle, Washington, https://www.aboutamazon.com/news/sustainability/amazon-carbon-credit-decarbonization-sustainability (2025).

96. Amazon. Amazon supports the world's largest deployment of direct air capture technology to remove carbon from the atmosphere; Amazon.com, Inc., Seattle, Washington, https://www.aboutamazon.com/news/sustainability/amazon-direct-air-capture-investment-fights-climate-change (2023).

97. Amazon. Amazon helps mobilize $1 billion to protect rainforests worldwide; Seattle, Washington, https://www.aboutamazon.com/news/sustainability/amazon-helps-mobilize-1-billion-to-protect-rainforests-worldwide (2021).

98. Mombak. Microsoft signs mega carbon credit deal with Brazilian company Mombak; São Paulo, Brazil, https://mombak.com/news/microsoft-signs-mega-carbon-credit-deal-with-brazilian-company-mombak/ (2024).

99. Mark Segal. Microsoft Signs Largest-Ever DAC Carbon Credit Purchase Agreement with Oxy's 1PointFive; ESG Today, New York, New York, https://www.esgtoday.com/microsoft-signs-largest-ever-dac-carbon-credit-purchase-agreement-with-oxys-1pointfive/ (2024).

100. ESG Today. Microsoft Signs Largest-Ever Permanent Carbon Removal Purchase Agreement; New York, New York, https://www.esgtoday.com/microsoft-signs-largest-ever-permanent-carbon-removal-purchase-agreement/ (2025).

101. Jennifer L. Google, Meta, and Others Invest $41M in Carbon Removal Credits; CarbonCredits.Com, Houston, Texas, https://carboncredits.com/google-meta-and-others-invest-41m-in-carbon-removal-credits-frontier/ (2025).

102. Carbon Direct & Microsoft. Criteria for High-Quality Carbon Dioxide Removal (2025 Edition); New York, New York and Redmond, Washington, https://www.carbon-direct.com/criteria/2025-edition (2025).

103. Climate and Clean Air Coalition (CCAC). Super Pollutants; UN Environment Programme (UNEP), Paris, France, https://www.ccacoalition.org/news/super-pollutants (2024).

104. Zachary Skidmore. Google partners with Recoolit and Cool Effect to remove 25,000 tons of super pollutants by 2030; Data Center Dynamics (DCD), London, United Kingdom, https://www.datacenterdynamics.com/en/news/google-partners-with-recoolit-and-cool-effect-to-remove-25000-tons-of-super-pollutants-by-2030/ (2025).

4 Accelerating Low-Carbon Power with AI Data Centers

Ayse Coskun, Varun Sivaram and Swasti Jain

A. Introduction: The Double-Edged Sword of AI's Energy Consumption

i. Where AI meets the grid

Over the past decade, the proliferation of digital technologies has quietly reshaped global electricity demand. Now, with the exponential growth of artificial intelligence (AI), the digital sector—specifically data centers powering AI workloads—is rapidly emerging as one of the most transformative forces in global energy systems. As deployment accelerates across sectors, from finance and healthcare to manufacturing, hyperscale AI models are triggering surging demand for computational power and, by extension, electricity. Behind every model training run and wave of new applications sits a physical footprint of racks, wires, cooling loops and a rising, highly concentrated draw on the grid. Globally, data center electricity use is now about 1.5% of demand and, in the International Energy Agency's (IEA) Base Case Scenario, doubles to ~3% by 2030. Much of this load is clustered in areas where the scale of data center electricity demand strains local power adequacy and complicates long-term grid planning.[1] (See Chapter 1(C) of this Roadmap.) According to the IEA's Energy and AI Observatory, notable clusters include Loudoun County in Virginia, Dublin in Ireland, Singapore, Tokyo and the Amsterdam–Frankfurt–London corridor (see Figure 4-1).[2] While many of these clusters are urban or metropolitan hubs, large AI campuses are rising in more rural or remote regions where the relative strain on smaller grids can be even more pronounced. In places like rural Grant County in Washington (home to a cluster of hyperscale facilities around Quincy), the public utility has capped data center load growth because demand is outpacing local transmission capacity, underscoring how concentrated build-outs can strain smaller grids.[3]

Figure 4-1. *International Energy Agency (IEA) analysis of data center locations, powerlines and submarine fiber optic network data. The data center hubs shown below represent the capacity-weighted centroid of a cluster of data centers within 100 km of each other and total over 500 MW of installed capacity.*

What fills this near-term gap is not pre-ordained. Some planners lean on familiar dispatchable options, such as natural gas or coal-fired power plants, to guarantee reliability. Yet recent experience tells a more nuanced story. In the United States, solar and utility-scale batteries led new capacity additions in 2024. This shift signals multiple viable pathways to serve new loads and create workforce and institutional capacity to deploy these systems more quickly and at lower cost.[2,4] While capacity additions in 2023 were dominated by wind and solar, forward-looking integrated resource plans (IRPs) in high-growth states now include substantial natural gas capacity. This strategy is driven in part by differing views on the ability of firmed renewables, including wind and solar paired with batteries, to serve load with similar reliability to fully dispatchable generation. Crucially, evidence from real systems shows that a clean, flexible portfolio performs well under stress, successfully contributing to meeting reliability requirements. In this context, flexible portfolios are combinations of energy resources that adapt output or consumption to the grid's needs. They pair wind and solar with battery storage, and demand response as variable renewables grow. For example, during the Electric Reliability Council of Texas' (ERCOT's) record-hot summer of 2024, mid-day solar met a large share of load, and evening battery discharge bridged the ramp as the sun set. This maintained reliability and avoided conservation appeals or load shedding even as new demand records were set. This example serves as an operational proof point that firm service can emerge from the combination of renewables, storage and market design rather than from one resource alone.[5]

Globally, the IEA forecasts that renewables (primarily wind, solar and hydro) will be used to meet nearly 50% of the additional electricity demand from data centers through 2030, while natural gas and coal together will supply over 40%.[6] Specifically, in the United States, much of the near-term capacity for announced mega-campuses will be powered by fossil fuels. For example, Meta's $10 billion Louisiana campus is slated to be served by multiple new gas-fired plants, while the planned Stargate AI campus in Abilene, Texas includes a 360 MW on-site natural gas facility.[7] In contrast, markets like Canada, Japan and Brazil are responding to compute demand with renewable-forward infrastructure planning.[8]

With transparent forecasting and grid modernization investments, jurisdictions can meet rising compute loads while bending the supply mix toward low-carbon and flexible portfolios, rather than locking in outcomes by default.[9]

ii. The opportunity

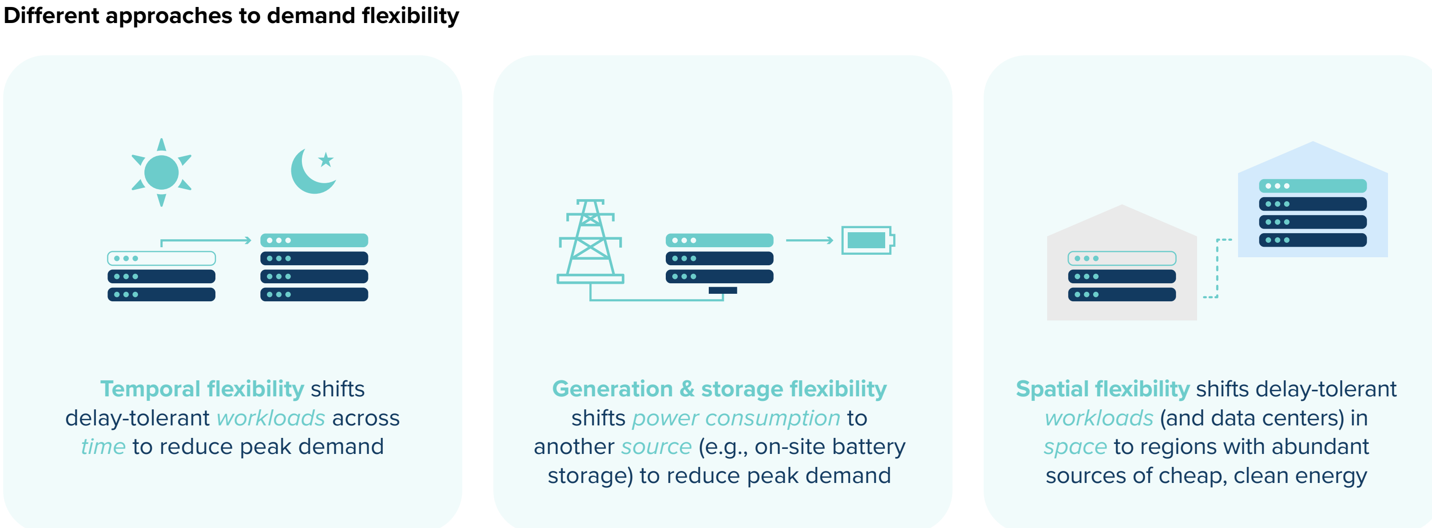
On the ground, developers are bringing data centers online wherever firm power is available—even as they aim to run on renewables or nuclear later—because clean additions and transmission take time, and near-term supply remains largely fossil-based as a result.[6] In the United States, interconnection timelines have stretched from under two years in the 2000s to over four years recently, with a median of five years for projects completed in 2023, creating a real pace gap between compute build-out and a fully clean and reliable grid.[10]

This chapter advances a central thesis: the immense and concentrated power demand of AI data centers could be a primary driver to accelerate the clean energy transition. By proactively aligning AI computational use and infrastructure growth with clean energy deployment, stakeholders can unlock scale, financing and innovation in four key ways:

1. Advanced Market Commitments (AMCs) for Clean-Firm Power: AMCs are purchase agreements intended to de-risk an innovative technology. AMCs differ from power purchase agreements (PPAs), which are long-term commitments to buy power regardless of whether the generation source is innovative. This distinction matters as compute demand is outpacing clean capacity in many regions where availability-based offtake can de-risk and scale a broader portfolio. In this context, clean firm power refers to low-carbon resources that can generate electricity, independent of weather, such as enhanced geothermal, advanced nuclear, carbon capture and storage (CCS). Clean firm power also includes green hydrogen to complement the large renewable PPAs that hyperscalers (i.e., ≥50 MW at a site or ≥100 MW portfolio) already sign.[11] Even so, near-term reliability gaps remain: a planned 1.8 GW AI campus near Cheyenne, Wyoming will use dedicated natural gas with proposed carbon capture to meet immediate needs—underscoring why AMC-style deals that pull forward new clean-firm projects are essential to avoid long-term fossil lock-in.[12]

2. Demand Flexibility as a Grid Asset: Demand flexibility refers to the demand side, such as a data center, adjusting its power use in response to grid needs while maintaining quality of service (QoS). Many data centers today behave as always-on loads (e.g., Google and Meta fleet data show diurnal power swings of only approximately 4% on average).[13] Workloads like AI training and other batch analytics have hours-to-days completion windows. This enables temporal flexibility, shifting non-urgent workloads to another time, such as overnight or during demand-response events (see Figure 4-2). It also enables spatial flexibility, shifting workloads across regions or sites, to follow abundant low-carbon power and avoid local grid stress.[14] For example, Google's carbon-intelligent platform already shifts non-urgent compute to lower-carbon hours and has been used for targeted demand response during grid stress.[15]

Figure 4-2. *Demand flexibility means shifting delay-tolerant workloads in time, shifting power consumption to on-site resources like batteries, and shifting workloads or siting across regions to follow abundant, low-cost clean energy. (Source: Rocky Mountain Institute (RMI))*[14]



3. Siting for Sustainability: Siting refers to where new capacity is built; this is distinct from real-time workload shifting within the current data center footprint. In principle, latency-tolerant workloads (e.g., AI training, batch analytics) can be sited in renewable-rich regions, but in practice, many mega-campuses go where power is available fastest. Thus, near-term supply often tends to be fossil-based. Given multi-year interconnection timelines and constraints on moving data, near-term siting should prioritize grids with headroom and curtailment, pair builds with storage and use availability-based clean offtake to pull forward new clean-firm capacity.

4. AI as the Architect of a Clean Energy Future: Machine learning can optimize grid operations, forecast renewable output and accelerate materials discovery for next-generation energy technologies. Focusing AI applications on opportunities and challenges can reduce costs, environmental stresses and speed to market.[16]

This transformation of aligning the AI compute build-out with clean, reliable power will not occur automatically. It will require knowledge-sharing across firms and utilities by publishing anonymized load-shape datasets and sharing procurement playbooks for flexible, clean capacity. Yet the window of opportunity is now: as trillions of dollars flow into AI infrastructure, strategic intervention can shape the energy future at a global scale.

iii. Global lens

AI data center growth and its energy implications occur in a global context, creating new opportunities. However, most capacity remains centered in the United States, given that approximately 54% of global hyperscale data center capacity and North American build has clustered in Northern Virginia, Phoenix, Dallas–Fort Worth and Atlanta.[17] Northern Virginia alone has approximately 4-5 GW of current data center load, with growth outpacing generation and transmission additions.[18] In parallel, near-term supply for several marquee US campuses is being met with fossil dispatch or planned new gas capacity, such as Project Stargate's on-site ~360 MW gas plant in Abilene, Texas.[19]

Against this backdrop, material projects and policies are advancing outside the United States. In Japan, the Ministry of Economy, Trade and Industry (METI) has launched public-private partnerships to co-develop AI data centers with integrated offshore wind and floating solar infrastructure.[20] In Brazil, clean-power strategies for data centers are being led by renewable PPAs, signed by companies such as Scala, while true hydroelectric co-location exists at Itaipu's utility-campus data center.[21] Malaysia's state of Johor houses the Green Data Center Park, powered by 500 MW of on-site solar, positioning Southeast Asia as a competitive region for clean information technology (IT) growth.[22] In Canada, hydro-rich provinces—especially Québec and British Columbia—offer low-carbon power for AI-class data centers and have recently tightened and streamlined how large new loads connect. Québec's 2025 energy reform created a ministerial authorization regime for large connections, while British Columbia enacted a streamlined permitting law, opened consecutive clean-power procurements, and kept a pause on new crypto-mining interconnections to reserve capacity for strategic industrial customers.[23,24]

Even in power-constrained regions like sub-Saharan Africa, the possibility of AI data center hubs is driving renewed interest in distributed solar and grid modernization. However, challenges related to reliability, financing and workforce capacity remain significant. For example, Microsoft and G42 announced a $1 billion investment in data centers in Kenya designed to run on geothermal: an example of pairing latency-tolerant AI with a firm, low-carbon supply at the source.[25] In Namibia, the European Union's Global Gateway is backing giga-scale renewables for green hydrogen and even green iron—an "energy-at-the-resource" model that data center developers could emulate.[26] These global projects underscore that data center planning and investment are starting to influence national energy planning well beyond the traditional technology hub geographies.

B. Mechanism 1: Advanced Market Commitments as a Catalyst for Clean-Firm Power

i. The power of the purse: how hyperscalers move markets

Large hyperscalers have used PPAs for many years to expand deployment of solar and wind by agreeing to buy power for 10-20 years at predictable terms (long-tenor offtakes) to enable project financing. These hyperscalers are now the largest corporate buyers of renewables.[27] PPA prices generally clear near market rates, so pricing is relatively well established. Sometimes the PPAs also incorporate a green premium, which is the additional cost of a choosing cleaner technology over one that emits more greenhouse gases. More recently, hyperscalers have begun using AMCs, usually tied to technology maturation or performance demonstrations to signal demand for electricity from novel technologies. However, AMCs are more speculative, with limited precedent on price points and timelines. Thus, they function as pioneer instruments to support technology development, rather than commoditized instruments simply intended to secure power.

A visible example is Google's multi-year offtake with Fervo Energy, announced in May 2021, which has allowed Fervo Energy to develop a next-generation geothermal project. Fervo began delivering carbon-free electricity to NV Energy's system in late 2023.[28] Further, after Nevada regulators approved a Clean Transition Tariff in May 2025, Fervo is slated to deliver 115 MW of round-the-clock enhanced geothermal to NV Energy's grid, which NV Energy will sell to Google at a set rate. With deals such as this one, 2024 set a global record for corporate clean-energy deals.[11,29]

In parallel, for advanced geothermal, federal analyses and new resource assessments show a large upside if progress continues. The Department of Energy's (DOE's) Enhanced Geothermal Shot targets $45/MWh by 2035, and the US Geological Survey (USGS) estimates the potential of Enhanced Geothermal Systems (EGS) is approximately 135 GW in the Great Basin upper crust to 6 km.[30] Independent modeling finds the national potential is more than 60-100 GW by mid-century with technology advances.[31] Even with this potential, significant hurdles remain: costs remain high, long-term performance is unproven at scale, seismic risks must be managed and permitting risk still hinders financing. As a result, AMCs like the Google-Fervo deal demonstrate greater bankability.

ii. Case study: nuclear renaissance

Small modular reactors (SMRs) are fission technologies that follow nuclear-plant licensing. They aim to gain cost and schedule control through standardization and factory fabrication. DOE has an active SMR support and solicitation pathway to progress near-term designs. Fusion, by contrast, operates under a different US regulatory path. For example, the Nuclear Regulatory Commission (NRC) placed fusion under the byproduct-materials framework rather than power-reactor rules, potentially enabling faster first-plant timelines.[32] Recent corporate offtakes illustrate both lanes. In the case of fission, Microsoft-Constellation's 20-year PPA to restart Three Mile Island Unit 1 was announced September 20, 2024. Constellation is targeting a 2028 return to service, pending approvals and execution (see Figure 4-3).[33-35] In fusion, Microsoft and Helion announced their agreement on May 10, 2023, aiming for an initial ~50 MW of power generation in 2028. Google and Commonwealth Fusion Systems (CFS) announced their 200 MW offtake on June 30, 2025. This offtake targets early-2030s power from the CFS plant in Virginia.[36,37] Of note, these fusion and restart deals are forward-dated commitments rather than active deliveries, whereas the Google-Fervo project is already online and scaling under the approved tariff.

Figure 4-3. *Three Mile Island Nuclear Power Plant. (Source: American Nuclear Society)*[35]

iii. Scaling pathways and spillover effects beyond data centers

AMCs can launch first-of-a-kind (FOAK) projects. However, the doubling rate of installed capacity is what primarily drives cost declines (via learning-by-doing, supply-chain maturation and finance de-risking). Thus, spillover benefits only materialize when deployments repeat and cumulative capacity grows. In modular clean technologies, meta-analyses consistently find experience-curve cost declines.

Specifically, as synthesized by Rubin et al. (2015) across 11 electricity-supply technologies, these cost declines are typically 10-30% per doubling of installed capacity.[38]

DOE's Liftoff analysis suggests that moving from FOAK to subsequent builds of the same design could reduce overnight capital costs (the up-front cost of construction excluding financing) by roughly 30-40%.[39,40] This is contingent on delivering FOAK projects competently and then achieving approximately 10-20 serial deployments with reactor-to-reactor learning rates of roughly 12-15%. (Note that this is a modeled potential, as opposed to being evidence from standardized SMR fleets to date.)[41] Likewise, major assessments note that SMR economic competitiveness is still unproven at a commercial scale. By contrast, where standardization and serial build have been achieved in large reactor programs, costs and schedules have improved, offering a directional precedent but not a guarantee for SMRs.[42] In this context, structuring hyperscaler AMCs as both multi-unit and multi-year offtakes is best viewed as an enabling mechanism to increase the odds that cumulative deployment doublings occur. For example, pairing AMCs with policy support, finance backstops, liability frameworks and workforce pipelines further improves the chances that cost trajectories bend downward rather than stall.

There is also a clear policy analogue: the United Kingdom's Contracts-for-Difference (CfD), a public AMC, has delivered multi-GW auction rounds at declining strike prices over time, demonstrating how bankable revenue certainty mobilizes private capital at scale. The latest round awarded 9.6 GW across 128 projects, illustrating how demand-certainty mechanisms can push technologies down their cost curves.[43] For AI buyers, the lesson is strategic: catalyzing spillovers today helps secure scarce, reliable clean power near key load clusters and drive long-run costs down as technologies learn. Many hyperscalers even have explicit carbon-free energy (CFE) commitments, such as Google which aims to run on 24/7 carbon-free energy by 2030.[44] Unlike most large loads, hyperscalers are already helping co-design utility structures, such as Nevada's Clean Transition Tariff that enables the Google and Fervo geothermal scale-up (as discussed above).[45,46]

C. Mechanism 2: The Flexible Data Center: Transforming Power Grids

i. From rigid consumer to grid stabilizer

The next wave of owner-operated hyperscale clouds already modulates load in practice. Google's carbon-aware computing shifts movable batch work across data centers based on hourly carbon-free energy forecasts, and its "elastic training" and checkpointing tools show how some training jobs can pause and resume without losing progress.[47] Facilities can also flex on the non-IT side. For example, Microsoft's Dublin campus is using data center uninterruptible power supply (UPS) batteries for grid frequency services, demonstrating a pathway to make sites grid-responsive.[48] By contrast, most co-location operators—companies that lease space, power, and cooling to multiple tenants—face a split-incentive. The operators run the facility, but tenants control their own servers and workloads, which limits operator-led orchestration beyond facility-side measures.[49]

Duke University's Nicholas Institute estimates that US balancing authorities could integrate nearly 100 GW of new flexible large loads using existing capacity if those loads curtailed briefly during the most stressed hours. Specifically, these loads would have to curtail an average of 0.5% of annual uptime and roughly 2 hours on average during events.[50] This finding speaks to available headroom in today's system if new loads are flexible. Rather than claiming that more renewables must be built to serve these large flexible loads, the study shows modest flexibility can reduce the need for near-term capacity additions. Further, the same flexibility also improves renewable utilization when wind and solar are added by cutting curtailment when renewable resources are highly available and still trimming load during tight hours.[50]

Beyond Google and Microsoft, operators such as Digital Realty and Equinix in Dublin, Aeven in Denmark, and Basefarm in Norway have already deployed grid-interactive UPS or demand-response systems. These systems provide frequency regulation, which means quickly adjusting power to keep the grid stable, while maintaining quality of service (QoS) guarantees. [51-54]

On the modeling side, the National Renewable Energy Laboratory's (NREL's) Electrification Futures work finds demand-side flexibility cuts operating costs and eases the need for variable renewable energy (VRE) integration in highly electrified systems. However, its flexible-load assumptions are drawn mainly from electric vehicle (EV) charging and flexible industrial end uses.[55] This insight is still relevant to data centers because they can also match this profile when owners time-shift batch training or non-urgent inference and when sites use UPS batteries for grid services.[56] Of note, a portion of modern data centers now reach the 100+ MW class, comparable to classic single-site giants such as aluminum smelters (hundreds of MW), electric arc furnaces (EAF) and electrified steam crackers (hundreds of MW). However, data centers are increasingly software-defined, fast-ramping, and already equipped with battery and UPS assets that can provide services.[57-59] These features make owner-operated clouds uniquely positioned to catalyze clean energy and flexibility at scale, even though multi-tenant co-location sites will remain less controllable at the workload level.

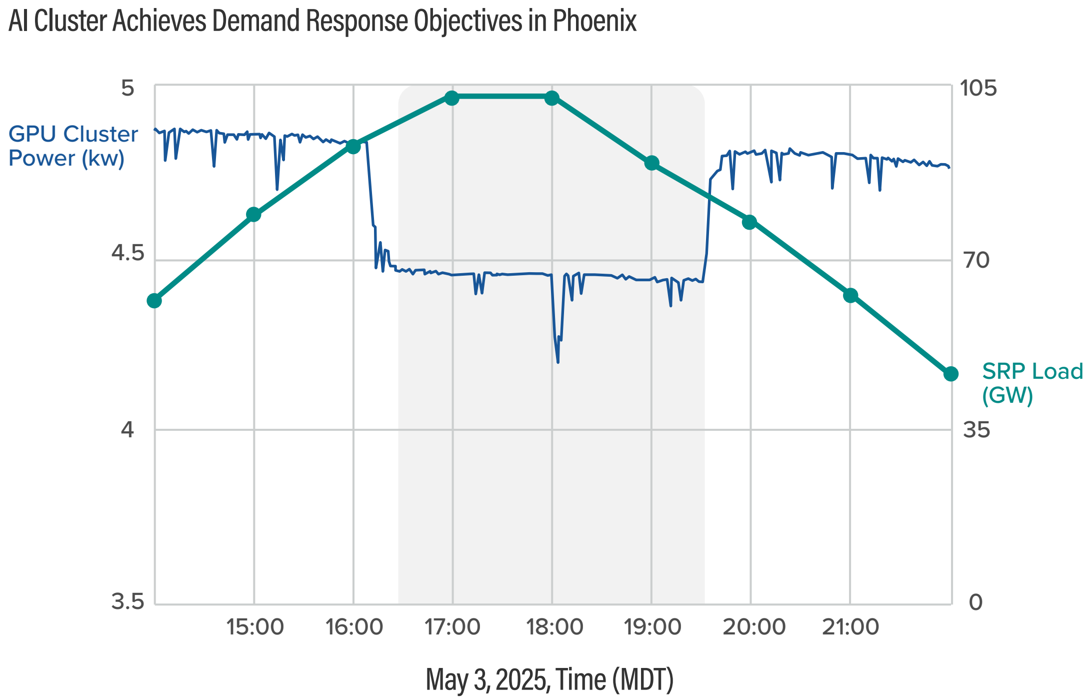
ii. Innovations in action

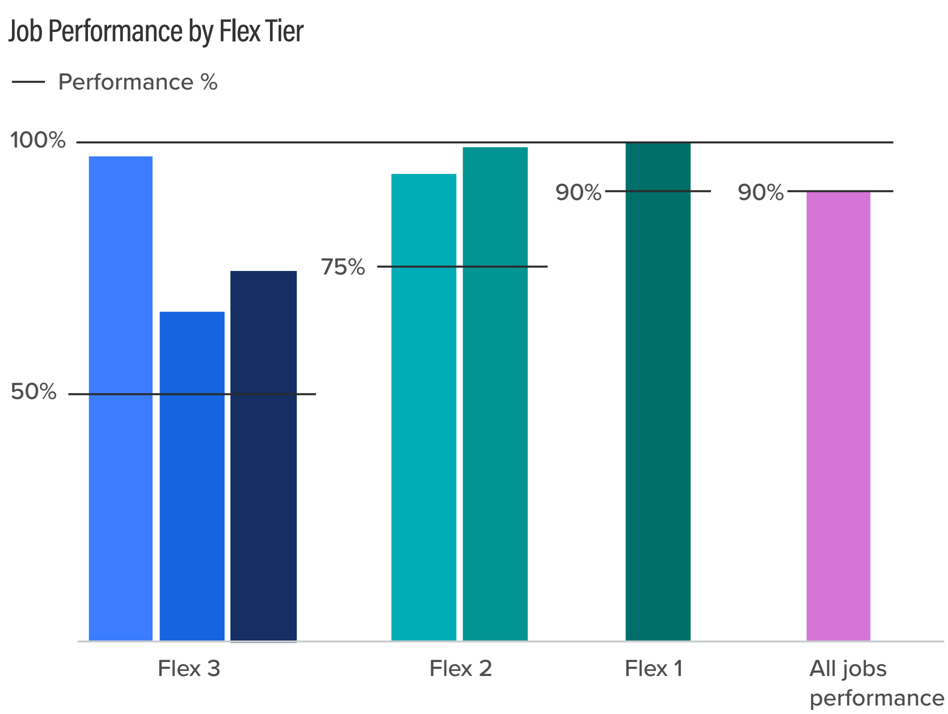
First movers are already showing the playbook. For example, Google's carbon-aware computing shifts non-urgent workloads across hours and, where possible, across locations based on hourly carbon-free energy forecasts.[60] However, only a subset of organizations can do this in practice: owner-operated hyperscale clouds with global fleets and batchable jobs can orchestrate load, whereas most multi-tenant co-location sites cannot directly control tenant workloads.[61] In parallel, new platforms are emerging to translate that flexibility into operational grid support. For example, in May 2025, Emerald AI[a] (backed by NVIDIA's venture arm) conducted a field experiment with Salt River Project (SRP) in Phoenix and sustained a 25% power reduction for three hours on a live AI graphics processing unit (GPU) cluster (see Figure 4-4). In the Phoenix demonstration, AI jobs were tagged by how much performance they can temporarily give up during grid stress: Flex 1 (up to 10% performance reduction), Flex 2 (up to 25%) and Flex 3 (up to 50%). Given coincident-peak (CP) and capacity charges hinge on only a few stressed hours, a 25% curtailment capability typically yields about $40,000 to $70,000 per MW per year in ERCOT's 4 Coincident Peak (4CP), and about $17,000 to $160,000 per MW per year in PJM's 5 Coincident Peak (5CP). Therefore, a 100 MW site that can flex 25 MW often sees low seven figure annual bill reductions.[62-65]

This three-hour test translates to 24/7/365 operations because the peaks are brief and forecastable, so software can pre-stage checkpoints and apply power caps only during those intervals while running normally the rest of the year.[66] Emerald AI's Phoenix demonstration serves as a significant proof point for grid-resilience and cost reduction by shaving critical peaks without violating compute performance targets.[67] Together, these approaches map to what IEA expects the parts of the sector that are capable of shifting load to deliver this decade: flexible demand that helps integrate more renewables while easing immediate supply-side pressures in hot-spot regions.[6]

a Co-author Varun Sivaram is the founder and CEO of Emerald AI. Co-authors Ayse Coskun and Swasti Jain are affiliated with Emerald AI, as well.

Figure 4-4. *Emerald AI's Phoenix demonstration shows the power reduction curve and applications with different flexible service-level agreements (SLAs) meeting their performance requirements. (Source: ArXiv)[68]*

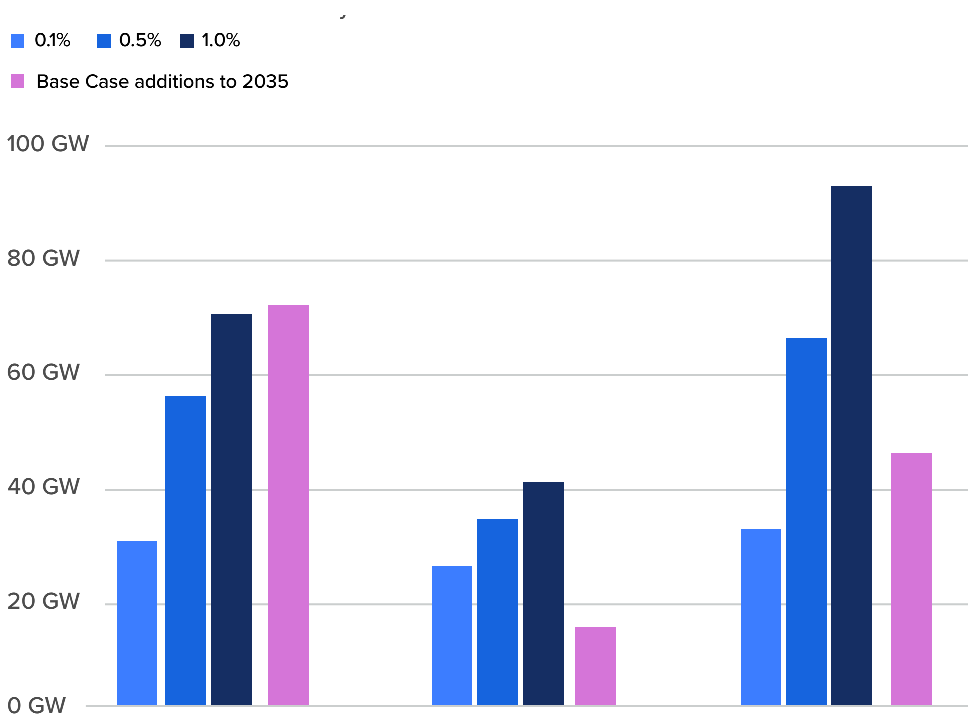




iii. A paradigm shift for grid management

As shown in Figure 4-5 below, the IEA expects flexible demand to absorb essentially all (~100%) projected data center additions through 2035 if operators are flexible for just ~0.1-1% of hours, complementing conventional flexibility from natural gas where it remains system-critical.[69] System context matters: in Japan, for example, tight reserve margins (a small amount of extra generation capacity), the division of the electric grid into two regions with different frequencies (50 Hz in the east and 60 Hz in the west) and policy design mean large, routine industrial curtailment is more constrained. In this case, Japanese market operators emphasize securing decarbonized supply, while demand response is being used primarily as a resource during tight hours in capacity markets.[70] In Japan, demand response participation is concentrated among large commercial and industrial (C&I) customers, such as high-voltage factories, campuses and big buildings, typically enrolled through aggregators certified by Japan's Energy Ministry. These aggregators also bundle behind-the-meter (BTM) batteries, combined heat and power, EVs and building controls.[71] This approach succeeds because there are clear, paying routes to market in the capacity market and the supply-demand balancing market. In 2022, for example, about 2.3 GW of demand response won roughly 60% of the "Power Source I" auction.[72]

Figure 4-5. *Data center capacity additions through 2035 and feasible integration into the current electricity system under different flexibility cases. (Source: IEA)*[69]



Current electricity systems can already integrate all data centre additions to 2035 if a mix of back up activation and workload management reduces grid demand 1% of the time.

iv. Enhanced renewable integration

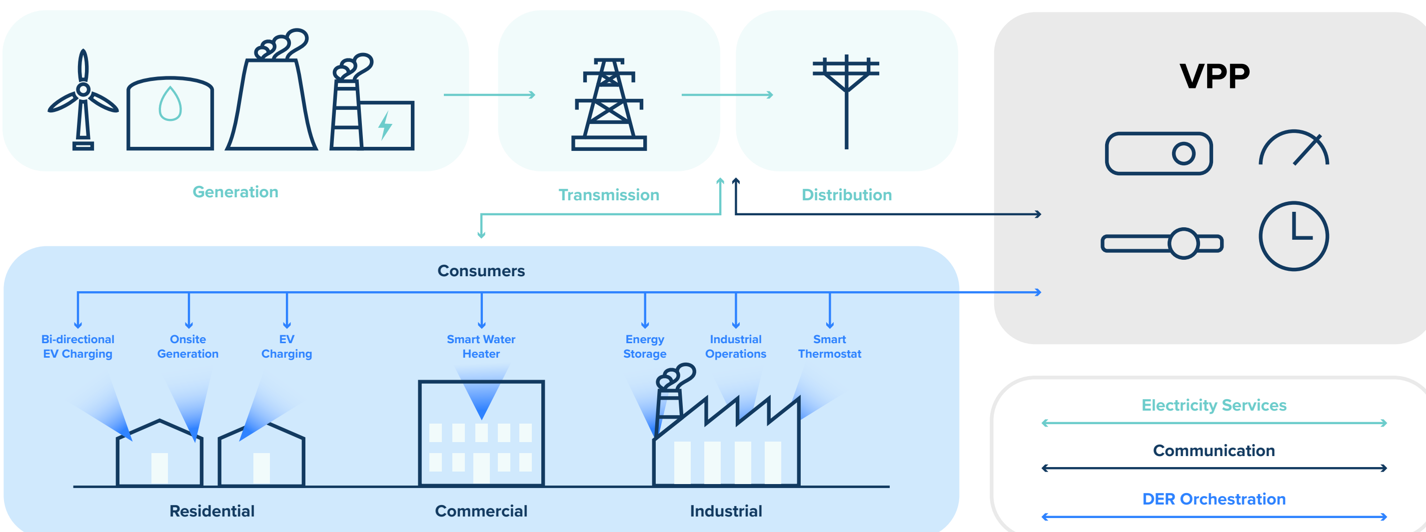
Flexible demand enhances renewable integration by absorbing surplus solar and wind when it is plentiful and standing down during constrained hours. In doing so, flexible demand reduces the need for curtailment, enables deeper renewable penetration, and avoids overbuilding supply for rare peaks.[73] In California ISO's (Independent System Operator) grid, renewable curtailment reached the Terawatt-hour (TWh) scale in 2024, concentrated in high solar shoulder months: energy that shiftable loads (including data centers) could take up, especially when paired with short-duration storage and carbon-aware scheduling.[74] The IEA quantified the system impact with the US example, showing only ~0.3% of annual grid electricity needing active management during short 3-5 hour events.[69]

Recent research from the Association for Computing Machinery (ACM) Journal and ETH Zurich (Federal Institute of Technology Zurich) shows that QoS-aware scheduling and power-capping can deliver demand response and even regulation services without breaching service levels and that carbon-aware orchestration across sites is practical at scale.[75-77] Field evidence now complements this literature: Emerald AI's May 2025 demonstration on a 256-GPU cluster in a hyperscale cloud reduced power by 25% for three hours using a software-only approach while maintaining QoS.[68] On the value side, IEA modeling indicates that modest, time-bound flexibility reduces the need for curtailment and defers some peak-driven supply build, consistent with system-level findings.[69] The forward question is therefore not whether natural gas can provide flexibility, since it does, but what blend of flexible demand, storage and dispatchable supply delivers the best reliability and cost profile in each region.[78]

v. Grid resilience and cost reduction

When dispatched at the right times and places, flexible load can downsize networks and peaking investments, improving resilience and affordability. As shown in Figure 4-6 below, DOE defines virtual power plants (VPPs) as aggregations of distributed energy resources (DERs), such as rooftop solar with customer-sited batteries, EVs, and smart buildings that can balance electricity supply and demand and deliver utility-grade services like a power plant.[79] This chapter treats data center flexibility as virtual capacity that is compatible with virtual power plants, so DOE's VPP analysis is a useful proxy for system value. DOE estimates that tripling VPP capacity to approximately 80-160 GW by 2030 would reduce overall grid costs by roughly $10 billion per year by avoiding use of peakers (power-generating facilities that run during highest-demand periods), deferring grid investments and lowering operating costs.[79] In the United Kingdom, the National Energy System Operator's (NESO's) Demand Flexibility Service delivered 3.3 GWh of verified peak-time reductions in winter 2022 and 3917.7 MWh across 44 events in winter 2024 under its merit-based design.[80,81] As a result, NESO demonstrates a practical route to reliability at lower balancing cost.[82] At the distribution edge, National Grid—the investor-owned utility that owns the high-voltage transmission network in England and Wales—reports £80 million of investment deferral via 17 GWh of procured flexibility and greater than 70,000 registered flexible assets.[83] This is a different operational model than data centers. Aggregating thousands of devices is complex, while one or two large data centers can deliver the same order of magnitude of response with fewer parties and direct telemetry. This serves as evidence that market-based flexibility can substitute for some wires-and-steel upgrades. Although results will vary by system design and siting, the direction is consistent: well-designed flexibility reduces the need for both curtailment and high-cost build-outs.

Figure 4-6. *Virtual power plants (VPPs) aggregate distributed energy resources, such as rooftop solar with customer-sited batteries, electric vehicles (EVs) and chargers, smart buildings and equipment and their controls, and flexible commercial and industrial (C&I) loads. (Source: Department of Energy Loan Programs Office (DOE LPO))*[79]



vi. Accelerated AI innovation (why flexibility can be a win-win)

From the grid's perspective, flexible data centers shave critical peaks, reduce the need for curtailment and defer costly reinforcements—improving reliability and affordability. Likewise, from the operator's perspective, flexibility can be a win-win where accelerated interconnection is a binding constraint. In other words, transmission operators may advance interconnection dates or offer ramped and conditional interconnections when projects commit to flexible operating modes. Not every market faces the same pressure, but the challenge is particularly acute in the United States, where the median time from interconnection request to operation reached about five years for projects built in 2023.[10] NESO has reached a point where it can bring forward 20 GW of connections in approximately 4 years via queue reforms and flexible arrangements.[84] In the United States, Federal Energy Regulatory Commission (FERC) Order 2023 is streamlining how new projects connect to the grid by studying many requests together in “clusters” and imposing withdrawal penalties to shorten queues. In parallel, some regions are trying “flexible interconnection,” through which very large new loads can connect sooner if they agree to operate within limits and share real-time data. For example, in Texas, ERCOT’s large-load process requires telemetry and allows voluntary curtailment during grid stress.[63] On the customer side, making this work usually takes three simple building blocks:

1. Power caps: Software limits that quickly lower server or site power for short periods.

2. Thermal buffers: Modest on-site chilled water or ice storage lets facilities store cooling when supply is plentiful and draw it down when the grid is constrained.[85]

3. Real-time telemetry tied to Service-Level Agreements (SLAs): Continuous, machine-readable telemetry linked to clear SLAs verifies the response for the grid while ensuring applications still meet their performance targets.

D. Mechanism 3: Siting for Sustainability

i. Siting for clean, reliable compute

Training and many batch inference jobs are tolerant to network delay, so operators can place new capacity in regions with strong renewable resources and then schedule work to follow local peaks. In the United States, the Great Plains offers exceptional wind potential and the desert in the Southwest offers high solar irradiance, which makes these regions natural candidates for “compute-next-to-resource” strategies.[86] Geothermal is also expanding beyond traditional fields as enhanced geothermal systems mature, widening the map for firm clean supply.[87] Remote siting often runs into limited grid headroom, which is why some projects consider BTM generation at the campus. This option remains the exception because most hyperscalers prefer utility or third-party power under bankable tariffs or PPAs rather than owning plants at scale. Recent experience in Pennsylvania shows how co-located or BTM nuclear supply raised regulatory and cost-allocation questions and ultimately shifted toward a standard grid-connected structure.[88,89] This preference aligns with the broader corporate market’s use of utility green tariffs and long-term contracts to add new clean capacity.[89]

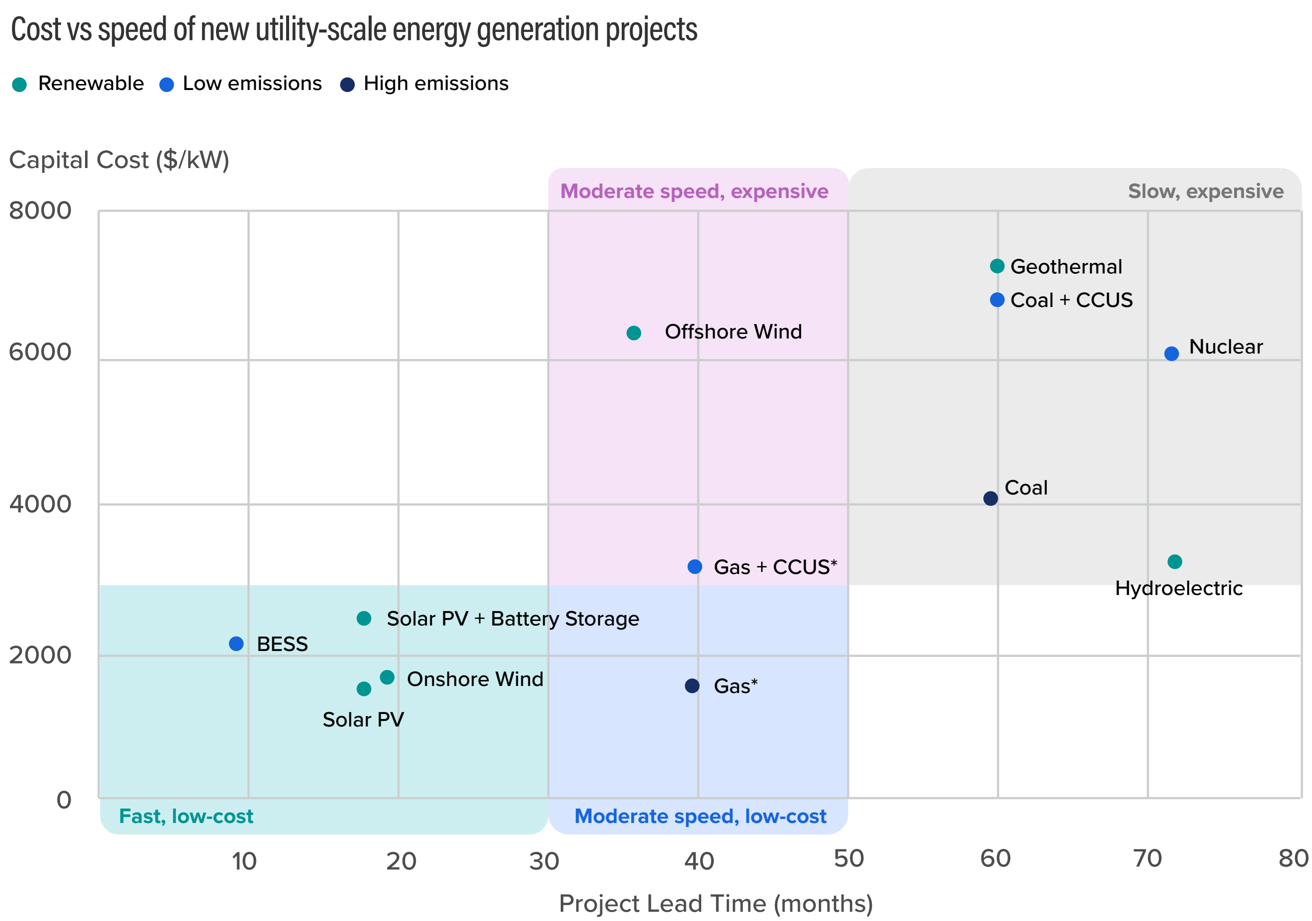
When the goal is to accelerate new clean energy, siting choices work best when paired with enabling mechanisms that bring additional projects onto the grid. One pathway is utility tariffs designed for hourly matched portfolios, such as Nevada's approved structure that enables Google to source new geothermal and storage through NV Energy.[87,90] A second pathway is risk-reduction finance for firm and long-duration storage so remote renewable hubs can deliver around the clock. Federal loan guarantees have begun backing multi-hour storage that firms renewables and lowers system costs. Finally, grid modernization investments, such as advanced conductors, dynamic line ratings and new high-capacity corridors, expand the ability of resource-rich zones to host large flexible loads.[91-93]

ii. Catalyzing new clean energy development

Large AI data center projects, if not planned carefully, can drive up costs for other ratepayers and strain grid stability. However, relocatable data center loads offer a unique opportunity to accelerate low-carbon power deployment when aligned with renewable energy growth. By being flexible on location, these energy-hungry data centers can be placed in resource-rich regions, such as deserts or windy plains, where new solar and wind farms can be built to supply them.[86] This strategic siting turns data centers into demand catalysts for additional clean generation capacity. In practice, a relocatable AI campus can serve as an anchor offtaker that makes a new renewable project financially viable, bringing extra wind or solar online faster than traditional grid planning might.[14] One concrete example is Microsoft's New Zealand cloud region. Microsoft signed a 10-year agreement tied to Contact Energy's new 51.4 MW Te Huka III geothermal plant, and Contact explicitly says the contract supported its investment decision to build the project.[104] The plant began delivering to the grid in October 2024, illustrating how a relocatable cloud load can underwrite firm clean capacity that serves both the data center and the broader system. In short, when AI data centers are free to move to the power (instead of always bringing power to the load), they can help drive an accelerated build-out of wind, solar and other low-carbon resources.

Critically, realizing these benefits depends on project design and grid integration. Reaping community-wide gains requires that the clean electricity procured for a data center also strengthen the public grid. For example, using utility green tariffs or CFDs ensures that new renewable capacity is not just behind-the-fence—on the customer side of the meter, serving only the facility—but also contributes energy to everyone over time.[94,95] As Figure 4-7 shows, relocatable data centers should pair any new renewables with energy storage and embrace flexible operations (e.g., drawing less power at grid peak times or using on-site batteries), so their presence reduces the need for curtailment and eases grid congestion instead of exacerbating it.[14] This strategy defers the need for certain network upgrades and helps ensure reliability is maintained with renewables. Only when power procurement, interconnection terms and community agreements are structured for system-wide value will a relocatable load accelerate the clean energy transition.[96] In practice, this means the data center's clean power investments should lower overall system costs over time (by adding new capacity and reducing fuel use), and its operators should coordinate with utilities to avoid saddling other customers with infrastructure expenses. When done right, a flexible, renewables-driven data center can act as a springboard for more clean energy on the grid to deliver not just low-carbon electricity for its own operations, but cheaper and more reliable power for the surrounding region as well.[14]

Figure 4-7. *Solar and onshore wind (especially when paired with batteries) are the fastest, lowest-cost new capacity, while coal/CCUS (carbon capture, utilization and storage), hydroelectric, geothermal and offshore wind are slower and more expensive. (Source: Rocky Mountain Institute (RMI))*[14]

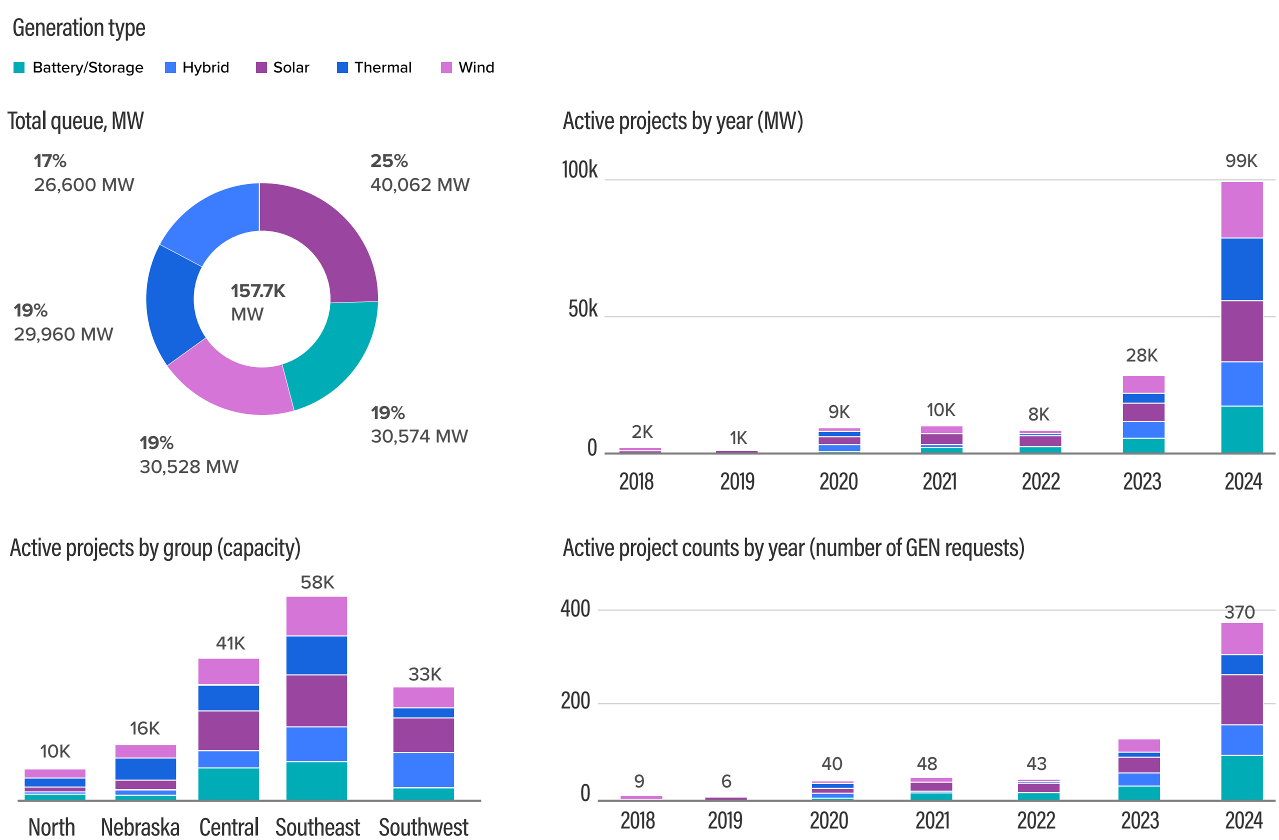


E. Mechanism 4: AI as the Architect of a Clean Energy Future

i. Grid optimization

One near-term, high-impact role for AI is speeding interconnection and transmission studies: an "offline" use that directly unlocks renewable projects. As Figure 4-8 shows, Southwest Power Pool's (SPP's) generator-interconnection (GI) queue is facing significant backlogs. Therefore, SPP is partnering with Hitachi Energy to apply AI to generator-interconnection queue triage and reliability assessments so planners can screen power-flow cases, prioritize upgrades and shorten study timelines.[97] Hitachi and SPP will use an integrated NVIDIA-based compute and AI platform to automate processes, add predictive analytics and pilot AI-augmented simulation, with systems acceleration and data-management optimization due by December 2025. These tools complement both FERC Order 2023 queue reforms and DOE's Grid Modernization Initiative by automating contingency ranking, hosting-capacity mapping, and topology and stability screening. All these partnerships help connect renewables and storage sooner, while reducing the need for curtailment and fuel burn.[97]

Figure 4-8. *Southwest Power Pool's (SPP's) generator-interconnection (GI) queue lists 659 projects totaling ~157.7 GW, with ~380 currently in study and clusters still open from 2018. Source: RTO Insider[97]*



ii. Materials discovery and design

AI is compressing the discovery-to-device cycle for clean energy materials. For example, instead of testing materials one by one, AI learns from big materials databases and predicts which compositions and crystal structures are most promising. It can sift through tens of millions of options and surface a few thousand good candidates for battery electrodes, solid electrolytes, photovoltaic absorbers and catalysts, among other materials.[98] These models are coupled with high-throughput first-principles calculations and open data infrastructures, such as The Materials Project, to validate stability and properties and to hand off the best leads to synthesis.[99] On the lab side, AI tools mine the literature for synthesis recipes and processing conditions, feeding autonomous platforms that can make and test candidates rapidly: shortening iteration times for solar, battery and CO_2-conversion technologies while lowering cost-to-proof.[100]

iii. Improved forecasting

AI is also closing the gap between weather hazards and power-system impacts by improving both the speed and the skill of forecasts that feed unit commitment, dispatch, storage scheduling and outage preparedness. Thus, current versions of AI are already building on the “hazard-to-impact” framing highlighted in last year’s ICEF Roadmap chapter on extreme-weather response, which emphasized more actionable signals for heat waves, wind ramps and atmospheric rivers.[101] Today’s AI models deliver state-of-the-art medium-range skill at a fraction of the compute and energy cost of traditional physics-based forecasts on supercomputers. The European Centre for Medium-Range Weather Forecast’s (ECMRWF) now operates an Artificial Intelligence Forecasting System (AIFS) in parallel with its traditional physics-based Integrated Forecasting System (IFS). Peer-reviewed AI models, such as GraphCast demonstrate strong 10-day global forecast performance, giving grid operators more frequent, higher-skill inputs for renewable scheduling and contingency analysis.[102] The AI outputs are trained and calibrated against authoritative observational datasets from the National Oceanic and Atmospheric Administration (NOAA) and other government weather services, which remain the backbone for assimilation and verification.[103]

Recent American Meteorological Society (AMS) work shows that machine-learning corrections to models can dramatically reduce wind-forecast error, which directly lowers reserve needs, and the need for curtailment and balancing costs for renewable-heavy grids. Taken together, these three AI applications are beginning to reinforce one another. AI for interconnection and other offline studies helps planners clear queues and move clean projects to "shovel-ready" status faster. AI-enhanced operational forecasts improve day-ahead and intraday scheduling of wind, solar and storage. Lastly, major weather centers are now deploying probabilistic AI ensembles. As a result, this combination creates a virtuous cycle that speeds grid build-out, strengthens extreme event readiness, and enables deeper renewable penetration and more reliable service.[103]

F. Recommendations

1. *Utilities and independent power producers (IPP) should:*

    - ***Deploy advanced control tools** to accelerate interconnection and grid studies and to operate flexible portfolios. These tools include model-predictive control, enhanced forecasting and, where appropriate, AI.*

    - ***Adopt staged or ramped interconnections for large loads** (within standard planning cycles) and require telemetry and fast power-capping from data centers and consider on-site storage to provide demand response and regulation while maintaining service-level objectives.*

    - *Use these tools to **prioritize non-wires alternatives** and to reduce curtailment in renewable-rich zones.*

2. *Electricity regulators should:*

    - ***Establish clear 24/7 carbon-free energy procurement pathways** that treat storage and clean-firm resources as first-class options alongside renewables.*

    - *Enable **advanced market commitments (AMCs)** that allow multi-buyer participation, recognize hourly matching, and credit verifiable flexible-load performance.*

3. *National governments, regulators, and utilities should:*

    - ***Expand targeted public-private risk-sharing** to lower the cost of* *firm, low-carbon supply while keeping rates affordable amid rising public concern about electricity bill impacts from data center–driven capacity additions.*

- ***Pair corporate offtake with loan guarantees**, liability and fuel frameworks (where relevant), and long-duration storage demonstrations.*
- ***Adapt CfD-style mechanisms to clean-firm resources and storage** so FOAK projects are followed by repeat builds of the same design.*

4. *Large data center operators with load flexibility, hyperscalers and procurement authorities should:*
   - ***Commit to portfolio-based, 24/7 carbon-free procurement** that include renewables, storage and clean-firm resources where available.*
   - ***Publish transparent hourly performance** and adopt grid-supportive operating modes, such as fast power caps and brief curtail on-signal, to **unlock faster interconnection and lower system costs**.*
   - ***Prioritize deliverable power** to the public grid (when siting in resource-rich regions), rather than exclusively behind-the-fence supply.*
5. *National and local governments should:*
   - ***Link siting incentives for new AI campuses to system and community value.***
   - *Require additional deliverable **clean capacity, storage co-procurement and community benefit plans** that include workforce pipelines, water stewardship and shared transmission upgrades.*
6. *National governments and utilities, including public power and transmission owners, should **invest in grid modernization**, including advanced transmission, system visibility and congestion management, so resource-rich zones can host large and flexible loads without unnecessary overbuild. Regulators should authorize these investments, set incentives and ensure timely cost recovery.*
7. *Academic experts and system operators should **advance operations-ready forecasting for clean grids**. Priorities include post-processing and downscaling of weather models (including via AI) for wind and solar, probabilistic products that feed unit commitment and storage scheduling, and open benchmarks that **connect forecast improvements to avoided reserves, reduced curtailment, and emissions reductions**.*

G. References

1. International Energy Agency (IEA). Chapter 3: Energy Demand from AI in Energy and AI (IEA, Paris, France, 2025, https://www.iea.org/reports/energy-and-ai/energy-demand-from-ai).

2. International Energy Agency (IEA). Energy and AI Observatory; IEA, Paris, France, https://www.iea.org/data-and-statistics/data-tools/energy-and-ai-observatory?tab=Energy+for+AI (Accessed August 2025 (Last updated June 2025).

3. Cheryl Schweizer. Grant PUD places limits on electrical demand from data centers; Columbia Basin Herald, Moses Lake, Washington, https://columbiabasinherald.com/news/2025/mar/31/grant-pud-places-limits-on-electrical-demand-from-data-centers/ (2025).

4. Office of Energy Statistics Staff. Solar, battery storage to lead new U.S. generating capacity additions in 2025; US Energy Information Administration (EIA), Washington, DC, https://www.eia.gov/todayinenergy/detail.php?id=64586 (2025).

5. Garrett Golding. Solar, battery capacity saved the Texas grid last summer; an uncertain future awaits; Federal Reserve Bank of Dallas, Dallas, Texas, https://www.dallasfed.org/research/economics/2025/0114 (2025).

6. International Energy Agency (IEA). Chapter 4: Energy Supply for AI in Energy and AI (IEA, Paris, France, 2025, https://www.iea.org/reports/energy-and-ai/energy-supply-for-ai).

7. Kristin Toussaint. Meta is building a new data center in Louisiana—and this Senate committee wants to know why it's being powered by gas (exclusive); Fast Company, New York, New York, https://www.fastcompany.com/91334846/meta-is-building-a-new-data-center-in-louisiana-and-this-senate-committee-wants-to-know-why-its-being-powered-by-gas-exclusive (2025).

8. Julian Cruz. Brazil's Data Center Boom: A Renewable-Fueled Infrastructure Play in a Strategic Market; AInvest, New York, New York, https://www.ainvest.com/news/brazil-data-center-boom-renewable-fueled-infrastructure-play-strategic-market-2505/ (2025).

9. Arman Shehabi, Alex Newkirk, Sarah J Smith, Alex Hubbard, Nuoa Lei, Md Abu Bakar Siddik, Billie Holecek, Jonathan Koomey, Eric Masanet & Dale Sartor. 2024 United States Data Center Energy Usage Report (Report #: LBNL-2001637); Lawrence Berkeley National Laboratory, https://doi.org/10.71468/P1WC7Q; Retrieved from https://escholarship.org/uc/item/32d6m0d1 (2024).

10. Joseph Rand, Nick Manderlink, Will Gorman, Ryan H Wiser, Joachim Seel, Julie Mulvaney Kemp, Seongeun Jeong & Fritz Kahrl. Queued Up: 2024 Edition, Characteristics of Power Plants Seeking Transmission Interconnection As of the End of 2023; Lawrence Berkeley National Laboratory Energy Analysis and Environmental Impacts Division, Energy Markets and Policy Department Berkeley, California, https://emp.lbl.gov/sites/default/files/2024-04/Queued%20Up%202024%20Edition_1.pdf (2024).

11. Tony Lenoir. Corporate PPA leaderboard – Microsoft leap cuts into Amazon lead; S&P Global, New York, New York, https://www.spglobal.com/market-intelligence/en/news-insights/research/corporate-ppa-leaderboard-microsoft-leap-cuts-into-amazon-lead (2025).

12. Mead Gruver & Matt O'Brien. Cheyenne to host massive AI data center using more electricity than all Wyoming homes combined; The Associated Press (AP), New York, New York, https://apnews.com/article/ai-artificial-intelligence-data-center-electricity-wyoming-cheyenne-44da7974e2d942acd8bf003ebe2e855a (2025).

13. Bilge Acun, Benjamin Lee, Fiodar Kazhamiaka, Kiwan Maeng, Manoj Chakkaravarthy, Udit Gupta, David Brooks & Carole-Jean Wu. Carbon Explorer: A Holistic Approach for Designing Carbon Aware Datacenters. arXiv:2201.10036 (2022). https://doi.org/10.48550/arXiv.2201.10036.

14. Yuki Numata, Alexandra Gorin, Laurens Speelman, Lauren Shwisberg & Chiara Gulli. Fast, Flexible Solutions for Data Centers; Rocky Mountain Institute (RMI), Basalt, Colorado, https://rmi.org/fast-flexible-solutions-for-data-centers/ (2025).

15. Emma Penrod. Google taps 'carbon-intelligent' computing platform to help maintain grid reliability in power crises; Utility Dive (Informa TechTarget), Newton, Massachusetts, https://www.utilitydive.com/news/google-carbon-intelligent-computing-platform-system-reliability-demand-response-grid-emergency/698958/ (2023).

16. David Sandalow, Colin McCormick, Alp Kucukelbir, Julio Friedmann, Michal Nachmany, Hoesung Lee, Alice Hill, Daniel Loehr, Matthew Wald, Antoine Halff, Ruben Glatt, Philippe Benoit, Kevin Karl et al. Artificial Intelligence for Climate Change Mitigation Roadmap (Second Edition) (ICEF Innovation Roadmap Project, November 2024); https://doi.org/10.7916/2j4p-nw61 (2024).

17. Synergy Research Group. Hyperscale Data Center Count Hits 1,136; Average Size Increases; US Accounts for 54% of Total Capacity; Reno, Nevada, https://www.srgresearch.com/articles/hyperscale-data-center-count-hits-1136-average-size-increases-us-accounts-for-54-of-total-capacity (2025).

18. Matt Busse. Data centers driving 'immense increase' in Virginia's energy demand, report says; Cardinal News, Roanoke, Virginia, https://cardinalnews.org/2024/12/10/data-centers-driving-immense-increase-in-virginia-energy-demand-report-says/ (2024).

19. Lancium. Lancium and the Stargate Project in Abilene, TX: Bringing Hyperscale Campuses to Texas; Energy Systems Integration Group (ESIG), Reston, Virginia, https://www.esig.energy/wp-content/uploads/2025/05/ESIG_LLTF_PresentationLancium.pdf (2025).

20. Trade and Industry (METI) and the Ministry of Internal Affairs and Communications (MIC) Ministry of Economy. Public-Private Advisory Council on Watt-Bit Collaboration to Launch; METI, Tokyo, Japan, https://www.meti.go.jp/english/press/2025/0318_001.html (2025).

21. Scala Data Centers. Scala Data Centers and Serena Announce the Largest Renewable Energy Partnership in Latin America's Data Center Industry; São Paulo, Brazil, https://scaladatacenters.com/en/scala-data-centers-and-serena-announce-the-largest-renewable-energy-partnership-in-latin-americas-data-center-industry/

22. Azeem Azhar. Big Tech Will Scour the Globe in Its Search for Cheap Energy; WIRED Magazine, San Francisco, California, https://www.wired.com/story/big-tech-data-centers-cheap-energy/ (2024).

23. Hydro-Québec Stratégies et Finances. Action Plan 2035: Towards a Decarbonized and Prosperous Québec (English Version); Hydro-Québec, Montreal, Québec, Canada, https://www.hydroquebec.com/data/a-propos/pdf/action-plan-2035.pdf (2023).

24. Dominique Rolland, Patricia Larouche, Shannon Corbeil Bathurst & Mariane Gagné. Québec Government Adopts Bill 69 to Modernize Energy Legislation; Stikeman Elliott LLP Toronto, Ontario, Canada, https://stikeman.com/en-ca/kh/canadian-energy-law/quebec-government-adopts-bill-69-to-modernize-energy-legislation (2025).

25. Darryl Willis. Microsoft partners celebrate AI innovation; Microsoft, Redmond, Washington, https://www.microsoft.com/en-us/industry/blog/energy-and-resources/2024/06/27/microsoft-partners-celebrate-ai-innovation (2024).

26. Directorate-General for International Partnerships (DG-INTPA). Global Gateway: Namibia becomes a pioneer for Africa's green transition; European Commission DG-INTPA, Brussels, Belgium, https://international-partnerships.ec.europa.eu/news-and-events/news/global-gateway-namibia-becomes-pioneer-africas-green-transition-2025-04-11_en (2025).

27. Lorenzo Moavero Milanesi, Tjark Freundt, Yuito Yamada, Fridolin Pflugmann & Marc Ludwig. How hyperscalers are fueling the race for 24/7 clean power; McKinsey & Company, New York, New York, https://www.mckinsey.com/industries/electric-power-and-natural-gas/our-insights/how-hyperscalers-are-fueling-the-race-for-24-7-clean-power (2024).

28. Michael Terrell. With new geothermal project, it's full steam ahead for 24/7 carbon-free energy; Google Cloud, Mountain View, California, https://cloud.google.com/blog/products/infrastructure/google-fervo-geothermal-project-creates-carbon-free-energy (2021).

29. Emma Penrod. NV Energy seeks new tariff to supply Google with 24/7 power from Fervo geothermal plant; Utility Dive (Informa TechTarget), Newton, Massachusetts, https://www.utilitydive.com/news/google-fervo-nv-energy-nevada-puc-clean-energy-tariff/719472/ (2024).

30. Erick R. Burns, Colin Williams & Jacob DeAngelo. Enhanced geothermal systems electric-resource assessment for the Great Basin, southwestern United States. Report No. 2025-3027, 4 (US Geological Survey (USGS); Reston, VA, 2025), https://doi.org/10.3133/fs20253027.

31. Chen Chen, Daniel Merino-Garcia, Timothy D. G. H. Lines & Daniel S. Cohan. Geothermal power generation potential in the United States by 2050. Environmental Research: Energy 1, 025003 (2024). https://dx.doi.org/10.1088/2753-3751/ad3fbb.

32. US Nuclear Regulatory Commission (NRC). Backgrounder on Nuclear Insurance and Disaster Relief; US NRC, Washington, DC, https://www.nrc.gov/reading-rm/doc-collections/fact-sheets/nuclear-insurance.html (Accessed August 2025 (Last updated August 2025).

33. World Nuclear News (WNN). Constellation to restart Three Mile Island unit, powering Microsoft; WNN, London, United Kingdom, https://www.world-nuclear-news.org/articles/constellation-to-restart-three-mile-island-unit-powering-microsoft (2024).

34. Brian Martucci. Constellation plans 2028 restart of Three Mile Island unit 1, spurred by Microsoft PPA; Utility Dive (Informa TechTarget), Newton, Massachusetts, https://www.utilitydive.com/news/constellation-three-mile-island-nuclear-power-plant-microsoft-data-center-ppa/727652/ (2024).

35. Matt Wald. Resurrecting Three Mile Island; Nuclear Newswire, American Nuclear Society (ANS) Westmont, Illinois, https://www.ans.org/news/2025-01-24/article-6714/resurrecting-three-mile-island/ (2025).

36. Commonwealth Fusion Systems (CFS) and Google. Google and Commonwealth Fusion Systems Sign Strategic Partnership; CFS and Google, Devens, Massachusetts and Mountain View, California, https://cfs.energy/news-and-media/google-and-commonwealth-fusion-systems-sign-strategic-partnershipGoogle (2025).

37. David Kirtley. Announcing Helion's fusion power purchase agreement with Microsoft; Helion Energy, Everett, Washington, https://www.helionenergy.com/articles/announcing-helion-fusion-ppa-with-microsoft-constellation/ (Accessed August 2025).

38. Edward S. Rubin, Inês M. L. Azevedo, Paulina Jaramillo & Sonia Yeh. A review of learning rates for electricity supply technologies. Energy Policy 86, 198-218 (2015). https://doi.org/10.1016/j.enpol.2015.06.011.

39. US Department of Energy (DOE) Office of Nuclear Energy. Commercializing Advanced Nuclear Reactors Explained in Five Charts; Washington, DC, https://www.energy.gov/ne/articles/commercializing-advanced-nuclear-reactors-explained-five-charts (2023).

40. Julie Kozeracki, Chris Vlahoplus, Katheryn Scott, Melissa Bates, Billy Valderrama, Erica Bickford, Tim Stuhldreher, Andrew Foss & Tom Fanning. Pathways to Commercial Liftoff: Advanced Nuclear; US Department of Energy (DOE), Washington, DC, https://www.energy.gov/sites/default/files/2025-07/LIFTOFF_DOE_Advanced-Nuclear.pdf (2023).

41. Kelly Cummins, Melissa Klembara, Alexa Thompson, Olivia Corriere, Emily Goldfield & William Dean. Portfolio Insights: Learning from Case Studies: Financing and Development Approaches from Recent First-of-a-Kind Projects; Office of Clean Energy Demonstration (OCED), Washington, DC, https://www.energy.gov/sites/default/files/2024-11/FOAK%20Financing%20and%20Development%20Approaches_112024_vf.pdf (2024).

42. Ryan Spangler, Sunming Qin, Levi M Larsen, Chandu Bolisetti, Abdalla Abou-Jaoude, Mehdi Asgari, Koroush Shirvan, W. Robb Stewart, James Krellenstein & Garrett Wilkinson. Potential Cost Reduction in New

Nuclear Deployments Based on Recent AP1000 Experience: Systems Analysis & Integration Campaign; US Department of Energy (DOE) Systems Analysis & Integration Campaign, Washington, DC, https://inldigitallibrary.inl.gov/sites/STI/STI/Sort_173162.pdf (2025).

43. Department for Energy Security and Net Zero. Guidance: Contracts for Difference (CfD) Allocation Round 6: results GOV.UK, London, United Kingdom, https://www.gov.uk/government/publications/contracts-for-difference-cfd-allocation-round-6-results (2024).

44. Google Sustainability. 24/7 by 2030: Realizing a Carbon-free Future; Mountain View, California, https://sustainability.google/reports/247-carbon-free-energy/ (2020).

45. Lisa Martine Jenkins. The 'clean transition tariff' won approval in Nevada. What's next for Fervo?; Latitude Media, Somerville, Massachusetts, https://www.latitudemedia.com/news/the-clean-transition-tariff-won-approval-in-nevada-whats-next-for-fervo/ (2025).

46. Vicki M. Baldwin (Parsons Behle & Latimer). Public Utilities Commission of Nevada Electronic Filing RE: Docket No. 23-04016 - Investigation regarding Sierra Pacific Power Company d/b/a NV Energy's Progress of Developing a Clean Transition Tariff and its Appropriateness for Development as a New Tariff Offering Pursuant to the Modified Final Order Issued in. Docket Nos. 22-06014, 22-06015 and 22-06016; Public Utilities Commission of Nevada, Carson City, Nevada, https://pucweb1.state.nv.us/PDF/AxImages/DOCKETS_2020_THRU_PRESENT/2023-4/26664.pdf (2023).

47. Deepak Patil. Train AI for less: Improve ML Goodput with elastic training and optimized checkpointing; Google Cloud, Mountain View, California, https://cloud.google.com/blog/products/ai-machine-learning/elastic-training-and-optimized-checkpointing-improve-ml-goodput (2025).

48. John Roach. Microsoft datacenter batteries to support growth of renewables on the power grid; Microsoft, Redmond, Washington, https://news.microsoft.com/source/features/sustainability/ireland-wind-farm-datacenter-ups (2022).

49. Alex C. Newkirk, Nichole Hanus & Christopher T. Payne. Expert and operator perspectives on barriers to energy efficiency in data centers. Energy Efficiency 17, 63 (2024). https://doi.org/10.1007/s12053-024-10244-7.

50. Tyler H. Norris, Tim Profeta, Dalia Patino-Echeverri & Adam Cowie-Haskel. Three Key Takeaways: Rethinking Load Growth in U.S. Power Systems; Nicholas Institute for Energy, Environment & Sustainability at Duke University, Durham, North Carolina, https://nicholasinstitute.duke.edu/articles/three-key-takeaways-rethinking-load-growth-us-power-systems (2025).

51. Will Reynolds & Emma Long. Digital Realty and Enel X Partner to Support Ireland's Renewable Energy Transition; Digital Realty, Dublin, Ireland, https://www.digitalrealty.com/about/newsroom/press-releases/123259/digital-realty-and-enel-x-partner-to-support-irelands-renewable-energy-transition (2024).

52. Georgia Butler. Schneider Electric & Aeven to deliver excess power to Danish grid; Data Center Dynamics, London, United Kingdom, https://www.datacenterdynamics.

com/en/news/schneider-electric-aeven-to-deliver-excess-power-to-danish-grid/ (2023).

53. Vera Grishchenko. Eaton UPS-as-a-Reserve Proven in Data Center Pilot; Eaton, Espoo, Finland, https://www.eaton.com/us/en-us/company/news-insights/news-releases/2018/UPSasR-pilot-Norway.html (2018).

54. Equinix. Response by Equinix to Cru Consultation: Large Energy Users Connection policy; Redwood City, California, https://consult.cru.ie/en/system/files/materials/200/CRU202504ac%20-%20Equinix.pdf (2024).

55. National Renewable Energy Laboratory (NREL). Flexible Loads and Renewable Energy Work Together in a Highly Electrified Future; NREL, Golden, Colorado, https://www.nrel.gov/manufacturing/news/program/2021/flexible-loads-and-renewable-energy-work-together-in-a-highly-electrified-future (2021).

56. Ella Zhou & Trieu Mai. Electrification Futures Study: Operational Analysis of U.S. Power Systems with Increased Electrification and Demand-Side Flexibility (NREL/TP-6A20-79094); National Renewable Energy Laboratory (NREL), Golden, Colorado, https://docs.nrel.gov/docs/fy21osti/79094.pdf (2021).

57. Andy Home. US aluminium smelters vie with Big Tech for scarce power; Reuters, London, United Kingdom, https://www.reuters.com/markets/commodities/us-aluminium-smelters-vie-with-big-tech-scarce-power-andy-home-2025-05-22/ (2025).

58. Ltd. (SRDL) Sanrui Electric Furnace Co. How Much Power Does an Electric Arc Furnace Require? ; SRDL, Xi'an, Shaanxi Province, China, https://www.srfurnace.com/how-much-power-does-an-electric-arc-furnace-require.html (2025).

59. Jim Middleton. Decarbonisation of steam crackers; Decarbonisation Technology (Crambeth Allen Publishing Ltd), Croydon, United Kingdom, https://www.ten.com/sites/energies/files/2023-05/decarbonisation_of_steam_crackers_technipenergies.pdf (2021).

60. Ana Radovanovic. Our data centers now work harder when the sun shines and wind blows (Blog); Google, Mountain View, California, https://blog.google/inside-google/infrastructure/data-centers-work-harder-sun-shines-wind-blows/ (2020).

61. Heather Klemick, Elizabeth Kopits & Ann Wolverton. Data Center Energy Efficiency Investments: Qualitative Evidence from Focus Groups and Interviews; US Environmental Protection Agency (EPA), Washington, DC, https://www.epa.gov/environmental-economics/data-center-energy-efficiency-investments-qualitative-evidence-focus-groups (2017).

62. CPower. 4CP Management System: A Dynamic Solution to Managing 4CP Charges; Baltimore, Maryland, https://cpowerenergy.com/wp-content/uploads/2016/12/ERCOT_4CP_Web_Download.pdf Accessed August 2025).

63. Electric Reliability Council of Texas (ERCOT). ERCOT Four Coincident Peak Calculations; ERCOT, Taylor, Texas, https://www.ercot.com/mktinfo/data_agg/4cp (2024).

64. Rodan Energy Solutions. Everything You Need to Know About 5CP In PJM (Blog); Princeton, New Jersey, https://rodanenergy.com/everything-you-need-to-know-about-5cp-in-pjm (2024).

65. Resource Adequacy Planning. PJM Manual 19: Load Forecasting and Analysis (Revision: 35); PJM (Pennsylvania-New Jersey-Maryland) Interconnection, Valley Forge, Pennsylvania, https://www.pjm.com/-/media/DotCom/documents/manuals/archive/m19/m19v35-load-forecasting-and-analysis-12-31-2021.pdf (2021).

66. Grid Status Exports. ERCOT's 4CP Summer Demand Roller Coaster Takes Off as Storage Flips Outcomes (Blog); Chicago, Illinois, https://blog.gridstatus.io/ercot-4cp-2025-june/ (2025).

67. Marc Spieler. How AI Factories Can Help Relieve Grid Stress; NVIDIA, Santa Clara, California, https://blogs.nvidia.com/blog/ai-factories-flexible-power-use/ (2025).

68. Philip Colangelo, Ayse K. Coskun, Jack Megrue, Ciaran Roberts, Shayan Sengupta, Varun Sivaram, Ethan Tiao, Aroon Vijaykar, Chris Williams, Daniel C. Wilson, Zack MacFarland, Daniel Dreiling, Nathan Morey et al. Turning AI Data Centers into Grid-Interactive Assets: Results from a Field Demonstration in Phoenix, Arizona. arXiv:2507.00909 (2025). https://doi.org/10.48550/arXiv.2507.00909.

69. International Energy Agency (IEA). Energy and AI; IEA, Paris, France, https://www.iea.org/reports/energy-and-ai (2025).

70. Enel X Japan. Demand Response; Tokyo, Japan, https://www.enelx.com/jp/en/demand-response (Accessed August 2025).

71. Agency for Natural Resources and Energy (ANRE). Guidelines for Energy Resource Aggregation Business; ANRE, Ministry of Economy, Trade and Industry (METI), Tokyo, Japan, https://www.enecho.meti.go.jp/en/category/vpp_dr/data/guidelines_for_energy_resource_aggregation_business.pdf (2015 (Last revised June 2020)).

72. Emi Bertoli, Vida Rozite, Brendan Reidenbach, Anthony Vautrin, Sungjin Oh & Matthieu Suire. Demand Response; International Energy Agency (IEA), Paris, France, https://www.iea.org/energy-system/energy-efficiency-and-demand/demand-response (Accessed August 2025 (Last updated July 2023)).

73. International Energy Agency (IEA). Executive Summary in Integrating Solar and Wind (IEA, Paris, France, 2024, https://www.iea.org/reports/integrating-solar-and-wind/executive-summary).

74. California Independent System Operator (ISO). Monthly Renewables Performance Report: July 2024; California ISO, Folsom, California, https://www.caiso.com/documents/monthly-renewables-performance-report-july-2024.html (2024).

75. Ali Jahanshahi, Nanpeng Yu & Daniel Wong. PowerMorph: QoS-Aware Server Power Reshaping for Data Center Regulation Service. ACM Trans. Archit. Code Optim. 19, Article 36 (2022). https://doi.org/10.1145/3524129.

76. Sophie Hall, Francesco Micheli, Giuseppe Belgioioso, Ana Radovanović & Florian Dörfler. Carbon-Aware Computing for Data Centers with Probabilistic Performance Guarantees. arXiv:2410.21510 (2024). https://ui.adsabs.harvard.edu/abs/2024arXiv241021510H.

77. C. A. Silva, R. Vilaça, A. Pereira & R. J. Bessa. A review on the decarbonization of high-performance computing centers. Renewable and Sustainable Energy Reviews 189, 114019 (2024). https://doi.org/10.1016/j.rser.2023.114019.

78. Lori Aniti. Solar and wind power curtailments are increasing in California; US Energy Information Administration (EIA), Washington, DC, https://www.eia.gov/todayinenergy/detail.php?id=65364 (2025).

79. Jigar Shah. DOE Releases New Report on Pathways to Commercial Liftoff for Virtual Power Plants (Blog); US Department of Energy (DOE) Loan Programs Office, Washington, DC, https://www.energy.gov/lpo/articles/doe-releases-new-report-pathways-commercial-liftoff-virtual-power-plants (2023).

80. National Energy System Operator (NESO). Demand Flexibility Service (DFS) (Report); NESO, Warwick, United Kingdom, https://www.neso.energy/document/363911/download (2025).

81. National Energy System Operator (NESO). Demand Flexibility Service delivers electricity to power 10 million households; NESO, Warwick, United Kingdom, https://www.neso.energy/news/demand-flexibility-service-delivers-electricity-power-10-million-households (2023).

82. National Energy System Operator (NESO). Demand Flexibility Service (DFS) (Website); NESO, Warwick, United Kingdom, https://www.neso.energy/industry-information/balancing-services/demand-flexibility-service-dfs (Accessed August 2025).

83. Solar Power Portal. National Grid DSO releases two-year expansion plan; Solar Power Portal. Informa PLC, London, United Kingdom, https://www.solarpowerportal.co.uk/energy-policy/national-grid-dso-releases-two-year-expansion-plan (2024).

84. National Energy System Operator (NESO). Homepage; NESO, Warwick, United Kingdom, https://www.neso.energy/ (Accessed August 2025).

85. A. Arteconi, N. J. Hewitt & F. Polonara. State of the art of thermal storage for demand-side management. Applied Energy 93, 371-389 (2012). https://doi.org/10.1016/j.apenergy.2011.12.045.

86. Manajit Sengupta, Yu Xie, Anthony Lopez, Aron Habte, Galen Maclaurin & James Shelby. Solar Resource Maps and Data (https://www.nrel.gov/gis/solar-resource-maps) from The National Solar Radiation Data Base (NSRDB). Renewable and Sustainable Energy Reviews 89, 51-60 (2018). https://doi.org/10.1016/j.rser.2018.03.003.

87. Bloom Energy. Survey: 27% of data centers are expected to run entirely on onsite power by 2030; Latitute Studios, Parkville, Maryland, https://www.latitudemedia.com/news/survey-27-of-data-centers-are-expected-to-run-entirely-on-onsite-power-by-2030/ (2025).

88. Ethan Howland. FERC rejects interconnection pact for Talen-Amazon data center deal at nuclear plant (Dive Brief); Utility Dive (Informa TechTarget), Newton, Massachusetts, https://www.utilitydive.com/news/ferc-interconnection-isa-talen-amazon-data-center-susquehanna-exelon/731841/ (2024).

89. Avi Salzman. How Amazon's Nuclear Deal Could Solve a Touchy Political Problem; Barron's (Dow Jones & Company), New York, New York, https://www.barrons.com/articles/amazon-nuclear-deal-talen-energy-42e8b78b (2025).

90. Zachary Skidmore. Diversity of power - the biggest data center energy stories of 2024; Data Center Dynamics, London, United Kingdom, https://www.datacenterdynamics.com/en/analysis/diversity-of-power-the-biggest-data-center-energy-stories-of-2024/ (2025).

91. Grid Deployment Office. Building a Better Grid Awards | January 2025; US Department of Energy (DOE), Washington, DC, https://www.energy.gov/gdo/articles/building-better-grid-awards-january-2025 (2025).

92. Loans Program Office. LPO Financing Programs; US Department of Energy (DOE), Washington, DC, https://www.energy.gov/lpo/lpo-financing-programs (Accessed August 2025).

93. Mack Ramsden & Joshua Rosen. Publication: Order No. 1920: A Guide to FERC's Landmark Transmission Planning Order; Foley Hoag LLP, Boston, Massachusetts, https://foleyhoag.com/news-and-insights/publications/alerts-and-updates/2024/may/order-no-1920-a-guide-to-fercs-landmark-transmission-planning-order/ (2024).

94. GOV.UK. Collection: Contracts for Difference London, United Kingdom, https://www.gov.uk/government/collections/contracts-for-difference (Accessed August 2025).

95. US Environmental Protection Agency (EPA). Utility Green Tariffs; EPA, Washington, DC, https://www.epa.gov/green-power-markets/utility-green-tariffs (Accessed August 2025).

96. Cindy Sabato. State Corporation Commission Must Set Fair Rates and Protect Virginia Residents from Subsidizing Data Center Infrastructure; Piedmont Environmental Council, Warrenton, Virginia, https://www.pecva.org/work/energy-work/state-corporation-commission-must-set-fair-rates-and-protect-virginia-residents-from-subsidizing-data-center-infrastructure/ (2025).

97. Tom Kleckner. SPP, Hitachi Partner to Use AI in Clearing GI Queue; RTO Insider LLC (Yes Energy), Boulder, Colorado, https://www.rtoinsider.com/107367-spp-hitachi-partner-ai-clearing-gi-queue/ (2025).

98. Amil Merchant, Simon Batzner, Samuel S. Schoenholz, Muratahan Aykol, Gowoon Cheon & Ekin Dogus Cubuk. Scaling deep learning for materials discovery. Nature 624, 80-85 (2023). https://doi.org/10.1038/s41586-023-06735-9.

99. The Materials Project. Materials Explorer: MgV_2O_5; Lawrence Berkeley National Laboratory (LBNL), Berkeley, California, https://next-gen.materialsproject.org/materials/mp-19003 (Accessed August 2025).

100. Colin McCormick. Chapter 13: Materials Innovation in Artificial Intelligence for Climate Change Mitigation Roadmap (Second Edition) (ICEF Innovation Roadmap Project; https://doi.org/10.7916/2j4p-nw61) (2024, Chapter 3 found at https://www.icef.go.jp/wp-content/themes/icef_new/pdf/roadmap/2024/13_ICEF2.0%20Materials%20Innovation_stand%20alone.pdf).

101. Alice C. Hill and Colin McCormick. Chapter 14: Extreme Weather Response in Artificial Intelligence for Climate Change Mitigation Roadmap (Second Edition) (ICEF Innovation Roadmap Project; https://doi.org/10.7916/2j4p-nw61) (2024, Chapter 14 found at https://www.icef.go.jp/wp-content/themes/icef_new/pdf/roadmap/2024/14_ICEF2.0%20Extreme%20Weather%20Response_stand%20alone.pdf).

102. Remi Lam, Alvaro Sanchez-Gonzalez, Matthew Willson, Peter Wirnsberger, Meire Fortunato, Ferran Alet, Suman Ravuri, Timo Ewalds, Zach Eaton-Rosen, Weihua Hu, Alexander Merose, Stephan Hoyer, George Holland et al. Learning skillful medium-range global weather forecasting. Science 382, 1416-1421 (2023). https://doi.org/10.1126/science.adi2336.

103. National Oceanic and Atmospheric Administration (NOAA) National Centers for Environmental Prediction (NCEP). Global Forecast System (GFS); NOAA NCEP, College Park, Maryland, https://www.ncei.noaa.gov/products/weather-climate-models/global-forecast (Accessed August 2025).

104. Contact Energy. Contact signs ten-year renewable energy agreement with Microsoft on new geothermal power station at Te Huka; Contact Energy Ltd, New Zealand, https://contact.co.nz/investor-centre/news/2023/contact-signs-ten-year-renewable-energy-agreement-with-microsoft (2023).

5 Data Center Water Use

Julio Friedmann, Angela Yuan and David

As data center capacity grows globally and interest in AI surges, data centers' water use has emerged as a critical concern in some regions. In part, this concern is due to the location of some data centers in water-scarce regions such as the US Southwest and Persian Gulf. This concern reflects the fundamental importance of water as a natural resource. As nations, communities and companies manage data center demand growth, understanding these facilities' water use is critical to making sound investment and policy decisions.

Data on the water consumption of data centers are poor. The authors were unable to find standardized data sets for water use at these facilities at the national, state, county or municipal level. Much of the data that do exist lack regional context, including whether water resources at a data center are scarce. This lack of data can make it difficult to evaluate and select potential solutions to water-related challenges at data centers, whether at the site-specific, regional or global level.

This chapter provides background on data center water use, placing it in the context of water used by other sectors. The chapter discusses the challenges created by data center water use, corporate initiatives in this area, differences between data centers' direct and indirect water use and options to reduce data centers' water footprint. The chapter concludes with recommendations.

A. Definitions

i. Water usage effectiveness (WUE)

The most common measure of water consumed at a data center is water usage effectiveness (WUE)—a data center's water use divided by the energy its information technology (IT) equipment uses. WUE is typically expressed in liters per kilowatt-hour (L/kWh).[1] WUE is straightforward to estimate; however, like power usage effectiveness (PUE), it is an incomplete metric at best. For example, WUE provides no information on several important topics such as water consumed by off-site electricity generation used to power a data center (which can be substantial) or water scarcity in the region where a data center is located.

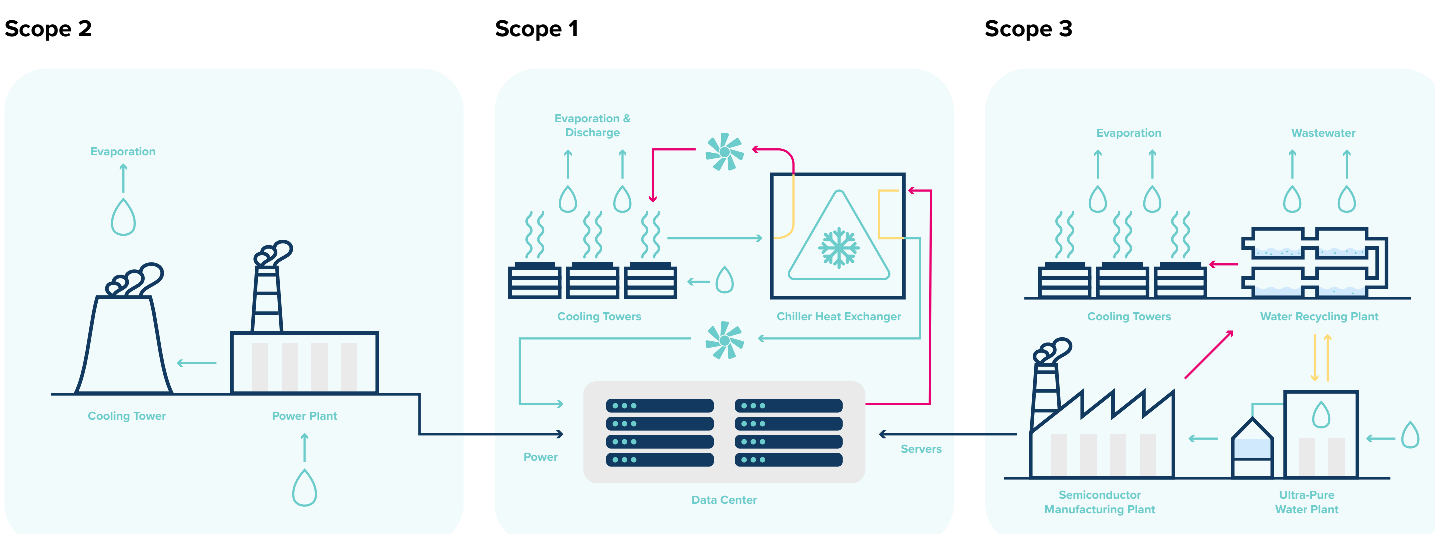
ii. Scope 1, 2 and 3 water use

Water use by data centers can be divided into three categories:

- **Scope 1:** Water used in direct operations (e.g., for cooling).
- **Scope 2:** Water used indirectly for power generation (e.g., evaporative cooling at thermal power plants).
- **Scope 3:** Water used indirectly in producing building materials, such as steel or cement ("embodied water").

These categories parallel the framework for emissions accounting under the Greenhouse Gas Protocol: Scope 1 (direct emissions), Scope 2 (energy-related emissions) and Scope 3 (embodied emissions).[2] See Figure 5-1.

Figure 5-1. *Scope 1, 2 and 3 water consumption.*



In general, direct water consumption at data centers (Scope 1) is modest. In contrast, indirect water consumption from thermal power generation for data centers (Scope 2) can be very large. (In most cases this occurs offsite at remote power stations.) Embodied water use (Scope 3) is likely to be minimal and distant from data center operations.

iii. Water use versus water consumption

Water use and water consumption are similar but distinct concepts.[3-5]

- Water use (often called water withdrawal) refers to freshwater taken from a watershed (ground water or surface water) and then used for purposes such as farming or drinking.

- Water consumption is the difference between water withdrawal and water returned to the same watershed. Water consumed is "evaporated, transpired, incorporated into products or crops, or otherwise removed from the immediate water environment."[5]

iv. Water-stress

A water-stressed region is one where water resources are insufficient to meet demands of human consumption. This can be due to a combination of factors, including limited rainfall, high water needs or inadequate infrastructure.
There are several frequently used definitions of water-stress, each with strengths and weaknesses.[6] Estimates of people worldwide exposed to water-stress vary greatly by methodology, ranging from 0.5 to 4.3 billion people.[7]

- The Falkenmark Water-Stress Indicator[8] measures the amount of renewable freshwater available per person per year. While straightforward, this definition has significant limitations. For example, it fails to account for the accessibility of water resources (whether some water in a region is deep underground or heavily polluted). It also fails to include man-made sources of freshwater, such as desalination plants.

- The Withdrawal-to-Availability Ratio measures the ratio of water withdrawal in a region to the total available water (as the name implies). This approach also has limitations. Like the Falkenmark Indicator, it does not include man-made sources of freshwater, water reclamation or water recycling.

There are other definitions, many of which are limited because they are static—they fail to take account of conditions that may change the extent of water-stress in a region over time.[9] Analyses of population growth, agricultural trends, industrial development plans, potential climate change impacts and other factors are necessary to fully assess water-stress in a region.[10]

B. Putting Data Center Water Use in Context

Data center water use is tiny globally but can be significant locally.
A comparison with agricultural water use is striking. Agriculture accounts for 70% of freshwater withdrawals globally, using roughly 20.4 trillion liters (5.48 trillion gallons) per day.[11,12] Data centers, in contrast, are projected to use roughly 1.7 billion liters (~450 million gallons) per day globally by 2030.[13] **Current global water use by agriculture is therefore over 12,000 times greater than projected global water use by data centers in 2030.** Put differently, projected water use by data centers in 2030 is roughly 0.008% of water used by agriculture today.

Box 5-1

ChatGPT versus Hamburgers: Comparing Water Consumption

To help put data center water use in context, a comparison between the water consumed by ChatGPT and hamburgers may be helpful. Our research suggests that, using conservative assumptions:

- 19,000 ChatGPT-3 queries and the process of producing one hamburger consume roughly the same amount of water.
- ChatGPT-4o takes roughly 537 days—almost a year-and-a-half—to consume as much water as needed to produce the hamburgers McDonald's sells in one day.

These estimates are based on the following.

- According to data in one study, roughly 328 liters (87 gallons) of water are consumed to produce one hamburger in the United States (mainly for growing feed). Data in another study suggest the figure is higher—roughly 580 liters[5] (153 gallons) per hamburger.

- A recent study found that a single ChatGPT-3 query in the United States used an average of roughly 17 milliliters of water, taking into account the water used in training and inference, as well as Scope 1 and Scope 2 emissions. The lower figure for water consumed to produce a hamburger just above (328 liters) and basic division (328 liters divided by 17 milliliters = 19,294) yields the first estimate: 19,000 ChatGPT-3 queries and one hamburger consume roughly the same amount of water.

- McDonald's sells roughly 6.5 million hamburgers each day.[13] The lower figure for water consumed to produce a hamburger just above (328 liters) and basic multiplication (6.5 million hamburgers x 328 liters) suggest that roughly 2.13 billion liters (565 million gallons) of water are consumed to produce the hamburgers McDonald's sells in one day.

- Another recent study found that GPT-4o's annual water consumption is roughly 1.45 billion liters (383 million gallons).[14] Comparing that figure to the rough annual water consumption needed for McDonald's hamburgers and doing basic division (2.13 billion liters divided by 1.45 billion liters = 1.47; 1.47 years = 537 days) yields the second estimate: ChatGPT4o takes roughly 537 days—almost a year-and-a-half—to consume as much water needed for the hamburgers McDonald's sells in one day.

Several caveats are required with respect to the foregoing estimates. First, the quality of data with respect to data center water use is poor, as emphasized at the beginning of this chapter. Second, the estimates above could understate ChatGPT's water consumption in comparison to hamburgers for several reasons. (For example, more recent large language models likely consume more water than ChatGPT-3.) On the other hand, the estimates above could also overstate ChatGPT's water consumption in comparison to hamburgers for several reasons. (For example, the estimates use a conservative assumption with respect to water consumed to produce hamburgers, given the study cited above that found that almost twice as much water is needed.)

We invite critiques of the ChatGPT/hamburger water consumption comparisons above. Yet even if the comparisons are off by an order of magnitude, they still provide insight into the rough scale of data center water use as compared to the water use of a familiar everyday product. More data collection, research and public dialogue on these topics are needed.

Comparing data center water use to natural water fluxes is helpful as well. Data centers represent a tiny fraction of natural water flows regionally.

The following comparisons provide context.

- Projected data center water demand globally in 2030 is roughly 0.02% of the Amazon River's daily discharge at its mouth.[12,17]

- Data centers in Northern Virginia, the largest global data center hub, use approximately 5.1 M gal/day[15]—0.07% of the Potomac River's daily flow (~7000 M gal/day) at Washington, D.C.[18] In Phoenix, Arizona, data centers consume roughly 177 M gal/day, roughly 1.36% of the Colorado River's average daily flow at Lees Ferry, Arizona and 22.2% of Maricopa County's daily water use.[23,24,25]

- A 20-50 MW facility can use 11-19 million liters per day (3-5 M gal/day), similar to the daily water use for a city of 30,000-40,000 people.[19]

The table below provides data on water use by data centers and total water flow in select regions. However, this context changes notably in regions of acute water-stress, as discussed below.

C. Data Centers in Water-Stressed Regions

Table 5-1. *Water flows in some geographies compared with water consumption from data centers.*

Region	Use Type	Average Daily Flow/Use in M L/day (M gal/day)
Global	**World's largest river by volume —Amazon River (at the river's mouth) (flow)**	**~14,679,600 (~3,877,800)[20]**
	Global data center 2030 forecast (use)	~3288 (~869)[13]
Virginia/Washington, D.C.	**Potomac River (at Washington, D.C.) (flow)**	**~26,498 (~7000)[21]**
	Virginia surface and groundwater withdrawals for agriculture (use)	**~121 (~32)[22]**
	Northern Virginia data centers (use)	~19 (~5)[18]

China	**Yangtze River (at Datong station) (flow)**	**~2,653,747 (~700,894)**[20]
	China's agricultural use (use)	**~1,055,616 (~278,880)**[21]
	China's data centers (use)	~3557 (~940)[22]
Phoenix/Maricopa County (Arizona)	**Colorado River Basin (at Lees Ferry, AZ) (flow)**	**~49,339 (~13,034)**[23]
	Phoenix/Maricopa County's residential and commercial (use)	**~3017 (~797)**[24]
	Phoenix/Maricopa County's data centers (use)	~670 (~177)[25]
Singapore	**Singapore's water consumption (use)**	**~1666 (~440)**[29]
	Singapore's data centers (use)	~134 (~35)[30]

Water scarcity is inherently a local issue. In water-stressed regions, even modest water demands from data centers can exacerbate local tensions.

Data centers are frequently clustered in dry areas due to land and energy availability. Two-thirds of data centers built or planned in the United States since 2022 are in high or extremely high water-stress areas. Almost three-quarters of these projects are in five US states: Virginia, Texas, Arizona, Illinois and California.[13] Under the Chinese government's East Data, West Computing program, many new data centers will be built in arid provinces, including Inner Mongolia, Gansu and Ningxia. Many of the world's regions with growing tech infrastructure, such as California's Central Valley, Arizona's Sonoran Desert and northern China, already face chronic water shortages.

In these regions, data centers sometimes compete with agricultural water demand, which can be significant. For example, California's Central Valley grows 20% of US agricultural production, and the agriculture sector overwhelmingly dominates water use. In California, recent and recurring droughts,[31] court-ordered constraints on water withdrawal,[32] state executive orders on water conservation[33,34] and climate change impacts[35] have magnified concerns about additional stressors, including data center demands.

In Arizona, though data centers account for <0.1% of the state's total water use (~9.5 M liters per day or ~2.5 M gal/day), they are concentrated in fast-growing suburbs where supply is under pressure. In Chandler, Arizona, city officials passed a 2015 ordinance to restrict data center water usage, reflecting tensions around limited groundwater.[36]

Similarly, northern China faces severe water-stress. Beijing's per capita water availability is far below the international water shortage threshold, intensifying the need to regulate industrial water usage strictly.[37] In Beijing, municipal and industrial daily water use totals around 6942 M L/day (1834 M gal/day).[38] While detailed local data-center-specific figures are not publicly available, national-level data indicate that 3557 M L/day (940 M gal/day) are used for all Chinese data centers, suggesting that the relative share of data centers in water-scarce regions like Beijing could be significant.[25] Indeed recent regulations in Beijing and surrounding provinces require data centers to enhance WUE significantly.[39] Such proactive policies help mitigate the additional water-stress posed by the rapidly expanding digital economy in these already strained areas.

In Singapore, a substantial share of freshwater is used by data centers in comparison to municipal uses. Due to limited freshwater availability, Singapore depends on desalination, water imports and advanced wastewater recycling.[30,40]

D. Direct Water Consumption (Scope 1)

Most on-site (direct) water is consumed by data center cooling.[41,42] High-quality data on direct water use are difficult to find. Individual facility estimates or direct data are very scarce. Estimates for direct water consumption come with high uncertainty and limited ability to extrapolate for many reasons:

- Data mixing: Direct and indirect (power related) water consumption is reported together without separation or distinction.[43-45]

- Poor transparency: Few operators—less than one third[1]—disclose data directly. For example, while both Microsoft and Google share these data each year, other US operators do not.

- Regulatory range: Data centers operate across a wide range of regulatory actors and jurisdictions. Most often, water is managed by cities and municipalities, yielding a wide range of mandates, requirements and standards. This makes aggregation difficult and disclosure enforcement spotty or difficult.

- Water source: WUE does not differentiate between municipal water, reclaimed water or desalinated water.[46]

- Diverse geography: Data centers operate across a wide range of geographies. Cool-climate data centers require less cooling and less water. The converse is true for warm climates, humid or dry.

In addition, modern data centers, including those under construction, come equipped with a diverse set of cooling options with dramatically different water consumption indexes: open-loop cooling, closed-loop cooling, adiabatic cooling, air cooling or advanced cooling technologies (see Chapter 2.3 of this Roadmap). Some systems also recover direct-use water. Most estimates based on power consumption or PUE do not represent this range of technology or water efficiency at all.

Table 5-2. *Estimated water consumption by data centers*

Data Center	Estimated direct water (total consumed)	Estimated WUE (rate)	Additional notes
Microsoft: 2022,[47] 2023[48]	6.9 B L/year (1.7 B gal/year)	0.3 L/kWh	~40% reduction from 2021–2023
Google: self-reported, 2021[49]	620 M L/year (164 M gal/year)		
Amazon Web Services: self-reported, 2024[50]		0.15 L/kWh	~40% reduction from 2021–2024
Meta: self–reported, 2024[51]		0.20 L/kWh	~33% reduction from 2020–2023
Industry average WUE in 2024[52]		1.85 L/kWh	
Practical range of WUE in 2024[41]		0.0–2.5	
Google: GSERB[53]	16.3 B		Unclear if this also includes indirect (power-related) water
2015 vintage[54] 1-MW data center	25 M L/year (6.75 M gal/year)		An extreme case for an old system; it does not include efficiencies of scale
Northern Virginia in 2023[18]	7 B L/year (1.85 B gal/year)		Approximately 50 data centers, not all
Japan in 2025[55]	89 B L/year	53 L/kWh	Very high estimated WUE–unclear if correct; most likely includes both indirect and embodied water in estimate across
UK in 2024[56]	~10 B L/year (~2.6 B gal/year)		231 data centers
Global estimate in 2023[18]	~720 B L/year (~190 B gal/year)		Data source: Bluefield Research[57]

The range of estimates complicates the task of estimating the likely on-site ("Scope 1") water use for a 200 MW data center.

- One estimate of average 2024 WUE is 1.85 L/kWh.[52] Using this WUE, an average 200 MW data center would consume ~3.1 M L/year (0.82 M gal/year).

- The IEA published an estimate for a typical US data center.[58] However, they did not separate direct and indirect water consumption in their calculation:
    - The inclusive estimate (i.e., including direct, indirect and embodied water use) was 2 M L/year (~0.5 M gal/year) for a 100 MW facility or 4 M L/year (~1.1 M gal/year) for a 200 MW facility. Assuming a 95% capacity factor, this would be a WUE of 2.5 L/kWh.

    - In IEA's figure of 5.27, cooling consumed roughly 20% of the total consumption in 2023. This would yield an estimate of direct water use of 0.8 M L/year (~0.2 M gal/year) for a 200 MW facility, or a WUE of 0.5.

- Using their published number (WUE of 0.3 L/kWh), an average 200 MW Microsoft facility with a 95% capacity factor would consume 0.5 M L/year (0.14 M gal/year).

- Meta, LuxConnect and Amazon claim an average of 0.2 L/kWh or less. An average 200 MW data center for one of these firms would consume ~0.33 M L/year (~0.09 M gal/year) or less.

Chapter 2.3 of this Roadmap describes a range of technologies that could reduce direct water consumption at data centers. Innovation in this space, driven by industrial priorities, is both rapid and profound. Before designing any targeted research, development and demonstration (RD&D) program, a government census of current technologies and their likely water resource benefit would help avoid investment in wasteful efforts and dead ends. This effort would require data sharing from companies and greater transparency—a priority for decision makers.

E. Indirect Water Consumption - Energy-Related (Scope 2)

i. Background

Water use from power generation cooling represents a significant portion of the indirect water footprint for data centers. This Scope 2 water use can account for the majority of a data center's total water impact, especially in fossil-fuel dominated grids.[59] Thermal power plants (coal, nuclear and natural gas generation) consume significant water volumes for cooling. Air cooling systems consume relatively little water but represent a small fraction of existing plants and an equally small fraction of recent power generation.

In general, generation withdraws and consumes freshwater, commonly from rivers, lakes and groundwater, although most coastal power plants withdraw seawater for cooling. Water withdrawals are commonly much larger than water use; for example, US thermal power withdrawals in 2015[50] were roughly 500 B L/day (133 B gal/day) but consumption was only 10.2 B L/day (2.7 B gal/day). Decision makers should take care in distinguishing between these two in data and analysis.

Transitioning from thermal to renewable energy sources can dramatically cut the indirect water use associated with data centers. Wet-cooled thermal power plants, for example, typically consume approximately 2900-3000 liters (766-793 gallons) per megawatt-hour,[61] while renewables consume virtually none (Table 5-3). This highlights renewables' dual climate and water conservation benefits.

***Table 5-3.** Water consumption by generation type—US average data*

Units are L/kWh, which are equal to m^3/MWh and tonne/MWh. Units in parentheses are gallons per MWh. Data sources: Thunder Said Energy[62] and National Renewable Energy Laboratory (NREL).[5]

Generation Type	Thunder Said Units: L/kWh (gal/MWh)	NREL, median value Units: L/kWh (gal/MWh)
Nuclear	2.1 (554)	2.54 (672)
Coal	2.0 (528)	2.6 (687)
Gas – natural gas combine cycle (NGCC)	1.2 (317)	0.7 (198)
Gas – open cycle	0.1-0.2 (26-54)	--
Solar photovoltaic (PV)	0.01-0.1 (2.6-26)	0.1 (26)
Wind	0.0000001 (0.000026)	0 (0)

Table 5-4. Power-related water consumption by data centers in key nations. Units are billions of liters (in parentheses are billions of gallons)

Country	Annual water consumption Billions of l (billions of gals)	Daily water consumption Billions l/day (billions g/day)	National % of daily water use (for power)
United States	~5280 (~1390)	14.5 (3.8)	~12%[63]
China	~12,700 (3350)	34.8 (9.2)	~23%[64,65]
India	~5920 (1563)	16.2 (4.3)	~18%[66]
France	~855 (225)	2.3 (0.6)	~10[67]
United Kingdom	~292 (77)	0.8 (0.2)	~9[67]

In cases where data center loads receive power from new generation, water use will depend on the technology being used. This varies significantly by country and region. New US thermal generation is mostly gas, for example, while new Chinese thermal generation is mostly coal, on top of renewable generation.

Estimates of US data center power consumption vary widely. The IEA (2025) estimate is 175 TWh total in 2023, reflecting significant recent growth (see Chapter 1 of this Roadmap for details).

ii. Local regulatory constraints on water use for power generation

Global regulatory approaches vary widely by region, but they often focus on three key levers:

- **Incentives or mandates for adopting more water-efficient cooling systems**, such as dry or hybrid cooling as opposed to traditional once-through cooling, aim to lower water consumption.
- **Limits on water discharge temperatures** aim to minimize adverse thermal impacts on local aquatic life.
- **Restrictions on water withdrawals** have become critical for protecting water bodies and aquatic ecosystems, especially during periods of heat waves and droughts.

The following sections detail significant restrictions on how water withdrawals and total consumption for thermal power generation are regulated and restricted. These restrictions do not apply directly to water withdrawals and consumption for data centers per se. However, they could affect data centers indirectly if their power supply comes from thermal generation.

United States: In the United States, federal regulations under the Clean Water Act 316(a) and 316(b) provide a comprehensive framework that helps guide the location, design and operation of water cooling structures in power generation. Section 316(a) limits water discharge temperatures, while 316(b) mandates that Best Technology Available (BTA), including closed-cycle cooling systems, water reuse, recirculating systems and other technologies, be adopted to minimize adverse environmental impacts from water cooling systems.[68] Additionally, 316(b) requires National Pollutant Discharge Elimination System (NPDES) permits for withdrawal for larger facilities, and the US Environmental Protection Agency's (EPA's) 2004 Phase II rule required existing power plants to reduce withdrawals to levels comparable to closed-cycle systems. Furthermore, during periods of high heat and drought, regulations become even more restrictive. These restrictions can limit both access to cooling water and its use, especially for power generation.

These federal regulations are reinforced at the state level, and regulations vary considerably by state. For example, California's 2010 Once-Through Cooling Policy banned once-through cooling at coastal power plants to reduce water consumption.[69] The state also strictly regulates discharged cooling water, requiring that it not exceed receiving water temperatures by more than 20 °F (11.1 °C).[70]

China: China's heavy dependence on thermal power generation exacerbates regional water-stress, sometimes significantly. This has resulted in stringent power plant water-use regulations. Since the mid-2000s, China has required new coal-fired power plants in water-stressed regions to adopt dry-cooling technologies, resulting in 70-80% reductions in water use compared to conventional wet-cooling systems.[71] The 2013 Water Allocation Plan for Coal Bases introduced regional caps on water use and accelerated wastewater recycling in the power sector in water-scarce regions.[72] China's 2015 Water Pollution Prevention and Control Action Plan also mandates discharge temperature limits, typically <35 °C or ≤3-5 °C above intake levels.[73] Further, China's Three Red Lines national policy sets binding limits on water use, withdrawal efficiency and wastewater discharge for power generation.[74] Finally, many provinces, like Hebei and Guangdong, have added restrictions in more densely populated or ecologically sensitive regions.

India: In India, 40% of thermal power plants are in regions with high water-stress.[75] Since 1999, India has required that all thermal power plants using freshwater to meet withdrawal limits and adopt closed-cycle cooling systems to reduce water use and thermal pollution.[76] India's Central Pollution Control Board (CPCB) limits the temperature of condenser cooling water from thermal power plants to a maximum of 7°C above ambient water temperature for new coastal plants and 10°C above the temperature of inlet water for existing plants.[77] These rules are particularly important in water-stressed states like Maharashtra and Rajasthan, where water demand for agriculture, energy and municipal supply already exceeds availability.

As these regulatory frameworks evolve, they shape the sustainability of data centers' indirect water use by influencing the water-efficiency of local power generation. This underscores the importance of grid decarbonization, not just for emissions, but for water security as well.

F. Indirect Water Consumption - Embodied (Scope 3)

Much less water is used in manufacturing materials for data centers than for direct cooling of a data center or for cooling the thermal generation used by data centers. In part, this is because the physical materials going into a data center are functional for years (chips) or decades (concrete, structural steel, rebar).

Even given these characterizations, the embodied water use at data centers merits greater understanding and attention in light of these facilities' growth, size and rate of deployment. This analysis focuses on four elements of embodied water consumption: concrete, structural steel, rebar and chips. To simplify, the unit we consider is a 200 MW data center. Physical footprints of data centers vary chiefly by energy density, with low-energy density footprints of ~120,000 m^2 (~1.3 million ft^2) and high-energy density footprints of ~60,000 m^2 (~650,000 sq ft). Numbers will also vary somewhat by geography and vintage of production.

i. Concrete

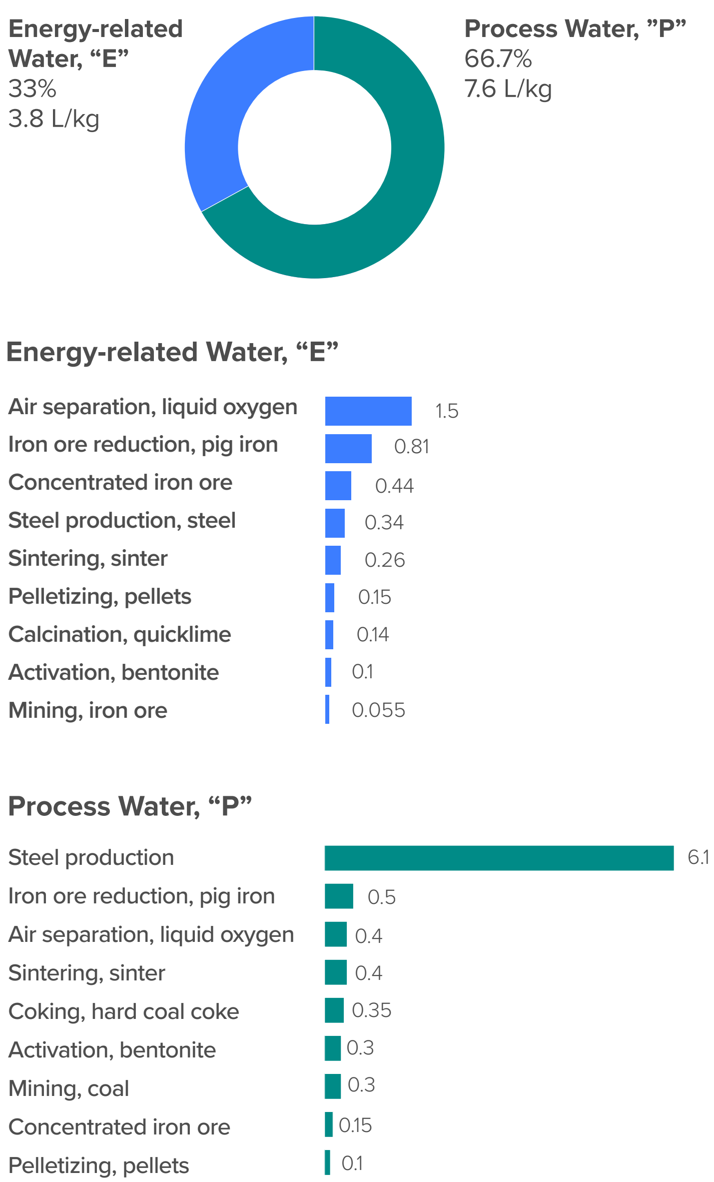
Concrete production and use require significant water.[78] Estimates range from 2000-2500 liters per m^3 of concrete. This includes ~500 L/ton for production of clinker, limestone mining and washing as well as a significant fraction (up to 80%) as mixing water, which is necessary to convert raw Portland cement into the binder of concrete.[78]

A 200 MW data center can require a wide range of concrete volumes as a function of building code, seismic requirements and design. Estimates range between 55,000 and 500,000 m^3. Assuming 300,000 m^3 of concrete for the physical plant, including foundations, walls, slabs and service areas, a data center will use roughly 700 million liters of water for concrete (ranging from 110-1250 million liters), equivalent to roughly 2% of the water consumed by US data centers in a single day.

ii. Steel: structural steel and rebar

Primary steel production requires significant water inputs.[78] For conventional blast-furnace/basic oxygen furnace operation (BF-BOF), water is consumed in mining, upgrading ores, oxygen production for blast furnace use, cooling and energy inputs. Although iron ore mining occurs in water-stressed regions (e.g., Australia, South Africa), most steel is produced in water-abundant areas (e.g., Hubei province, China, Japan, Korea and Germany) and most of the water consumption is associated with production.[78] On average, primary production consumes 11.8 liters of water per kg of steel (11.8 m^3/tonne or 3.1 gal/tonne).

Figure 5-2. *Water consumption in steel production.*



Wedges labeled “P” are primary production (direct water consumption) and wedges labeled “E” are water consumed by energy inputs (indirect). Source: Gerber-Leenes et al., 2018,[78] including supplemental material.[79]

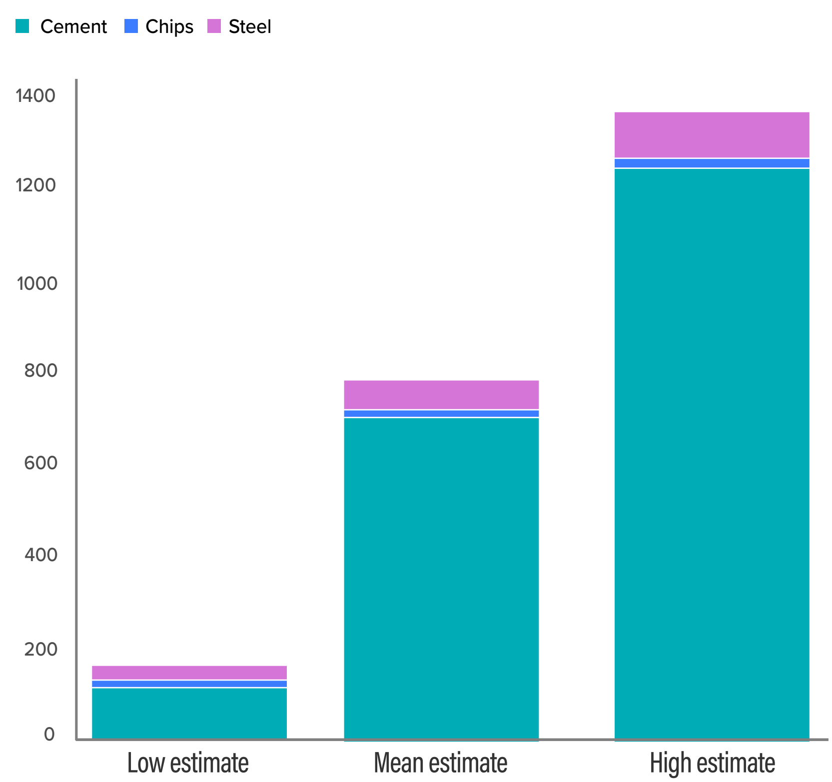
Robust, validated estimates for structural steel and rebar use in data centers are scarce. Assuming 50-75 kg/ft^2 of floor (535-802 kg/m^2), estimates of steel water footprint would range from ~34 million to ~98 million liters (9 to 26 million gallons) of water per 200 MW data center.

iii. Chips

Advanced semiconductor production (chips) consumes significant water resources. An integrated circuit can require 2200 liters per unit, with a large chip fabrication facility ("fab") consuming as much as 18.2 million L/day (4.8 million gal/day) to produce 40,000 wafers each month. In addition, much of this water must be "ultra-pure" water, which itself consumes water to produce (ultrapure water is used[80] to rinse wafers, cool special manufacturing equipment and dilute chemicals for manufacturing). Published estimates indicate that 1400-1600 liters of water[81] are needed to produce 1000 liters of ultrapure water, with 1500 liters per unit of ultrapure water needed to produce a single 30-cm wafer.

Assuming 5000-6000 wafers are needed to stock a 200 MW data center, roughly 2.3-3.3 million liters (580,000-870,000 gallons) are required to provide the initial set of chips. If each set of chips lasts 3-5 years[82] and a data center operates for 15 years, then 4 times that much water is needed, not including the water requirements for mining high-grade silica or copper, transportation, or upgrading of ores. This is a conservative estimate, since graphics processing units (GPUs) are replaced with higher frequency (2-3 years).

Figure 5-3. *Summary estimate of water (liters) required for building a 200 MW data center, high range and low range by material.*



G. Corporate Initiatives

Some tech firms self-report water usage. In 2023, Microsoft reported that 42% of its water withdrawals company-wide came from stressed areas, and Google reported 15%.[76] However, many other tech companies do not self-report.

Beyond reporting, key data center buyers and operators have made specific commitments regarding water use:

- Microsoft has committed to being "net water positive" by 2030, meaning it will generate or return a greater volume of water than is consumed
- Google has made a similar commitment, declaring its intent to produce or return 20% more water than it consumes by 2030.[77]
- Amazon Web Services (AWS) began sharing its annual WUE in 2022[78] and has also committed to being "net water positive" by 2030.[79]

Not all companies have been as proactive. A 2022 analysis[80] concluded that only 16% of data center operators have declared plans to manage water-stress. Although leading companies have reported low WUE values, the industry average is significantly higher (see Table 5-2)

H. Options for water footprint reduction

Companies (e.g., Microsoft,[81] Amazon Web Services[45] and Meta[82,83]) are demonstrating significant improvements in managing water consumption today, including reducing WUE. In most cases, these improvements come from applying advanced cooling technologies and increasing purchases of low-carbon electricity. At the same time, growing demand is increasing the volume of direct data center water consumption, which adds water-stress locally, while increasing use of natural gas and coal generation, which adds water-stress to some regions and communities. New, expanded and restarted nuclear facilities will consume significant water volumes and might add to regional water-stress.

These changes are driven by a combination of internal corporate policies, market demands, economic and operational trade-offs, and infrastructure limits. Given these dynamics, companies, governments and regulators have a relatively modest list of options to consider:

- **Siting:** Water is consumed where data centers, thermal generation or manufacturing operate, so siting plays an outsized role in water planning.

Many considerations affect siting choices, including economics, infrastructure access, climate and community concerns. Potential water-stress should factor into siting decisions as well.

- Unsurprisingly, placing key manufacturing sites (e.g., chip fabs) and large data centers in water scarce regions can stress natural resources and create political and economic challenges. Conversely, placing these facilities outside of water scarce regions minimizes regional stresses and friction.

- Secondary water supply options may help siting decisions, such as potential for water reclamation and reuse.[81] Similarly, companies may be able to make arrangements with agricultural or municipal water users to help reduce the companies' water consumption through interventions.

- **Direct:** Since the overwhelming majority of direct water consumption is associated with cooling during data center operations, the primary vector for water footprint reduction is alternative cooling pathways. (See Chapter 2.3 of this Roadmap.)

- **Indirect:** Since the majority of indirect water consumption is associated with using thermal electricity generation, the primary vector for water footprint reduction is to use non-thermal water sources, chiefly wind and solar with battery support. (See Chapter 3.2 of this Roadmap.)

- **Embodied:** Concrete, steel and chips all consume significant amounts of water (including ultra-pure water for chip-making). Producing these goods using water-saving approaches and technologies would help reduce water-stress. For example, this could include use of reclaimed water (e.g., in concrete mixing operations and at chip fabs).[81] Specifically for concrete, mixing water is an enormous part of the water used. Technologies that use CO_2 for binding concrete (e.g., CarbonCure or CarbonBuilt) can significantly reduce use of mixing water. Similarly, advanced steel production technologies, such as direct reduction of iron with electric arc furnace (DRI+EAF) production, would reduce water use in general. Ultimately, supply chain standards and procurement policies could help reduce embodied emissions.

The majority of these approaches involve economic and operational trade-offs. For example, water scarce regions often have lower land costs, which is a significant economic component of facility siting costs. Similarly, procuring building materials and chips with low-water footprints may involve paying a green premium or accessing limited supplies. Absent regulation or policy-based incentives, it may take considerable time for new technologies to enter the market and replace existing kit.

I. Conclusion

Significant gaps in data availability and quality limit understanding of data center water use. Estimates are often inconsistent, incomplete and lack a local or regional context, which hinders investors, policy makers and civil society in making sound decisions regarding data center water use. The scale and kinds of water demands from data centers must be considered in the context of other water demands in a region or nation, such as agricultural or municipal use. Similarly, tracking and quantifying the direct, indirect and embodied water consumption of data centers is important in developing management plans.

Technologies, such as improving efficiency, reclaiming and reusing water, and prioritizing renewable power over thermal electricity generation, can help reduce water-stress from data centers. Especially in regions of water scarcity and high water-stress, data center operators should prioritize the use of advanced cooling technologies and develop procurement standards for low-water footprint building materials and chips to minimize water impacts and strain. Finally, in-depth, systems-level analyses can identify the most useful technology or regulatory options and reveal the costs and trade-offs between water-conserving options. These analyses can serve as the basis for policy and private sector solutions.

J. Recommendations

1. *Governments should* ***assemble and share data*** *related to direct, indirect and embodied water consumption from data center construction and operation. Data center owners and operators should volunteer to* ***share site-specific water use and consumption data*** *proactively and invite third-party review. If necessary, governments should require disclosure of this information.*

2. *All stakeholders should* ***recognize that data center water use is tiny in relation to water use by other sectors globally but can be very significant in water-scarce regions.***

3. ***Before siting data centers****, data center owners and operators should* ***assess likely water impacts****, including in particular by consulting with local stakeholders. In water-scarce regions, companies should consider several steps to reduce likely water impacts:*

a. ***Apply advanced cooling approaches*** *to reduce direct water use, with potential additional expense.*

b. ***Assess the cost and viability of water reclamation and reuse*** *and of increasing water supply (e.g., through desalination).*

c. ***Maximize non-thermal power supplies****, including solar, wind and batteries, including potential overbuilding of variable renewable resources and hybrid load balancing using a mix of thermal and non-thermal generation.*

d. ***Develop procurement standards*** *for building materials and chips with low-water footprints. Where possible, procure low-water footprint materials, including the cost of a modest green premium.*

4. *Governments should* ***support the development of advanced technologies*** *that limit the water footprint associated with AI use.*

    a. *Most importantly, governments should* ***support replacing fossil generation*** *with non-thermal generation and should encourage use of air cooling in existing facilities, both of which would dramatically reduce indirect water use.*

    b. *Similarly, governments should* ***support novel cement, concrete and steel technology*** *that would reduce water consumption, as well as CO_2 emissions. Where possible, companies should accelerate adoption and procurement of low-water pathways.*

    c. *Governments should* ***undertake a cost-benefit analysis*** *based on the lifetime of operation and seek support to reduce risk and cost.*

5. *Governments* ***should undertake initial and then systematic analysis to understand the technology options, costs and trade-offs between water-conserving options****. These analyses can serve as the basis for policy, including regulation or incentives. Data center builders and operators should share their data with government agencies to help identify low-cost, large-volume options for water footprint reduction.*

K. References

1. David Mytton. Data centre water consumption. npj Clean Water 4, 11 (2021). https://doi.org/10.1038/s41545-021-00101-w.

2. World Resources Institute (WRI) and World Business Council for Sustainable Development (WBCSD). Standards & Guidance; WRI and WBCSD, Washington, DC and Geneva, Switzerland, https://ghgprotocol.org/standards-guidance (Accessed August 2025).

3. Pengfei Li, Jianyi Yang, Mohammad A. Islam & Shaolei Ren. Making AI Less "Thirsty": Uncovering and Addressing the Secret Water Footprint of AI Models. arXiv:2304.03271 (2023). https://doi.org/10.48550/arXiv.2304.03271.

4. Paul Reig. What's the Difference Between Water Use and Water Consumption?; World Resources Institute (WRI), Washington, DC, https://www.wri.org/insights/whats-difference-between-water-use-and-water-consumption (2013).

5. Jordan Macknick, Robin Newmark, Garvin Heath & KC Hallett. A Review of Operational Water Consumption and Withdrawal Factors for Electricity Generating Technologies (Technical Report: NREL/TP-6A20-50900); National Renewable Energy Laboratory, Golden, Colorado, https://docs.nrel.gov/docs/fy11osti/50900.pdf (2011).

6. Fitsume T. Wolkeba, Mesfin M. Mekonnen, Kate A. Brauman & Mukesh Kumar. Indicator metrics and temporal aggregations introduce ambiguities in water scarcity estimates. Scientific Reports 14, 15182 (2024). https://doi.org/10.1038/s41598-024-65155-5.

7. Fitsume T. Wolkeba, Mesfin M. Mekonnen, Kate A. Brauman & Mukesh Kumar. Supplementary Information for "Indicator metrics and temporal aggregations introduce ambiguities in water scarcity estimates". Scientific Reports 14, 15182 (2024). https://doi.org/10.1038/s41598-024-65155-5; Suppmentary Information can be found at https://static-content.springer.com/esm/art%3A10.1038%2Fs41598-024-65155-5/MediaObjects/41598_2024_65155_MOESM1_ESM.docx.

8. Malin Falkenmark, Jan Lundqvist & Carl Widstrand. Macro-scale water scarcity requires micro-scale approaches. Natural Resources Forum 13, 258-267 (1989). https://doi.org/10.1111/j.1477-8947.1989.tb00348.x.

9. Muhammad Azam Khan, Ijaz Uddin & Nor Salwati Othman. Factors determining water stress: do environmental distress, energy consumption and industrialization matter? Environment, Development and Sustainability (2024). https://doi.org/10.1007/s10668-024-04653-y.

10. Office of Clean Energy Demonstrations (OCED). Managing Water for All: An OECD Perspective on Pricing and Financing; OECD Studies on Water, OECD Publishing, Paris, France, https://doi.org/10.1787/9789264059498-en (2009).

11. Scientific and Cultural Organization (UNESCO) United Nations Educational. UN World Water Development Report: Statistics; UNESCO, Paris, France, https://www.unesco.org/reports/wwdr/en/2024/s (2024).

12. High Tide Technologies: Resources. Water Usage In The Agricultural Industry; High Tide Technologies, Nashville, Tennessee, https://www.htt.io/learning-center/water-usage-in-the-agricultural-industry. (Accessed August 2025).

13. Leonardo Nicoletti, Michelle Ma & Dina Bass. AI is Draining Water from Areas that Need it Most; Bloomberg Technology + Green, New York, New York, https://www.bloomberg.com/graphics/2025-ai-impacts-data-centers-water-data/ (2025).

14. C. Alan Rotz, Senorpe Asem-Hiablie, Sara Place & Greg Thoma. Environmental footprints of beef cattle production in the United States. Agricultural Systems 169, 1-13 (2019). https://doi.org/10.1016/j.agsy.2018.11.005.

15. Greg Thoma, Ben Putman, Marty Matlock, Jennie Popp & Leah English. Sustainability Assessment of U.S. Beef Production Systems; University of Arkansas Resiliency Center, Fayetteville, Arkansas, https://scholarworks.uark.edu/rescentfs/3 (2017).

16. Erin Metz. How Many Burgers Does McDonald's Actually Sell Each Day?; Food Republic, Fishers, Indiana, https://www.foodrepublic.com/1769851/how-many-burgers-mcdonalds-sells-daily/ (2025).

17. Nidhal Jegham, Marwan Abdelatti, Lassad Elmoubarki & Abdeltawab Hendawi. How Hungry is AI? Benchmarking Energy, Water, and Carbon Footprint of LLM Inference. arXiv:2505.09598 (2025). https://doi.org/10.48550/arXiv.2505.09598.

18. Camilla Hodgson. US tech groups' water consumption soars in 'data centre alley'; Financial Times, London, United Kingdom, https://www.ft.com/content/1d468bd2-6712-4cdd-ac71-21e0ace2d048 (2024).

19. Nicholas Barrowclough. Transforming data centre cooling for a sustainable future; Data Center Magazine, Norwich, Norfolk, United Kingdom, https://datacentremagazine.com/articles/transforming-data-centre-cooling-for-a-sustainable-future (2023).

20. Water Resources Division US Geological Survey (USGS). The Amazon, measuring a mighty river. (USGS, Washington, DC, 1967), https://doi.org/10.3133/70046378; https://www.usgs.gov/publications/amazon-measuring-a-mighty-river.

21. Interstate Commission on the Potomac River Basin (ICPRB). Potomac Basin Facts: Interstate Commission on the Potomac River Basin; ICPRB, Rockville, Maryland, https://www.potomacriver.org/potomac-basin-facts/ (Accessed August 2025).

22. Virginia Department of Environmental Quality (DEQ). Commonwealth of Virginia Work Plan: USGS Water Use Data and Research (WUDR) Program; DEQ, Richmond, Virginia, https://d9-wret.s3.us-west-2.amazonaws.com/assets/palladium/production/s3fs-public/atoms/files/Virginia-WUDR-Workplan.pdf (2016).

23. Ze-Nan Zhu, Xiao-Hua Zhu, Chuanzheng Zhang, Minmo Chen, Hua Zheng, Zhensheng Zhang, Jiwen Zhong, Lixin Wei, Qiang Li, Hua Wang, Shuming Li & Arata Kaneko. Monitoring of Yangtze River Discharge at Datong Hydrometric Station Using Acoustic Tomography Technology. Frontiers in Earth Science Volume 9 - 2021 (2021). https://doi.org/10.3389/feart.2021.723123.

24. US Central Intelligence Agency (CIA). The World Factbook (2021 Archive): Field Listing - Total water withdrawal; CIA, Washington, DC, https://www.cia.gov/the-world-factbook/about/archives/2021/field/total-water-withdrawal/ (2021).

25. Hazel Moises. China's Data Centers and AI Industry Pose Growing Water Supply Concerns; W.Media, Suntec Tower 1, Singapore, https://w.media/chinas-data-centers-and-ai-industry-pose-growing-water-supply-concerns/ (2024).

26. Navajo Nation Water Rights Commission (NNWRC). Mainstem Lower Colorado River Basin Arizona; NNWRC, Window Rock, Arizona, https://nnwrc.navajo-nsn.gov/Basin-Updates/Mainstem-Lower-Colorado-River-Basin-Arizona (Accessed August 2025).

27. University of Arizona College of Agriculture & Life Sciences Cooperative Extension Water Resources Research Center (WRRC). Arizona Water Factsheet: Maricopa County - Water in Arizona; University of Arizona WRRC, https://www.maricopa.gov/DocumentCenter/View/73858/Maricopa-County-Water-Factsheet-121721 (2021).

28. Felicity Barringer. Thirsty for power and water, AI-crunching data centers sprout across the West; & The West by Stanford University's Bill Lane Center for the American West, Stanford, California, https://andthewest.stanford.edu/2025/thirsty-for-power-and-water-ai-crunching-data-centers-sprout-across-the-west/ (2025).

29. Public Utilities Board (PUB): Singapore's National Water Agency. Water Conservation; PUB, Environment Building, Singapore, https://www.pub.gov.sg/public/waterloop/water-conservation (Accessed August 2025).

30. Mordor Intelligence. Study Of Data Center Water Consumption In Singapore Size & Share Analysis - Growth Trends & Forecasts (2025 - 2030); Hyderabad, India, https://www.mordorintelligence.com/industry-reports/study-of-data-center-water-consumption-in-singapore?utm_source=chatgpt.com (2025).

31. Marcia DeLonge. In California's Central Valley, Drought Is a Growing Threat to Farms, Food, and People; Union of Concerned Scientists, Cambridge, Massachusetts, https://blog.ucs.org/marcia-delonge/in-californias-central-valley-drought-is-a-growing-threat-to-farms-food-and-people/. (2022).

32. Jennifer Eastman. U.S. Supreme Court Protects the Snail Darter; Elton B. Stephens Company (EBSCO), Ipswich, Massachusetts, https://www.ebsco.com/research-starters/politics-and-government/us-supreme-court-protects-snail-darter (2023).

33. Rachel Becker. Hot, inland California cities face the steepest water cuts with new conservation mandate; CalMatters, Sacramento, California, https://www.ebsco.com/research-starters/politics-and-

government/us-supreme-court-protects-snail-darter (2024).

34. California Water Boards: State Water Resources Control Board. Water Conservation Emergency Regulations; Sacramento, California, https://www.waterboards.ca.gov/conservation/regs/emergency_regulation.html (Accessed August 2025).

35. Alvar Escriva-Bou, Josué Medellín-Azuara, Ellen Hanak, John Abatzoglou & Joshua Viers. Policy Brief: Drought and California's Agriculture; Public Policy Institute of California (PPIC), San Francisco, California, https://www.ppic.org/publication/policy-brief-drought-and-californias-agriculture/ (2022).

36. Brett Walton. Data Centers a Small, But Growing Factor in Arizona's Water Budget; Circle of Blue, Traverse City, Michigan, https://www.circleofblue.org/2025/supply/data-centers-a-small-but-growing-factor-in-arizonas-water-budget/ (2025).

37. Sida Liu, Yangxiao Zhou, Fatima Eiman, Michael E. McClain & Xu-sheng Wang. Towards sustainable groundwater development with effective measures under future climate change in Beijing Plain, China. Journal of Hydrology 633, 130951 (2024). https://doi.org/10.1016/j.jhydrol.2024.130951.

38. Beijing Daily. Beijing's Water-Use Efficiency of 2024 Ranks First in China; Beijing, China, https://english.beijing.gov.cn/latest/news/202502/t20250217_4012816.html?v%3Djt (2025).

39. Soon Chen Kang, Perkins Liu & Dan Thompson. Beijing steps up power and renewable energy regulations on datacenters (Market Insight Report Reprint); Vertiv, Westerville, Ohio, (2021). https://www.vertiv.com/49e7be/globalassets/documents/analyst-reports/451_reprint_beijingpowerrenewablesregs_13may2021_333454_0.pdf?utm_source=chatgpt.com (2025).

40. Public Utilities Board (PUB): Singapore's National Water Agency. Singapore's Water Loop: Managing Our Water Demand; PUB, Environment Building, Singapore, https://www.pub.gov.sg/public/waterloop#:~:text=Water%20demand%20in%20Singapore%20is,By%202065%2C (Accessed August 2025).

41. Alfonso Capozzoli & Giulio Primiceri. Cooling Systems in Data Centers: State of Art and Emerging Technologies. Energy Procedia 83, 484-493 (2015). https://doi.org/10.1016/j.egypro.2015.12.168.

42. Miguel Yañez-Barnuevo. Data Centers and Water Consumption; Environmental and Energy Study Institute (EESI), Washington, DC, https://www.eesi.org/articles/view/data-centers-and-water-consumption (2025).

43. Javier Farfan & Alena Lohrmann. Gone with the clouds: Estimating the electricity and water footprint of digital data services in Europe. Energy Conversion and Management 290, 117225 (2023). https://doi.org/10.1016/j.enconman.2023.117225.

44. Ganesh Hegde. Conventional vs AI Data Center Cooling Options and How Much Wastewater is Being Generated; KETOS, Sunnyvale, California, https://www.sciencedirect.com/science/article/pii/S019689042300571X?via%3Dihub (2024).

45. Rosie Frost. ChatGPT 'drinks' a bottle of fresh water for every 20 to 50 questions we ask, study warns; Euronews, Lyon, France, https://www.euronews.com/green/2023/04/20/chatgpt-drinks-a-bottle-of-fresh-water-for-every-20-to-50-questions-we-ask-study-warns (2023).

46. Andrew Higgins. What Is Water Usage Effectiveness (WUE) in Data Centers?; Equinix, Inc., Redwood City, California, https://blog.equinix.com/blog/2024/11/13/what-is-water-usage-effectiveness-wue-in-data-centers/ (2024).

47. Matt O'Brien & Hannah Fingerhut. Artificial intelligence technology behind ChatGPT was built in Iowa — with a lot of water; Associated Press (AP), New York, New York, https://apnews.com/article/chatgpt-gpt4-iowa-ai-water-consumption-microsoft-f551fde98083d17a7e8d904f8be822c4 (2023).

48. David Chernicoff & Matt Vincent. Microsoft's 'Sustainable By Design' Initiative Seeks to Transform Data Center Water Efficiency; Data Center Frontier, Lawrenceville, New Jersey, https://www.datacenterfrontier.com/sustainability/article/55249649/microsofts-sustainable-by-design-initiative-seeks-to-transform-data-center-water-efficiency (2024).

49. Urs Hölzle. Our commitment to climate-conscious data center cooling (Blog); Google, Mountain View, California, https://blog.google/outreach-initiatives/sustainability/our-commitment-to-climate-conscious-data-center-cooling/ (2022).

50. Amazon. 2024 Amazon Sustainability Report; Amazon.com, Inc., Seattle, Washington, https://sustainability.aboutamazon.com/2024-amazon-sustainability-report.pdf (2025).

51. Meta. 2023 Meta Sustainability Report: 2023 Environmental Data Index; Meta, Menlo Park, California, https://sustainability.atmeta.com/wp-content/uploads/2023/07/Meta-2023-Environmental-Data-Index.pdf (2023).

52. Christopher Tozzi. A Guide to Data Center Water Usage Effectiveness (WUE) and Best Practices; Data Center Knowledge (Informa TechTarget), Boston, Massachusetts, https://www.datacenterknowledge.com/cooling/a-guide-to-data-center-water-usage-effectiveness-wue-and-best-practices?utm_source=chatgpt.com (2025).

53. JLL. The data center boom: Navigating growth and sustainability; Global Real Estate Sustainability Benchmark (GRESB), Amsterdam, Netherlands, https://www.gresb.com/nl-en/the-data-center-boom-navigating-growth-and-sustainability/ (2025).

54. Kevin Heslin. Ignore Data Center Water Consumption at Your Own Peril; Uptime Institute LLC, New York, New York, https://journal.uptimeinstitute.com/dont-ignore-water-consumption/ (2016).

55. Mordor Intelligence. Study Of Data Center Water Consumption In Japan Size & Share Analysis - Growth Trends & Forecasts (2025 - 2030); Hyderabad, India, https://www.mordorintelligence.com/industry-reports/study-of-data-center-water-consumption-in-japan? (2025).

56. Adam Vaughan. Thirsty data centres are sucking up Britain's scarce water supplies; The Times and The Sunday Times, London, United Kingdom, https://www.thetimes.com/uk/environment/article/data-centres-water-drought-pns6nqds2 (2025).

57. Bluefield Research. Water for Data Centers; Boston, Massachusetts, https://www.bluefieldresearch.com/data/water-for-data-centers/ (Accessed August 2025).

58. International Energy Agency (IEA). Energy and AI; IEA, Paris, France, https://www.iea.org/reports/energy-and-ai (2025).

59. Masaō Ashtine & David Mytton. Water-guzzling data centres; University of Oxford, Department of Engineering Science, Oxford, United Kingdom, https://eng.ox.ac.uk/case-studies/the-true-cost-of-water-guzzling-data-centres/ (2021).

60. Timothy Diehl Melissa Harris. Withdrawal and consumption of water by thermoelectric power plants in the United States, 2015; US Geological Survey (USGS) Tennessee Water Science Center; Lower Mississippi-Gulf Water Science Center, Nashville, Tennessee, https://doi.org/10.3133/sir20195103; https://www.usgs.gov/publications/withdrawal-and-consumption-water-thermoelectric-power-plants-united-states-2015 (2019).

61. Jagjit Singh Sehra & Michael Kost. How Thermal Power Plants Can Save 80% of Their Water; POWER Magazine, Houston, Texas, https://www.powermag.com/how-thermal-power-plants-can-save-80-of-their-water/ (2021).

62. Thunder Said Energy. Water intensity of power generation?; Austin, Texas, https://thundersaidenergy.com/downloads/water-intensity-of-power-generation/ (Accessed August 2025).

63. US Water Alliance. Water's Value; Federal Way, Washington, https://uswateralliance.org/resources/waters-value/. (Accessed August 2025).

64. T. David Zhou. China's 2022 Water Resources Bulletin Released; Yangtze Institute for Conservation and Development, Nanjing, Jiangsu, China, https://www.yicode.org.cn/en/chinas-2022-water-resources-bulletin-released/ (2023).

65. Xu Peng, Hong Chen, Honglin Zhong, Ruyin Long, Chao Zhang, Dandan Zhao, Guangfei Yang, Jingke Hong, Cuncun Duan, Xinxian Qi, Pengbang Wei, Pengfei Zhang & Jindao Chen. Water-saving co-benefits of CO₂ reduction in China’s electricity sector. iScience 26 (2023). https://doi.org/10.1016/j.isci.2023.106035.

66. Shresth Tayal, Sanchit S Agarwal, Saika Anjum & Saswata Chaudhury. Water use in power generation - Role of energy transition in India; International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria, https://acace-explorer.ece.iiasa.ac.at/story/india_water_transition/ (2023).

67. Alena Lohrmann, Michael Child & Christian Breyer. Assessment of the water footprint for the European power sector during the transition towards a 100% renewable energy system. Energy 233, 121098 (2021). https://doi.org/10.1016/j.energy.2021.121098.

68. US Environmental Protection Agency (EPA). Cooling Water Intakes; EPA, Washington, DC, https://www.epa.gov/cooling-water-intakes (Accessed August 2025).

69. California Water Boards: State Water Resources Control Board. Ocean Standards - CWA §316(b) Regulation; Sacramento, California, https://www.

waterboards.ca.gov/water_issues/programs/ocean/cwa316/ (Accessed August 2025).

70. California Water Boards: State Water Resources Control Board. Water quality control plan for control of temperature in the coastal and interstate waters and enclosed bays and estuaries of California; Sacramento, California, https://www.waterboards.ca.gov/water_issues/programs/ocean/docs/wqplans/thermpln.pdf (1998).

71. Connor Bebb. Water Rights Transfers and High-Tech Power Plants Hold off Energy-Water Clash in Northern China; Circle of Blue, Traverse City, Michigan, https://www.circleofblue.org/2011/world/from-agriculture-to-industry-efficiency-upgrades-transfer-water-use-rights-on-china%E2%80%99s-yellow-river/ (2011).

72. China Water Risk (CWR). MWR Announces Water-for-Coal Plan; CWR, Central, Hong Kong, https://cwrrr.org/notices/mwr-announces-for-coal-plan/ (2013).

73. Ministry of Ecology and Environment (MEE) of the People's Republic of China. Water Pollution Prevention and Control Law of the People's Republic of China (Amended in 2017); MEE, Beijing, China, https://english.mee.gov.cn/Resources/laws/environmental_laws/202012/t20201211_812662.shtml (2017).

74. Zongyong Zhang, Yuli Shan, Dandan Zhao, Martin R. Tillotson, Bofeng Cai, Xian Li, Heran Zheng, Cunxue Zhao, Dabo Guan, Junguo Liu & Yu Hao. City level water withdrawal and scarcity accounts of China. Scientific Data 11, 449 (2024). https://doi.org/10.1038/s41597-024-03115-4.

75. Tianyi Luo, Deepak Sriram Krishnan & Shreyan Sen. Parched Power: Water Demands, Risks, and Opportunities for India's Power Sector; World Resources Institute (WRI), Washington, DC, https://www.wri.org/parched-power-water-demands-risks-and-opportunities-indias-power-sector (2018).

76. Central Electricity Authority (CEA) of India. Report on Minimisation of Water Requirement in Coal Based Thermal Power Stations; CEA, New Delhi, India, https://cea.nic.in/old/reports/others/thermal/tetd/min_of%20water_coal_power.pdf (2012).

77. Central Pollution Control Board (CPCB). Environmental Standards for Ambient Air, Automobiles, Fuels, Industries and Noise (Pollution Control Law Series: PCLS/4/2000-2001); CPCB, Ministry of Environment & Forests, Goverment of India, Delhi, India, https://urbanemissions.info/wp-content/uploads/docs/CPCB-2000-Environmental-Standards.pdf (2000).

78. P. W. Gerbens-Leenes, A. Y. Hoekstra & R. Bosman. The blue and grey water footprint of construction materials: Steel, cement and glass. Water Resources and Industry 19, 1-12 (2018). https://doi.org/10.1016/j.wri.2017.11.002.

79. P. W. Gerbens-Leenes, A. Y. Hoekstra & R. Bosman. Supplemental Material from The blue and grey water footprint of construction materials: Steel, cement and glass. Water Resources and Industry 19, 1-12 (2018). https://doi.org/10.1016/j.wri.2017.11.002; Supplemental Material found at https://ars.els-cdn.com/content/image/1-s2.0-S2212371717300458-mmc1.docx.

80. Murugappa Water Technology & Solutions Pvt Ltd - (MWTS). How MWTS is Transforming the Semicon Industry; Chennai, India, https://murugappawater.com/how-mwts-is-transforming-the-semicon-industry/ (2024).

81. Alex Irwin-Hunt. Thirsty chip facilities under scrutiny in water stressed areas; fDi Intelligence, London, United Kingdom, https://www.fdiintelligence.com/content/c31f977a-a8b7-5ffc-9eaa-daa48a8d1d41 (2023).

82. Curvature Expert. How Long Do Servers Last?; Curvature, Santa Barbara, California, https://www.curvature.com/resources/blog/how-long-do-servers-last/ (2022).

83. Luke Barratt, Costanza Gambarini, Andrew Witherspoon & Aliya Uteuova. Revealed: Big tech's new datacentres will take water from the world's driest areas; The Guardian (US Edition), New York, New York, https://www.theguardian.com/environment/2025/apr/09/big-tech-datacentres-water (2025).

84. Heather Clancy. Sip or guzzle? Here's how Google's data centers use water; Trellis, Oakland, California, https://trellis.net/article/sip-or-guzzle-heres-how-googles-data-centers-use-water/ (2022).

85. Data Center Knowledge. AWS First CSP to Publish Data Center Water Usage Efficiency Metrics; Data Center Knowledge, San Francisco, California, https://www.datacenterknowledge.com/hyperscalers/aws-first-csp-to-publish-data-center-water-usage-efficiency-metrics (2022).

86. International Desalination and Reuse Association (IDRA). Amazon Web Services (AWS) Commits to Being Water Positive by 2030; IDRA, Danvers, Massachusetts, https://idadesal.org/wp-content/uploads/2024/06/Amazon-Water-Positive.pdf (2024).

87. Erin Johnson & Kata Molnar. ESG Risks Affecting Data Centers: Why Water Resource Use Matters to Investors; Sustainalytics, Amsterdam , Netherlands, https://www.sustainalytics.com/esg-research/resource/investors-esg-blog/esg-risks-affecting-data-centers-why-water-resource-use-matters-to-investors (2022).

88. Steve Solomon. Sustainable by design: Next-generation datacenters consume zero water for cooling; Microsoft, Redmond, Washington, https://www.microsoft.com/en-us/microsoft-cloud/blog/2024/12/09/sustainable-by-design-next-generation-datacenters-consume-zero-water-for-cooling/ (2024).

89. Meta. 2021 Sustainability Highlights Report; Menlo Park, California, https://sustainability.fb.com/wp-content/uploads/2022/07/2021-Meta-Sustainability-Highlights.pdf (2022).

90. Meta. 2024 Sustainability Report; Menlo Park, California, https://sustainability.atmeta.com/wp-content/uploads/2024/08/Meta-2024-Sustainability-Report.pdf (2024).

TEXT BOX

Data Center Electronic Waste

Angela Yuan and David Sandalow

As data centers expand globally, the amount of electronic waste (e-waste) they generate is growing. This waste includes discarded servers, networking devices, batteries, hard drives and other electronic components.

The total amount e-waste produced by data centers is unknown. One related study found that the artificial intelligence (AI) boom could add 1.2-5 million metric tons of e-waste cumulatively between 2020 and 2030.[1] According to the United Nations's (UN) Global E-waste Monitor, e-waste from all sources (including consumer electronics, lighting equipment and household appliances) was 62 million metric tons in 2022 and is projected to reach 82 million tons by 2030.[2]

Proper management of e-waste is crucial. If not handled correctly, hazardous materials in e-waste (such as lead, mercury and cadmium) can leach into soil and water and pose serious environmental and health risks.[2] Improper disposal of e-waste also causes loss of valuable materials including gold, iron, copper and rare-earth elements. (UN estimates placed the value of metals in e-waste in 2022 at approximately $91 billion.[2]) The majority of e-waste is not recycled and ends up in landfills or exported (in some cases to countries without proper recycling infrastructure or regulatory oversight).[1]

Some data center e-waste is recoverable and reusable, presenting a promising opportunity for circular economy strategies such as reuse, refurbishment and component harvesting.[3] In the United States, roughly 60% of the 20-70 million hard disk drives (HDDs) destroyed each year could be reused.[3] Some major tech companies are addressing these issues through refurbishment and reuse programs to extend hardware life, wiping data to allow for safe redeployment and partnering with certified recyclers for material recovery. For example, Microsoft operates a "Circular Center," which has achieved 90.9% of reuse and recycling for servers and components.[4,5] Google also runs programs to refurbish, reuse and recycle its e-waste.[6] One study found that implementation of circular economy strategies throughout the generative AI value chain could reduce e-waste generation by 16-86%.[1]

Many data center operators are adopting zero-waste certification standards and prioritizing circular economy principles in their designs.[7] However, security concerns remain challenging, leading to the destruction of reusable devices. A recent study estimated that 49% of devices from data centers were destroyed for privacy reasons, although 47% were still functional.[8] Further, when devices are recycled, different challenges arise. Careful management is needed to prevent adverse environmental and health impacts. Global recycling standards are lacking and international shipments often result in inadequate environmental oversight.[9,10]

Data center decommissioning is a critical stage for minimizing e-waste. Effective management of decommissioning can reduce e-waste, cut costs, and reduce legal, environmental and reputational risks, while still protecting sensitive data. However, there is currently an overall lack of standardization, transparency and oversight.

The AI boom will likely produce growing quantities of e-waste from data centers. A multifaceted approach and public-private partnerships can help manage this problem.

Recommendations

1. *Governments should:*
    a. ***Adopt and harmonize global standards for reuse, refurbishment and recycling of e-waste****, including for sanitization of data-bearing information technology (IT) equipment*
    b. ***Adopt and strengthen extended producer responsibility rules.***

2. *Data center operators should:*
    a. ***Refurbish, resell or donate retired equipment****; process that equipment through certified recyclers; and publicly report end-of-life outcomes*
    b. ***Prioritize measures to reduce equipment turnover****, such as regular preventative maintenance*
    c. ***Reduce equipment purchases where possible by using tools such as virtualization****, cloud computing and shared infrastructure*

3. *Manufacturers should:*
    a. ***Design modular equipment that is easily repaired and disassembled****, provide spare-part support, include clear recyclability labeling, and establish validated pathways for reuse, resale or refurbishment*
    b. ***Prioritize use of recycled or easily recyclable materials***
    c. ***Reduce or eliminate use of hazardous materials***

References

4. Peng Wang, Ling-Yu Zhang, Asaf Tzachor & Wei-Qiang Chen. E-waste challenges of generative artificial intelligence. Nature Computational Science 4, 818-823 (2024). https://doi.org/10.1038/s43588-024-00712-6.

5. Cornelis P. Baldé, Ruediger Kuehr, Tales Yamamoto, Rosie McDonald, Elena D'Angelo, Shahana Althaf, Garam Bel, Otmar Deubzer, Elena Fernandez-Cubillo, Vanessa Forti, Vanessa Gray, Sunil Herat, Shunichi Honda et al. Global E-waste Monitor 2024; International Telecommunication Union (ITU) and United Nations Institute for Training and Research (UNITAR), Geneva, Switzerland and Bonn, Germany, https://ewastemonitor.info/the-global-e-waste-monitor-2024/ (2024).

6. National Renewable Energy Laboratory (NREL). In a Circular Economy, Hard Drives Could Have Multiple Lives in the Future; NREL, Golden, Colorado, https://www.nrel.gov/news/detail/program/2021/in-a-circular-economy-hard-drives-could-have-multiple-lives-in-the-future (2021).

7. Microsoft. The Circular Center; Redmond, Washington, https://datacenters.microsoft.com/tour/circular-center/ (Accessed August 2025).

8. Libby Hargreaves. How Microsoft's Circular Supply Chain Hit 90.9% Reuse; Supply Chain Digital Magazine, Norfolk, United Kingdom, https://supplychaindigital.com/technology/how-microsoft-is-boosting-data-centre-component-circularity (2025).

9. Praneet Arshi & Kendal Smith. Harvesting hardware: Our approach to carbon-aware fleet deployment (Blog); Google, Mountain View, California, https://cloud.google.com/blog/topics/sustainability/hardware-harvesting-at-google-reducing-waste-and-emissions (2025).

10. US Green Building Council (USGBC). Applying TRUE to data center projects; USGBC, Washington, DC, https://support.usgbc.org/hc/en-us/articles/42159461305107-Applying-TRUE-to-data-center-projects (Accessed August 2025).

11. Blancco Technology Group. 2025 State of Data Sanitization Report; Bishop's Stortford, United Kingdom, https://blancco.com/resources/rs-data-sanitization-report/

12. Sensei .ie. Data Center Recycling: Limits of E-Waste Recycling Solutions; Human-I-T, Los Angeles, California, https://www.human-i-t.org/data-center-recycling/ (2024).

13. Rachel A. Meidl. Closing the Loop on the World's Fastest-growing Waste Stream: Electronics; Rice University's Baker Institute for Public Policy, Houston, Texas, https://www.bakerinstitute.org/research/closing-loop-worlds-fastest-growing-waste-stream-electronics (2023).

6 Government Policy

David Sandalow

Data centers are foundational infrastructure for the digital economy— powering email systems, video streaming, e-commerce, artificial intelligence (AI) models and more. As the digital economy has grown rapidly in recent years, governments around the world have paid increasing attention to data centers' energy use and environmental impacts. Electricity demand from data centers has received particular attention. Data centers' greenhouse gas emissions, local air pollutants, water consumption and e-waste have received attention as well.

Policymakers use a range of policy tools to address data centers' energy use and environmental impacts, including disclosure obligations, regulatory standard-setting and fiscal incentives. By far the most common policy is a requirement to achieve a stated level of power usage effectiveness (PUE)[a]. Standards for water usage effectiveness (WUE)[b] and disclosure obligations with respect to energy and water use are becoming more common as well.

Government policies on data centers' energy use and environmental impacts vary widely by jurisdiction. The European Union and several EU countries have disclosure obligations, energy efficiency standards and other policies aligned with their net-

a Power usage effectiveness (PUE) is a data center's total energy use divided by the energy used by its IT equipment.
b Water usage effectiveness (WUE) is a data center's water use divided by the energy used by its IT equipment. WUE is typically expressed in liters per kilowatt-hour (L/kWh).

Many thanks to Minjue Wu, Gareth Jones, Nadia Maunsell and Anne-Amélie Campant for their superb assistance with the research for this chapter.

zero greenhouse gas emissions goals. China, Japan and other Asian governments have energy efficiency standards and policies to encourage data centers to use renewable power. The policies of the US federal government change dramatically from administration to administration, with the current US administration emphasizing new data center construction as a high priority while deemphasizing environmental protection. (Some US states take a very different approach.)

Economy-wide policies (not focused on data centers) play an important role in data centers' energy use and environmental impacts. Policies to promote decarbonization of the electric grid, for example, help determine greenhouse gas emissions from data center power use. (See Chapter 3.2 of this Roadmap.) Energy efficiency standards for commercial buildings, which often apply to data centers, guide the technology choices and management practices of facility operators.

This chapter reviews government policies that specifically address data centers' energy use and environmental impacts. (Economy-wide policies not focused on data centers are mostly beyond the scope of this chapter.) Section A provides a global policy landscape, summarizing policies in key jurisdictions. Section B examines policies by topic, summarizing policies on disclosure, adequacy of power supplies, energy efficiency, renewable power, greenhouse gas emissions, local air pollution, water use and e-waste. Section C discusses the impacts of government data center energy and environment policies. Section D offers recommendations.

A. Global Policy Landscape

i. European Union (~10% of global data center capacity)[1]

European data center policies reflect the European Union's strong commitment to climate action and environmental protection more broadly. Under the European Green Deal and Energy Efficiency Directive (EED) recast of 2023, most data centers must report annually on energy and environmental metrics (including energy use, PUE, water use and waste-heat utilization), implement plans for continual energy efficiency improvement, and use or recover waste heat.[2,3] The European Commission has adopted a sustainability rating system for data centers to benchmark facilities on energy, carbon emissions and water efficiency.[2] The Climate Neutral Data Centre Pact (supported by the European Commission) commits data center operators to reach 100% carbon-free power by 2030 and meet ambitious energy efficiency targets (e.g., annual PUE ≤1.3 in cool climates and ≤1.4 in warm climates). While voluntary, the Pact was intended in substantial part to pre-empt stricter regulation.[4]

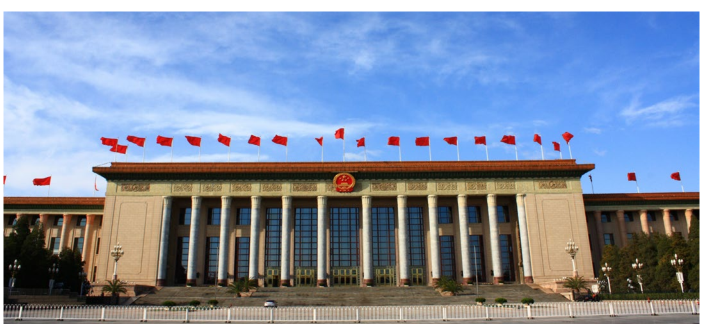
Several EU countries have their own policies to promote data center sustainability, including the following:

- The German government requires data centers to source at least 50% of their power from renewable sources. Starting in 2026, new data centers must have a PUE 1.2 or less and reuse at least 10% of their waste heat. Starting in 2027, all data centers in Germany must source 100% of their power from renewable sources, and new data centers must reuse at least 15% of their waste heat. Starting in 2030, all data centers in Germany must have a PUE of 1.3 or less, and new data centers must reuse at least 20% of their waste heat.[5-7]

- The Irish government imposed a moratorium on data center development and is now considering a proposal to require any new data center to provide dispatchable generation and/or storage capacity equal to its load on-site or nearby.[8-10]

- Denmark recently took several steps to promote the use of waste heat from data centers, including lifting a price cap that limited incentives for use of surplus waste heat and requiring data centers over 100 kW to assess the feasibility of heat recovery.[11-13]

ii. China (~28% of global data center capacity)[1]



The Chinese government has adopted a range of policies to improve the energy efficiency and reduce the environmental impact of data centers. All new large data centers were required to achieve a PUE ≤1.3 by the end of 2023.[14,15] The Chinese government aims to reduce the average PUE of data centers nationwide to 1.5 by the end of 2025.[16,17] The central government's July 2024 Special Action Plan for Green and Low-carbon Development of Data Centers calls for increasing data centers' renewable energy utilization by 10% each year.[17,18] Some provinces, including Inner Mongolia, encourage pairing data centers with renewable power sources.[19,20]

A signature policy is the "Eastern Data, Western Computing" initiative launched in 2022. To serve demand in dense eastern cities, Beijing is incentivizing new hyperscale data center clusters in western provinces where wind, solar and hydro power (as well as land) are more plentiful. By shifting workloads west, the Chinese government aims to decouple data growth from coal-heavy grids in the east.[21,22]

iii. United States (~47% of global data center capacity)[1]

On July 23, 2025, the Trump administration released its AI Action Plan and three related executive orders. These policy documents call for (1) rapid construction of new data center capacity in the United States, (2) streamlined environmental permitting, (3) priority interconnection of dispatchable power sources, including geothermal and nuclear power, (4) construction of data centers and power generation for data centers on federal lands and (5) greater use of AI to accelerate environmental reviews, among other policies.[23,24]

Many US government policies change significantly from administration to administration. Data center energy and environment policies are no exception. Although the Trump administration's strong support for new data center construction, geothermal power, nuclear power and the use of AI to accelerate environmental reviews are all broadly consistent with Biden administration policies, the current US administration takes a very different approach than its predecessor with respect to several issues including climate change, energy efficiency, solar power and wind power:

- The Biden administration directed all federal facilities, including data centers, to procure 100% carbon-free electricity by 2030. The Trump administration reversed this directive.[25,26]

- For many years, federal agencies ran voluntary energy efficiency programs related to data centers, such as ENERGY STAR for Data Centers (an efficiency certification) and DOE's Better Buildings Data Center Challenge. These programs are being cut significantly or eliminated under the Trump administration.[27]

- The Biden administration strongly supported solar and wind power, including for data center development.[28] The Trump administration does not.

Many US states have laws and policies related to data center energy use and environmental impacts. California, for example, offers incentives for data centers to use renewable energy and requires data centers to meet energy efficiency standards and disclose greenhouse gas emissions.[29-31] Washington exempts data centers that meet green building standards from sales and use taxes.[32] In June 2025, the Texas legislature passed a law requiring data centers and other large electric loads to pay for grid upgrades related to their power demand and to enable remote disconnection for grid operators to use in emergencies.[33-35]

iv. Japan (~4% of global capacity)[1]

In 2022, Japan's Energy Conservation Act was expanded to cover data centers, introducing mandatory energy management and reporting requirements for large facilities. Operators must report PUE and other metrics to Ministry of Economy, Trade and Industry (METI) annually and are expected to meet a "benchmark" PUE target of 1.4—a value set based on performance of the best 15% of facilities. Data centers that fail to make adequate efficiency improvements receive guidance, corrective orders and ultimately public disclosure or fines. Facilities achieving the PUE target may be recognized and eligible for energy-efficiency subsidies. This functions like a "top runner" program for data centers, continuously tightening as more operators hit the benchmark.[36,37]

Japan's "GX (Green Transformation) 2040 Vision" encourages locating data centers near low-carbon energy hubs (e.g., offshore wind farms and nuclear power plants). Announced in late 2024, this policy provides incentives (such as lowered electricity costs and tax rates) for companies to build data centers in regions with substantial renewable power.[38]

v. India (~3% of global capacity)[1]

The Indian government recognizes data centers as critical infrastructure, offering data center tax waivers and streamlined regulatory processes. State-specific policies encourage data center development as well, with some states offering additional incentives for data centers to use renewable energy.[39-41] Tamil Nadu, for example, provides a tax waiver to data centers that buy at least 30% of their power from renewable sources.[42] In 2010, the Bureau of Energy Efficiency (part of India's Ministry of Power) and the Confederation of Indian Industries released Energy Efficiency Guidelines and Best Practices for Indian Datacenters.[43] (The Indian Green Building Council (IGBC) has been promoting an ideal PUE of 1.4 to 1.5.)[44]

vi. United Kingdom (~2.5% of global capacity)[1]

The UK government requires data center operators to report their energy use and greenhouse gas emissions. Data center operators receive discounts on the Climate Change Levy—a tax on business energy use—if they meet energy efficiency targets.[45-47] These measures are part of broader efforts to align the sector with the UK's net-zero emissions target by 2050. To support this transition, the UK government has introduced the AI Energy Council, a cross-sector platform aimed at managing the energy demands of AI and data centers while meeting clean energy targets.[48]

vii. Singapore (~2% of global capacity)[1]

Singapore is a critical data center hub in Asia and one of the first to tie data center expansion to energy and environment issues. From 2019-2022, the Singapore government imposed a moratorium on new data centers due to power and land constraints. In 2022, the government lifted the moratorium but imposed strict standards under its Pilot Data Centre Sustainability Call. New facilities must meet "best-in-class" efficiency standards, including a design PUE of 1.3 or lower and obtain Green Mark Platinum certification. In 2024, the Singapore government released a Green Data Centre Roadmap. The Roadmap includes plans to add 300 MW of data center capacity, with 200 MW allocated to operators who use green energy. Singapore's data center best practices and building codes also include water efficiency measures (measured by WUE).[49-53]

viii. Gulf States (~1% of global capacity, but growing rapidly)[54]

Under its Vision 2030 program, the Saudi government is investing in large-scale renewable energy projects to support data center operations, including the $5 billion net-zero AI data center being developed at Neom's Oxagon hub. The Saudi government is also encouraging the growth of green data centers through tax incentives, grants and investments in renewable energy infrastructure.[55] The UAE government has established the National Team for Reviewing the Impact of Data Centers on the Energy Sector, which includes representatives from various ministries and digital authorities. This team is tasked with analyzing the impact of data centers on energy demand, evaluating the local market, and developing federal policies to regulate the operation of data centers.[56]

ix. Rest of the world

- **Australia and New Zealand (~1.5% of global capacity).**[1] Australia's government mandates that any data center hosting public sector workloads must achieve a PUE of 1.4 and a 5-star rating under the National Australian Built Environment Rating System.[57] New Zealand has no mandatory energy or emissions standards for data centers, but its clean electric grid (80% renewables) has attracted interest from hyperscalers committed to low-carbon data center development.[58,59]

- **South Korea (<1% of global capacity).**[1] South Korea's government requires new data centers to meet energy efficiency criteria as part of their power connection approvals.[60]

- **Southeast Asia (~1% of global capacity).**[1] The Malaysian government offers tax incentives for data centers that achieve several sustainability metrics, including a PUE of 1.4 and WUE of 2.2 m^3/MWh.[61,62] The Indonesian government announced a Green Data Center policy in 2022 and is exploring public-private partnerships to develop green data centers.[63,64]

- **South America, Central America and Caribbean (~1.5% of global capacity).**[1] Few countries in the region have policies specifically addressing data center sustainability, although many have policies promoting renewable energy and energy efficiency that affect data centers. The Mexican government provides subsidies for the use of renewable energy by data centers.[65]

- **Africa (~1% of global capacity).**[1] Africa's data center sector is expanding, especially in Kenya (where 90% of the power is from renewable sources) and South Africa. Governments in both countries are facilitating renewable power projects to meet the demand for cleaner energy in data centers.[66]

x. International organizations

The International Energy Agency (IEA) has emerged as a key global institution for collecting and sharing information on data centers' energy use and environmental impacts. The IEA's Energy and AI Observatory, launched in early 2025, is an important knowledge hub contributing to global policy development.[1] The International Technology Union also does work in this area.[67]

Under the Basel Convention (a treaty with more than 191 member countries), countries may not export e-waste to countries without authorized recycling facilities.[68]

B. Topics

Government policies with respect to data centers' energy and environmental impacts touch a wide range of topics. This section summarizes policies with respect to data centers' (1) disclosure of energy use and environmental impacts, (2) impacts on the adequacy of power supplies, (3) energy efficiency, (4) renewable energy use, (5) greenhouse gas emissions, (6) emissions of local air pollutants, (7) water use and (8) e-waste.

i. Disclosure

Many governments around the world require data center operators to disclose information on energy use, water use and greenhouse gas emissions. Some governments require corporate-level disclosures; others require disclosure at the facility level. Some governments release the disclosures publicly; others keep the disclosures private. Examples of such disclosure requirements include the following:

- The EU's Energy Efficiency Directive (EED Article 12) requires data center operators with an information technology (IT) power demand of at least 500 kW to monitor and report on their energy and water use at a facility level. This information is published in a public database.[17,69,70] The EU's Corporate Sustainability Reporting Directive (CSRD) requires data center operators to regularly report on energy and environmental indicators, including PUE, renewable power as a percentage of total power use (Renewable Energy Factor or REF), and carbon dioxide emissions per unit of energy consumed by IT equipment (Carbon Usage Effectiveness or CUE).

- Germany's Energy Efficiency Act (Energieeffizienzgesetz, EnEfG) requires data center operators to report annually to the government on both energy efficiency and total water consumption.[71,72]

- China's "Eastern Data, Western Computing" initiative requires participating data centers to report energy efficiency metrics to authorities.[73,74] Under the Special Action Plan for Green and Low-Carbon Development of Data Centers, Chinese data center operators must track and improve indicators, including PUE and renewable energy use. Facility-by-facility data are not made public.[19]

- Japan's Energy Conservation Act requires data center companies that exceed certain size thresholds to submit energy reports to METI.[75]

- California's Climate Corporate Data Accountability Act (SB 253) requires businesses with over $1 billion in annual revenues operating in California to report their Scope 1, 2 and 3 greenhouse gas emissions. This affects many data centers.[76,77]

Data center disclosure requirements are expanding to a growing number of jurisdictions and topics. Broad sustainability reporting mandates (covering energy and water use, greenhouse gas emissions and other topics) are becoming standard for large companies in Europe and the Asia-Pacific, requiring data center operators to publicly disclose their energy use and environmental impacts. A growing number of jurisdictions now require data centers to report energy use at the facility level. Several jurisdictions, including the EU and California, now require data center operators to report Scope 3 emissions (indirect emissions throughout the entire supply chain).[17,36]

ii. Impacts on adequacy of power supply

The rapid growth of data center electricity demand is placing strains on power grids in many regions, raising concerns about the adequacy of electric power supplies (known as “resource adequacy”). Policy responses include grid connection restrictions and grid upgrade requirements.

a. **Grid connection restrictions.** Several jurisdictions with grid constraints have paused or limited new data center connections:

    - The Singapore government imposed a moratorium on new data center projects from 2019 to 2022, citing constraints on energy resources. The government has since lifted the pause but now grants approvals only through a pilot program for a limited number of new projects that can demonstrate best-in-class efficiency and have designs that minimize burden on the grid.[78]

    - In 2022, the Irish government announced that it would no longer accept interconnection applications for data centers in Dublin.[79,80]

    - The city of Amsterdam has a near-total ban on new data centers in certain areas and strict requirements related to land use and grid impact for the few locations where data center development is permitted.[81]

    - In Northern Virginia, several county and municipal governments have imposed restrictions on new data center development. In March 2025, the Loudoun County Board of Supervisors eliminated “by-right” development of data centers (which means special exceptions are now required to develop any new data centers in the county). In September 2024, the Fairfax County Board of Supervisors approved an ordinance imposing strict locational, design, and noise reduction requirements on new data centers. [82,83]

b. **Grid connection requirements.** Some jurisdictions require data centers to fund infrastructure upgrades as a condition for interconnection. The Texas legislature passed such a law in June 2025 (also requiring that data centers allow remote disconnection in the event of grid emergencies).[33] In Ireland, the electricity regulator has proposed that new data centers will be required to provide generation and/or storage capacity to match the requested connection capacity.[84]

iii. Energy efficiency

Improving energy efficiency is a core objective of data center policies in many jurisdictions around the world. The policy toolkit includes energy efficiency standards, fiscal incentives and requirements to use waste heat. Many energy efficiency policies focus on PUE as a key metric.

a. **Energy efficiency standards**. Many jurisdictions require data centers to meet energy efficiency targets. PUE is overwhelmingly the most common metric

- The European Union sets PUE targets for new data centers (1.3 in cold climates and 1.4 in hot climates) as part of the Climate Neutral Data Centre Pact (a voluntary pact established in part to avoid more stringent regulation).[85]

- Under Germany's Energy Efficiency Act (Energieeffizienzgesetz, EnEfG), all data centers must progressively improve PUE. Starting July 1, 2026, new data centers must achieve PUE ≤1.2—one of the world's most stringent data center energy efficiency standards.[86]

- The Chinese government has a national average PUE target of 1.5 or below for domestic data centers.[16] The "Eastern Data, Western Computing" initiative funnels new data centers to western regions with ample land, significant renewable power and cool temperatures. These new data centers are required to be highly efficient (PUE well below 1.5) and use free cooling when possible.[87]

- The Singapore government has a PUE target of ≤1.3 at 100% IT load over the next 10 years.[51,88]

- The Australian government requires any data center housing federal agencies' data to achieve a 5-star rating under the National Australian Built Environment Rating System (NABERS), which includes a requirement to achieve a PUE roughly equivalent to 1.4.[17,89]

b. **Fiscal incentives**. Several governments offer reduced tax rates for data centers that meet energy efficiency benchmarks. The French government, for example, offers a reduced electricity tax rate for operators of energy efficient data centers that adopt certain best practices.[90] The UK government offers a discount on the Climate Change Levy (an energy tax) to data center operators that commit to meeting energy efficiency targets (often PUE or energy usage per rack).[91,92]

c. **Waste heat reuse.** Several jurisdictions impose requirements related to waste heat reuse in data centers. (See Chapter 2.4.) The European Union requires data centers to reuse waste heat where technically and economically feasible.[93] The German government imposes steadily increasing waste heat reuse obligations on new data centers in the years ahead.[86,94]

Governments take a range of other approaches to improving data center energy efficiency. The Japanese government encourages data centers to be energy efficient with technical guidance and periodic reviews.[36,37] In the United States, some state public utility commissions encourage utilities to run energy efficiency incentive programs targeting data centers. In California and Oregon, for example, power companies offer rebates for installing more efficient cooling or IT equipment.[36,95]

Energy efficiency policies for data centers are a mix of carrots and sticks, including energy efficiency standards (mainly PUE) and tax breaks. There is a trend toward more ambitious approaches, with steadily lower requirements for PUE and higher requirements for waste heat reuse (although the current US federal government is moving in the opposite direction—an important exception).

iv. Renewable energy

A growing number of governments require data centers to purchase renewable power or promote construction of data centers near renewable energy sources. Examples include the following:

- The German government requires data centers to cover 50% of their energy needs with unsubsidized renewable electricity, increasing to 100% by 2027.[86]
- The Chinese government aims to increase data centers' renewable energy utilization rate by 10% annually. Policies support direct transmission of renewable electricity to data centers and establishment of "green power industrial parks" with dedicated renewable sources and storage.[16,17,19] Some provincial and municipal governments in China now require new data centers to include renewable energy integration plans (like solar panels on site or agreements to buy wind power) when seeking approval.[96]

- The Irish government promotes co-location of renewable generation facilities with data centers and encourages advanced energy storage solutions. (85% of data centers in Ireland reportedly use renewable energy sources.[96])
- The Japanese government's GX 2040 Vision supports relocating tech industries near carbon-neutral energy hubs, including offshore wind farms and nuclear plants.[38]

v. Greenhouse gas emissions

A few jurisdictions have adopted greenhouse gas goals for data centers. Many jurisdictions have broader policies that affect data centers' greenhouse gas emissions, such as carbon taxes, emissions trading programs or economy-wide carbon neutrality goals. (See Chapter 3 for a broader discussion of data center greenhouse gas emissions.) The energy efficiency and renewable energy policies described above generally help reduce greenhouse gas emissions from data centers as well.[c]

a. **Data center carbon neutrality goals.** The Japanese government's Green Growth Strategy calls for data centers to be carbon neutral by 2040.[97] A European Commission white paper on Shaping Europe's Digital Future (February 2020) says that data centers "can and should become climate neutral by 2030."[98]

b. **Broader climate change programs.** Globally, at least 80 jurisdictions have implemented carbon taxes or emissions trading programs.[99] At least 139 countries representing more than 76% of global GDP have pledged carbon neutrality.[100,101] Although these programs generally do not mention data centers specifically, they play an important role in reducing data centers' greenhouse gas emissions – in particular from offsite power plants that sell electricity to data centers (Scope 2 emissions) and from manufacturers of iron, steel, cement, electronic equipment and other products in the data center supply chain (Scope 3 emissions). (See Chapters 3.2 and 3.3 of this Roadmap.)

vi. Local air pollution

Many data centers rely on diesel-fired generators for backup power. These generators are used sparingly but emit nitrogen oxides (NOx), particulate matter and other local air pollutants when used. Citizen groups are increasingly raising concerns about air pollution related to backup generators at data centers, especially in Virginia (which has the largest concentration of data centers in the world).[102-105] Governments have a range of policies to address this problem, including emissions standards, zoning rules and policies to promote cleaner backup power.

a. **Emissions standards.** In the United States, emergency backup generators are subject to relatively lenient federal emissions standards, provided they run during outages and are limited testing only.[106-108] In the EU, data center generators are generally subject to emissions limits under the Medium Combustion Plant Directive (MCPD). If a data center has a large diesel capacity that runs beyond emergency use, it could be regulated under the Industrial Emissions Directive (IED), which has stricter permitting rules.[109]

c In some situations, energy efficiency and renewable energy policies might not reduce greenhouse gas emissions from data centers. This could happen if (a) energy efficiency improvements lead only to more computation at a data center and not to less energy consumption (a 100% rebound effect) or (b) renewable power supplies in a region are fully utilized and unable to expand, so requiring data centers to use renewable power simply displaces businesses that would otherwise have used renewable power, forcing them to use higher-carbon power such as coal or natural gas instead. These situations are by no means impossible, but in most cases greater energy efficiency and more renewable power will reduce greenhouse gas emissions.

b. **Zoning.** The clustering of data centers has led to targeted zoning and permitting rules to mitigate hotspots of diesel exhaust and noise. In 2024, Fairfax County, Virginia limited the size of data centers in certain zones and included stricter site plan requirements (e.g., placing generators away from residential property lines) to address air pollution and noise concerns.[17,110]

Some governments are funding pilots and demonstrations to develop cleaner backup generators. The EU has funded projects exploring hydrogen fuel cell generators for large data centers, and some European data centers have tested fuel cells in place of diesel generators.[111]

vii. Water use

Governments are addressing data centers' water use with a range of policies including disclosure requirements, efficiency or effectiveness targets, and required use of reclaimed water.

a. **Disclosure requirements.** The European Union requires annual disclosure of data center water withdrawals and consumption (EED Article 12).[2,112] The German government requires data centers to report total water consumption and water use efficiency indicators (Energy Efficiency Act).[72,113]

b. **WUE requirements.** A growing number of governments are imposing requirements related to WUE or similar water use measures. (For a broader discussion of WUE and data center water use, see Chapter 5.) The Chinese government, for example, requires data centers serving government needs to use no more than 2.5 L/kWh of IT energy.[114,115] Governments in Beijing, Ningxia and Gansu mandate higher water use efficiency for data centers and are phasing out those with low water efficiency.[19] The Singapore government aims to reduce the median WUE for data centers from 2.2 m^3/MWh in 2021 to 2.0 m^3/MWh or lower by the early 2030s.[51,88]

c. **Reclaimed water.** In the western United States, some municipalities require use of reclaimed water for cooling and/or stipulate maximum water withdrawal amounts in permits.[116] The Loudon County, Virginia government supports a pipe network that provides reclaimed water to data centers.[117] Singapore encourages use of reclaimed water as well.[118]

Other policy tools for managing data center water use include the requirement to submit water management plans (such as in Singapore).[88] Water use policy for data centers is evolving, with more disclosure requirements, as well as incentives or mandates pushing facilities to adopt cooling solutions that drastically reduce freshwater consumption.[119]

viii. E-waste

The management of electronic waste (e-waste) generated by data centers is a growing concern. (See Text Box on E-Waste.) Key policies include the following:

a. **Hazardous materials restrictions.** Many jurisdictions—including the EU, China, South Korea and California—limit or prohibit the use of toxic substances, such as lead, cadmium and brominated flame retardants, in electronics.[120]

b. **E-waste recycling mandates.** Many jurisdictions have laws requiring electronic equipment to be collected and recycled at end-of-life. This obligation is often imposed on the manufacturer or importer (an approach known as "extended producer responsibility"). A leading example is the EU's Waste Electrical and Electronic Equipment Directive, which obligates manufacturers of electronic goods, including servers and IT hardware, to finance and facilitate the take-back and recycling of e-waste. In 2021, Singapore introduced an E-Waste Management System that places responsibility on producers and importers to collect and recycle e-waste. At least 25 US states have legislation addressing e-waste, with a range of provisions including mandatory recycling requirements and landfill bans.[121-125]

c. **Import and export bans.** China historically was a destination for global e-waste, but in 2017 it banned e-waste imports. Other countries, including Thailand, have adopted similar laws.[126-128] Under the Basel Convention, countries may not export e-waste to countries without authorized recycling facilities. More than 190 countries are Parties to the Basel Convention.[68,129]

C. Impacts of Government Policies

Government policy makers are paying increasing attention to data centers' energy use and environmental impacts, due largely to the surge in data center construction in recent years.[17,47,130,131] In the United States, for example, more than 500 bills related to data center energy use were introduced in state legislatures in the first seven months of 2025—an enormous increase in the number of such bills from prior years.[132] A series of new measures are under consideration in the European Union, Japan and elsewhere.[9,17]

Yet the literature evaluating the impact of data center energy and environment policies is sparse. A 2021 report for the Australian government reviewed studies on energy efficiency of data centers, concluding that "There is little evidence on the effectiveness of the metrics, policies and certifications described in this report."[133] A 2024 report for the IEA on policies regulating data center energy use found that "most have not been in place long enough for their effectiveness to be evaluated."[134] A search for this Roadmap yielded very few studies evaluating the impact of data center energy and environment policies.

Several factors make evaluating the impacts of government policies in this area difficult.

First, data are limited. Data center owners and operators do not typically disclose energy use at the facility level. Some companies volunteer such information for some sites, but facility-level disclosure of energy use at data centers is not standard industry practice. (Facility-level disclosure of water use is more common.)[134-137] Some jurisdictions require disclosure of data center energy use at the facility level, although the information may be closely held by governments and not available to the public. (See Section B.i of this chapter.)

Second, data center operators have incentives to improve energy and environmental performance independent of government policies. Energy efficiency measures can reduce costs and increase output. Use of solar and wind power can speed development of data centers and, in many locations, cut costs as well.[138] "Green" policies can help improve public acceptance of data centers, which are becoming increasingly controversial in some jurisdictions, and head off regulation through voluntary action.[139,140] These factors and others can make it difficult to determine which trends result from government policies and which result from other factors.
Third (and related to the foregoing), the data center industry has longstanding voluntary initiatives to promote clean energy and reduce environmental impacts.[141,142] Indeed the pledges and actions of some of the industry's leading companies with respect to renewable power and greenhouse gas emissions are significantly more ambitious than the policies of many governments.[143,144] (See Chapter 3 and text box on Voluntary Industry Initiatives in this Roadmap.) Figuring out when government policies

are additive to voluntary initiatives can sometimes be challenging.

Recognizing these challenges, this section considers three questions with respect to the impacts of data center energy and environmental policies.

First, do current government policies to promote data center energy efficiency have significant impacts?

There are reasons to be skeptical:

- Most government policies on data center efficiency focus on PUE; however, PUE is a very limited metric. PUE measures the energy efficiency of cooling, lighting and other systems that support servers at a facility, but not the energy use per unit of computation or work.[145] PUE is somewhat analogous to measuring the energy used by a car's air conditioning, lighting and other support systems, but not the car's miles-per-gallon or liters-per-kilometer.

- In addition, as noted above, many data center operators have significant incentives to improve energy efficiency independent of regulatory requirements. This is especially the case where electricity costs are high (such as in much of Europe) and where access to power is the biggest short-term constraint to data center development (such as in the United States today).

- Finally, rebound effects (Jevon's Paradox) significantly complicate assessments of the impact of energy efficiency policies related to data centers. More energy efficient equipment may lead to more computation at a site, not less energy consumption—indeed, it likely already has. Some important objectives of policies to improve the energy efficiency of data centers may not be achieved in whole or in part due rebound effects, although the topic needs considerably more study.[146-148] (See Chapter 1.)

Despite the foregoing, some government policies may help improve energy efficiency at data centers.[140] The US government's ENERGY STAR program has been praised for helping data center owners and operators improve energy efficiency and save money.[149] The EU's Code of Conduct for Energy Efficiency in Data Centers Energy Efficiency may have had similar impacts.[111] The 2024 IEA study noted above found that stricter PUE limits and other government policies would likely reduce energy use at data centers in the years ahead.[36]

Further study on this topic is needed to help guide policymakers.

Second, are governments using the right metrics in data center energy and environment policies?

Again, there are reasons to be skeptical.

A wide range of metrics are available to assess the energy and environmental impacts of data centers. Some of these metrics have been discussed and analyzed by experts for many years:

- In 2010, Japan's Green IT Promotion Council proposed a metric it labeled Data Center Performance per Energy (DPPE), which considers factors beyond the energy efficiency of a facility's cooling, lighting and other support systems, including the energy efficiency of the IT equipment (IT Equipment Energy Efficiency or ITEE), how effectively IT equipment in the data center is being used (IT Equipment Utilization or ITEU), and the proportion of renewable or green energy sources used (Green Energy Coefficient or GEC).[150,151]

- The Green Grid, an industry organization committed to resource efficient data centers since 2007, recently proposed two new metrics: Data Center Resource Effectiveness (DCRE), which "integrates multiple factors that include water usage, geographic consideration, and facility energy effectiveness," and IT Work Capacity (ITWC), which "equips operators to boost energy efficiency [and] supports compliance with global regulations."[152]

- Other relevant metrics include the percentage of energy captured for reuse (Energy Reuse Factor or ERF) and two metrics mentioned in section B.i of this chapter—renewable power as a percentage of total power use (Renewable Energy Factor or REF) and carbon dioxide emissions per unit of energy consumed by IT equipment (Carbon Usage Effectiveness or CUE).[153]

These metrics and others are important for capturing the full range of data centers' energy and environmental impacts. Yet despite a rich literature on these metrics, few government policies refer to metrics other than PUE. There are exceptions, including the EU's Corporate Sustainability Reporting Directive (CSRD), which requires reporting on metrics including REF and CUE, and the growing number of policies that refer to WUE. But the use of metrics other than PUE is not common, frequently limiting the impact of government policies.

Third, what impact do broad economy-wide policies have on data centers' energy and environmental performance?

Although rigorous studies are lacking, broad economy-wide policies appear to have significant impacts on data centers' energy and environmental impacts. Low-carbon goals for the power sector, for example, are central to reducing the carbon footprint of data centers buying electricity from the grid. As noted above, at least 80 jurisdictions have implemented carbon taxes or emissions trade programs[99]; at least 139 countries representing more than 76% of global GDP have pledged carbon neutrality.[100,101] Virginia, which hosts the world's largest data center hub, requires utilities to supply 100% carbon-free electricity by 2045.[154] Policies to speed the permitting of solar and wind power help data center developers meet their voluntary renewables procurement goals.[155]

Green finance policies are another example. As data centers adopt more sustainable practices, they have leveraged green finance products to help access financing and reduce costs.[156] Equinix, the largest global data center and colocation provider of cloud computing, has issued approximately $5.6 billion of green bonds to build sustainable data centers, including a colocation data center in Paris.[157] Fund manager and property owner Areim has raised a total of $971 million to support development of sustainable data centers in Nordic countries through its Areim DC Fund.[158] Singapore-based DayOne Data Centers secured $3.58 billion (SG$4.6 billion) in green financing to support the development of data centers in Malaysia.[159]

Government policies with respect to data centers' energy and environmental impacts are evolving rapidly. Many of these policies have important goals, including managing data centers' considerable energy footprint and minimizing adverse environmental impacts from data center development. But well-intentioned policies can sometimes fail to produce the desired results.[160,161] More work is needed to systematically assess data center policies and refine them for the years ahead.

D. Recommendations

1. *Governments should **collect and share data** on data centers' energy use and environmental impacts.*

2. *Governments should **build capacity to better understand fast-moving trends** with respect to data centers' energy use and environmental impacts.*

3. *Governments should **use a broad set of metrics when regulating data centers' energy use and environmental impacts**, including not just PUE.*

4. *Governments should **assist the rapid buildout of clean power capacity** to help meet growing data center power demand.*

5. *When governments procure data center services, they should **require vendors of data center services to disclose their energy use, water use and greenhouse gas emissions.**[57]*

6. *IEA Member governments should **expand the IEA's Energy and AI Observatory,** devoting additional resources to monitoring and reporting on data centers' energy use and environmental impacts, as well as policy trends with respect to data centers around the world. The Clean Energy Ministerial should **expand CEM's work on data centers** under its power sector and artificial intelligence initiatives.*

E. References

1. International Energy Agency (IEA). Energy and AI Observatory; IEA, Paris, France, https://www.iea.org/data-and-statistics/data-tools/energy-and-ai-observatory?tab=Energy+for+AI (Accessed August 2025 (Last updated June 2025)).

2. European Union – EUR-Lex. Directive (EU) 2023/1791 of the European Parliament and of the Council of 13 September 2023 on energy efficiency and amending Regulation (EU) 2023/955 (recast) (Text with EEA relevance) (Document 32023L1791); Publications Office of the European Union, Luxembourg, Luxembourg, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ%3AJOL_2023_231_R_0001&qid=1695186598766 (Accessed August 2025).

3. Tim Hickman, Thomas Burmeister, Erasmus Hoffmann, Dr. Petra Karin Iffert (née Kistner) & Aishwarya Jhal. Energy efficiency requirements under the EU AI Act; White & Case LLP, New York, New York, https://www.whitecase.com/insight-alert/energy-efficiency-requirements-under-eu-ai-act (2025).

4. Yevgeniy Sverdlik. Data Center Knowledge, San Francisco, California, https://www.datacenterknowledge.com/sustainability/preempting-regulation-eu-operators-strike-green-data-center-pact (2021).

5. Bundesministerium der Justiz und für Verbraucherschutz (BMJV). Gesetz zur Steigerung der Energieeffizienz in Deutschland1 (Energieeffizienzgesetz - EnEfG); BMJV, Berlin, Germany, https://www.gesetze-im-internet.de/enefg/BJNR1350B0023.html (2023).

6. Andy Lawrence. Resiliency vs low PUE: Could regulation spark innovation?; Data Center Dynamics, London, United Kingdom, https://www.datacenterdynamics.com/en/opinions/resiliency-vs-low-pue-could-regulation-spark-innovation/ (2024).

7. Rostyslav Telyatnykov, Thomas Burmeister & Dr. Petra Karin Iffert (née Kistner). Data center requirements under the new German Energy Efficiency Act; White & Case LLP New York, New York, https://www.whitecase.com/insight-alert/data-center-requirements-under-new-german-energy-efficiency-act (2023)

8. Commission for Regulation of Utilities (CRU). New Electricity Connection Policy for Data Centres; CRU, Dublin, Ireland, https://www.cru.ie/about-us/news/new-electricity-connection-policy-for-data-centre/ (2025).

9. Zachary Skidmore. Ireland plans legislation for data centers to build and run own power lines - report; Data Center Dynamics Ltd (DCD), London, United Kingdom, https://www.datacenterdynamics.com/en/news/ireland-plans-legislation-for-data-centers-to-build-and-run-own-power-lines-report/ (2025).

10. Barry Gross. What's the future for Ireland's data centres?; Simmons & Simmons LLP, London, United Kingdom, https://www.simmons-simmons.com/en/publications/cm9ipax7a0056upcsuyiymkc2/what-s-the-future-for-irelands-data-centres (2025).

11. Datacenter Forum. Denmark removes price cap on surplus heat; Mariestad, Sweden, https://www.datacenter-forum.com/datacenter-forum/denmark-removes-price-cap-on-surplus-heat (2025).

12. Hanne Kokkegård. Utilize waste heat from data centers in district heating; Technical University of Denmark (DTU), Kongens Lyngby, Denmark, https://www.dtu.dk/english/news/all-news/nyhed?id=b2e4c8f0-0c62-437d-afbf-d2a17c376ffa (2022).

13. Juan Jesús Jerez Monsalves, Claire Bergaentzlé & Martijn Backer. Regulatory Frameworks and Business Models for Data Centres Integrated to the Energy System: A comprehensive review of the rules and incentives affecting the provision of flexibility and waste-heat recovery by data centres in Denmark; Technical University of Denmark, Kongens Lyngby, Denmark, https://doi.org/10.11581/dtu.00000242 (2022).

14. Ministry of Industry and Information Technology of the People’s Republic of China (MIIT). Notice of the Ministry of Industry and Information Technology on the Issuance of the "Three-Year Action Plan for the Development of New Data Centers (2021-2023); MIIT, Beijing, China, https://www.miit.gov.cn/jgsj/txs/wjfb/art/2021/art_12cc04dc9daf4d57a7038811a57383b6.html

15. Peter Judge. Chinese government calls for 200 exaflops of data center compute by 2023; Data Center Dynamics, London, United Kingdom, https://www.datacenterdynamics.com/en/news/chinese-government-calls-for-200-exaflops-of-data-center-compute-by-2023/ (2021).

16. Xinhua. China sets green targets for data centers; English.gov.cn by The State Council of the People’s Republic of China, Beijing, China, https://english.www.gov.cn/news/202407/24/content_WS66a0b167c6d0868f4e8e96ba.html (2024).

17. Wylie Wong. Data Center Regulation Trends to Watch in 2025; Data Center Knowledge, San Francisco, California, https://www.datacenterknowledge.com/regulations/data-center-regulation-trends-to-watch-in-2025 (2024).

18. Ministry of Industry and Information Technology, National Data Administration, National Development and Reform Commission & National Energy Administration. Special Action Plan for Green and Low-carbon Development of Data Centers; China Water Risk (CWR), Central, Hong Kong, https://cwrrr.org/regulation/special-action-plan-for-green-and-low-carbon-development-of-data-centers/ (2024).

19. Yuan Ye. Explainer: How China is managing the rising energy demand from data centres; Carbon Brief Ltd, London, United Kingdom, https://www.carbonbrief.org/explainer-how-china-is-managing-the-rising-energy-demand-from-data-centres/ (2025).

20. General Office of the People's Government of Inner Mongolia Autonomous Region. General Office of the People's Government of Inner Mongolia Autonomous Region Supporting the Green Computing Power of the Horqin Cluster; Inner Mongolia Autonomous Region, China, https://www.nmg.gov.cn/zfbgt/zwgk/zzqwj/202406/t20240621_2527949.html (2024).

21. Xuemei Xie, Yuhang Han & Hao Tan. Greening China’s digital economy: exploring the contribution of the East–West Computing Resources Transmission Project to CO_2 reduction. Humanities and Social Sciences Communications 11, 466 (2024). https://doi.org/10.1057/s41599-024-02963-0.

22. China Daily. Computing power offers new chances for tech firms; English.gov.cn The State Council the People's Republic of China, Beijing, China. https://english.www.gov.cn/news/202306/19/content_WS648fc5a8c6d0868f4e8dcfa7.html

23. The White House. America’s AI Action Plan; Washington, DC, https://www.whitehouse.gov/wp-content/uploads/2025/07/Americas-AI-Action-Plan.pdf (2025).

24. The White House. Accelerating Federal Permitting of Data Center Infrastructure (Executive Order); Washington, DC, https://www.whitehouse.gov/presidential-actions/2025/07/accelerating-federal-permitting-of-data-center-infrastructure/ (2025).

25. Sabin Center for Climate Change Law. President Issues Executive Order Revoking Federal Sustainability Plan; Columbia University, New York, New York, https://climate.law.columbia.edu/content/president-issues-executive-order-revoking-federal-sustainability-plan-0 (Accessed August 2025).

26. Donald Trump. Executive Order 13834—Efficient Federal Operations; The American Presidency Project, Santa Barbara, California, https://www.presidency.ucsb.edu/documents/executive-order-13834-efficient-federal-operations (2018).

27. Robert Walton. Trump administration plans to end popular Energy Star program; Industry Dive, Washington, DC, https://www.utilitydive.com/news/trump-administration-epa-energy-star-program/747351/ (2025).

28. The White House. Executive Order on Advancing United States Leadership in Artificial Intelligence Infrastructure; Washington, DC, https://bidenwhitehouse.archives.gov/briefing-room/presidential-actions/2025/01/14/executive-order-on-advancing-united-states-leadership-in-artificial-intelligence-infrastructure/ (2025).

29. California Public Utilities Commission (CPUC). Homepage; CPUC, Sacramento, California, https://www.cpuc.ca.gov (Accessed August 2025).

30. Legislative Council Staff. State-by-State Data Center Regulation; North Dakota Legislative Council, Bismarck, North Dakota, https://ndlegis.gov/sites/default/files/resource/research-document/state-by-state-data-center-regulation-january-2025.pdf (2025).

31. California Energy Commission (CEC). Computer Rooms & Data Centers (Fact Sheet); CEC, Sacramento, California, https://www.energy.ca.gov/sites/default/files/2023-09/2022_CEC-Computer_Room_and_Data_Centers_ADA.pdf

32. Washington State Legislature. RCW 82.08.986: Exemptions—Eligible server equipment—Eligible power infrastructure—Computer data centers.; https://app.leg.wa.gov/Rcw/default.aspx?cite=82.08.986 (2024).

33. Texas State Legislature. Texas Senate Bill 6: Relating to the planning for, interconnection and operation of, and costs related to providing service for certain electrical loads and to the generation of electric power by a water supply or sewer service corporation.; Austin, Texas, https://legiscan.com/TX/bill/SB6/2025 (2025).

34. Robert A. James, Ashleigh K. Myers, Alicia M. McKnight, Sidney L. Fowler, M.C. Hammond & Polly Gomez. Texas SB6 Establishes New Transmission Fees and Interconnection Standards for Large Load Customers and Co-Located Loads Pillsbury Winthrop Shaw Pittman LLP, New York, New York, https://www.pillsburylaw.com/en/news-and-insights/texas-sb6-transmission-fees-interconnected-standards-large-load-customers-colocated-loads.html (2025).

35. Bill A Moore. Texas Legislature Adopts New Law Regarding Large Load Interconnection and Operation; Butler Snow LLP, Ridgeland, Mississippi, https://www.butlersnow.com/news-and-events/texas-legislature-adopts-new-law-regarding-large-load-interconnection-and-operation (2025).

36. EDNA Annex (Electronic Devices and Networks Annex) of the Technology Collaboration Programme on Energy Efficient End-Use Equipment (4E TCP). Policy development on energy efficiency of data centres; International Energy Agency (IEA), Paris, France, https://www.iea-4e.org/wp-content/uploads/2024/02/Policy-development-on-energy-efficiency-of-data-centres-draft-final-report-v1.05.pdf (2024).

37. Ministry of Economy Trade and Industry (METI). Guide to the Energy Conservation Act - FY 2023 Revised Edition; METI, Tokyo, Japan, https://www.enecho.meti.go.jp/category/saving_and_new/saving/media/data/shoene_tebiki_01.pdf (2023).

38. Zachary Skidmore. Japan unveils plan to relocate tech industries near low-carbon energy hubs; Data Center Dynamics, London, United Kingdom, https://www.datacenterdynamics.com/en/news/japan-unveils-plan-to-relocate-tech-industries-near-low-carbon-energy-hubs/ (2025).

39. Vibhuti Garg. India's power-hungry data centre sector at a crossroads; Institute for Energy Economics and Financial Analysis, Valley City, Ohio, https://ieefa.org/resources/indias-power-hungry-data-centre-sector-crossroads (2025).

40. Rajat Sethi, Raya Hazarika & Ananya Kumar. Data Centres in India: Opportunity and Incentives; S&R Associates, New Delhi, India, https://www.snrlaw.in/data-centres-in-india-opportunity-and-incentives/ (2023).

41. Mollshree Garg. India's data centre boom: Greener, smarter and poised to lead the world Environmental Resources Management (ERM) – India, Haryana, India, https://www.eco-business.com/press-releases/indias-data-centre-boom-greener-smarter-and-poised-to-lead-the-world/ (2025).

42. Information Technology Department Principal Secretary to Government of Tamil Nadu. Tamil Nadu Data Centre Policy 2021; Government of Tamil Nadu Information Technology Department, Chennai, Tamil Nadu, India, https://it.tn.gov.in/sites/default/files/2021-11/Connect%20BOOK.pdf (2021).

43. Confederation of Indian Industry and Bureau of Energy Efficiency (BEE). Energy Efficiency Guidelines and Best Practices in Indian Datacenters; Bureau of Energy Efficiency (BEE), Ministry of Power, Government of India, New Delhi, India, 2010).

44. Rashmi Singh. Green building regulations give impetus to sustainable data centers in India; Mongabay, Menlo Park, California, https://india.mongabay.com/2023/02/green-building-regulations-give-impetus-to-sustainable-data-centers-in-india/ (2023).

45. Environment Agency. Statutory guidance: Climate change umbrella agreement for the data centre sector Gov.UK, London, United Kingdom, https://www.gov.uk/government/publications/climate-change-umbrella-agreement-for-the-data-centre-sector (2014).

46. Pascal Holt. Navigating regulations and power challenges in the UK data center market; Data Center Dynamics, London, United Kingdom, https://www.datacenterdynamics.com/en/opinions/navigating-regulations-and-power-challenges-in-the-uk-data-center-market/ (2024).

47. Sean A. Lev, John A. Hodges & Jason E. Neal. How Global Data Center Regs May Influence U.S. Policies; HWG LLP, Washington, DC, https://hwglaw.com/2024/12/02/how-global-data-center-regs-may-influence-u-s-policies/ (2024).

48. Sophie Rice. UK AI Energy Council Addresses Data Centre Sustainability; Data Centre Magazine, Norwich, United Kingdom, https://datacentremagazine.com/technology-and-ai/how-uks-ai-energy-council-will-tackle-data-centre-emissions (2025).

49. Green Plan. Green Plan Overview: Our Sustainable Development Movement for a Better and Greener Future; Orchard Road, Singapore, https://www.greenplan.gov.sg/overview/ (Accessed August 2025).

50. Baker McKenzie. Global Compliance News Homepage (Blog); Chicago, Illinois, https://www.globalcompliancenews.com/ (Accessed August 2025).

51. Mark Wong. Singapore's Green Data Centre Roadmap – Representing a Necessary Intersection between Digital Infrastructure and Sustainability; Mayer Brown, Chicago, Illinois, https://www.mayerbrown.com/en/insights/publications/2024/08/singapores-green-data-centre-roadmap-representing-a-necessary-intersection-between-digital-infrastructure-and-sustainability (2024).

52. Brightlio. Homepage; Culver City, California, https://brightlio.com/ (Accessed August 2025).

53. Osmond Chia. Singapore to expand data centre capacity by at least one-third, pushes for green energy use; The Straits Times and Singapore Economic Development Board (EDB), News Centre, Singapore and Raffles City Tower, Singapore, https://www.edb.gov.sg/en/business-insights/insights/singapore-to-expand-data-centre-capacity-by-at-least-one-third-pushes-for-green-energy-use.html (2024).

54. Research and Markets: The World's Largest Market Research Store. GCC Data Center Portfolio Report 2025: Detailed Analysis of 106 Existing and 77 Upcoming Data Centers Across Bahrain, Kuwait, Oman, Qatar, Saudi Arabia and UAE; Research and Markets: The World's Largest Market Research Store and GlobeNewswire, Dublin, Ireland and Los Angeles, California, https://finance.yahoo.com/news/gcc-data-center-portfolio-report-080100291.html?guccounter=1&guce_referrer=aHR0cHM6Ly93d3cucGVycGxleGl0eS5haS8&guce_referrer_sig=AQAAADxnErlB4m9GHkRxPdJ8SKcUNPn1tr6WOZsOJuxn4y0RgITEAnuvzvXIBOdh5hEzfkyYzxGwCQYfp-C2SYy-VH9l_gEOknn8NovHVbw_9iJbMfKciLF9aW9GvBqTZDBmJSI7H9kbz6OklDcZdNQ_Rss9GE0qUIrsG39kQgXj75kF (2025).

55. Kristian Alexander. Digital Infrastructure, Strategic Power: The Gulf's Data Centre Boom; Observer Research Foundation Middle East (ORF ME), Dubai, United Arab Emirates, https://orfme.org/expert-speak/digital-infrastructure-strategic-power-the-gulfs-data-centre-boom/ (2025).

56. Emirates 24/7. National Team for Reviewing the Impact of Data Centers on Energy Sector Holds First Meeting; Dubai Media, Dubai, United Arab Emirates, https://www.emirates247.com/uae/national-team-for-reviewing-the-impact-of-data-centers-on-energy-sector-holds-first-meeting-2025-02-03-1.736958 (2025).

57. Matt Coleman, Kym Livesley, Michelle Segaert, Stephanie Hendy & Jett Lobb. Powering Australia's data centre boom: Navigating energy compliance

and opportunity; Dentons, Sydney and Melbourne, Australia, https://www.dentons.com/en/insights/articles/2025/july/3/powering-australias-data-centre-boom (2025).

58. Justin March, Emily Harris & Yvonne Chang. Data centers – A Case for New Zealand; DLA Piper, London, England, https://www.dlapiper.com/en/insights/publications/real-estate-gazette/real-estate-gazette-data-centers/new-zealand-data-centres-a-case-for-new-zealand (2022).

59. Russell Craig. Conscious cloud: our Aotearoa datacenter pledge Microsoft's datacenter region powered by 100% carbon free energy from day one; Microsoft, Redmond, Washington, https://news.microsoft.com/en-nz/2022/08/18/conscious-cloud-our-aotearoa-datacenter-pledge-microsofts-datacenter-region-powered-by-100-carbon-free-energy-from-day-one/ (2022).

60. Cushman & Wakefield. South Korea's Data Center Industry: Growth and Challenges Amid Power Supply and Regulatory Pressures; Chicago, Illinois, https://www.cushmanwakefield.com/en/south-korea/insights/south-korea-data-center-industry (2024).

61. PS Lee. Malaysia's Approach to Sustainable Data Centres: A Regional Benchmark; LinkedIn, Sunnyvale, California, https://www.linkedin.com/posts/leeps_sustainability-datacentres-malaysia-activity-7281503240951603202-NoWb/ (2025).

62. Dylan Butts. Eco-friendly data centers help drive $6.3 billion of green investment in Southeast Asia, but report shows more needed; CNBC (Consumer News and Business Channel), Englewood Cliffs, New Jersey, https://www.cnbc.com/2024/04/16/green-data-centers-help-increase-green-investment-in-southeast-asia-report.html (2024).

63. Green Data Center Indonesia (Greenex DC). Indonesian Government Policy Supports Green Data Center; Greenex DC, Jakarta, Indonesia, https://greenexdc.com/indonesian-government-policy-supports-green-data-center/ (2022).

64. Yuniar A. Indonesia partners private sector to invest in green data centres; GovInsider, One Raffles Place Mall, Singapore, https://govinsider.asia/intl-en/article/indonesia-partners-private-sector-to-invest-in-green-data-centres (2025).

65. Credence Research. Latin America Green Data Center Market By Component (Solution: Monitoring and Management System, Cooling System, Networking System, Power System, Others; Services: Installation & Deployment, Consulting, Support & Maintenance); By Data Center Type (Large Enterprises, Small & Medium Enterprises (SMEs)); By End User (Healthcare, Financial Services, Government, Telecom & IT, Others) – Growth, Share, Opportunities & Competitive Analysis, 2024 – 2032; San Jose, California, https://www.credenceresearch.com/report/latin-america-green-data-center-market (Accessed August 2025).

66. Simon Cudennec. Powering Africa's Digital Future: The Challenge of Energy for Data Center Development: Update; Bracewell LLP, Houston, Texas, https://www.bracewell.com/resources/powering-africas-digital-future-the-challenge-of-energy-for-data-center-development/ (2025).

67. International Telecommunication Union (ITU) and The World Bank. Green data centers: towards a sustainable digital transformation - A practitioner's guide; ITU, Geneva, Switzerland, https://www.itu.int/en/ITU-D/Environment/Pages/Toolbox/Green-data-center-guide.aspx (Accessed August 2025).

68. Basel Convention on the Control of Transboundary Movements of Hazardous Wastes and their Disposal. Parties to the Basel Convention on the Control of Transboundary Movements of Hazardous Wastes and their Disposal; Basal Convention, Basel, Switzerland, https://www.basel.int/Countries/StatusofRatifications/PartiesSignatories/tabid/4499/Default.aspx (Accessed August 2025).

69. European Union – EUR-Lex. Directive (EU) 2023/1791 of the European Parliament and of the Council of 13 September 2023 on energy efficiency and amending Regulation (EU) 2023/955 (recast) (Text with EEA relevance) (Document 32023L1791) at Chapter 3, Article 12; Publications Office of the European Union, Luxembourg, Luxembourg, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ%3AJOL_2023_231_R_0001&qid=1695186598766#art_12 (Accessed August 2025).

70. Catalyst. Sustainable data centers: ESG, compliance, and futureproofing for success; GRESB, Amsterdam, The Netherlands, https://www.gresb.com/nl-en/sustainable-data-centers-esg-compliance-and-futureproofing-for-success/ Sustainable data centers: ESG, compliance, and futureproofing for success (2025).

71. Bundesgesetzblatt. Bundesgesetzblatt Teil I: Gesetz zur Steigerung der Energieeffizienz und zur Änderung des Energiedienstleistungsgesetzes at Abschnitt 4, Sections 13 & 14; Food and Agriculture Organization of the United Nations (FAO), Rome, Italy, https://faolex.fao.org/docs/pdf/ger223337.pdf (2023).

72. Dr. Christoph Papenheim. Water usage and efficiency in German data centers: A regulatory overview; Dentons, Frankfurt, Germany, https://www.dentons.com/en/insights/articles/2023/november/22/water-usage-and-efficiency-in-german-data-centers-a-regulatory-overview (2023).

73. Ministry of Industry and Information Technology of the People's Republic of China (MIIT). Special Inspection Manual for Data-Center Energy Efficiency; MIIT, Beijing, China, https://www.gov.cn/zhengce/zhengceku/202404/P020240517496566744152.pdf (2024).

74. Caixin Global. Five Things to Know About China's Mega East-West Data Center Plan; Sixth Tone, Shanghai, China, https://www.sixthtone.com/news/1010207 (2022).

75. Ministry of Economy Trade and Industry (METI). Overview of the Energy Conservation Act, Energy-conservation information for business operators; METI), Tokyo, Japan, https://www.enecho.meti.go.jp/category/saving_and_new/saving/enterprise/overview/amendment/index.html (Accessed August 2024).

76. Sweep. California SB 253 – Key requirements and how to comply (Blog); Sweep SAS, Montpellier, France, https://www.sweep.net/blog/california-sb-253-key-requirements-and-how-to-comply (Accessed August 2025 (Last updated May 2025)).

77. Travis Talcott. ESG and Climate Accountability Reporting for U.S. Data Centers; EkkoSense, Nottingham, United Kingdom, https://www.ekkosense.com/resources/guides/esg-and-climate-accountability-reporting-for-u-s-data-centers/ (Accessed August 2025).

78. Robert Clark (Light Reading). Singapore Ends Data Center Pause As It Seeks Sustainable Growt; Data Center Knowledge, San Francisco, California, https://www.datacenterknowledge.com/cooling/singapore-ends-data-center-pause-as-it-seeks-sustainable-growth (2023).

79. Dan Swinhoe. EirGrid says no new applications for data centers in Dublin until 2028 - report; Data Center Dynamics, London, United Kingdom, https://www.datacenterdynamics.com/en/news/eirgrid-says-no-new-applications-for-data-centers-in-dublin-till-2028/ (2022).

80. Commission for Regulation of Utilities (CRU). CRU Direction to the System Operators related to Data Centre grid connection processing (Decision); Dublin, Ireland, https://cruie-live-96ca64acab2247eca8a850a7e54b-5b34f62.divio-media.com/documents/CRU21124-CRU-Direction-to-the-System-Operators-related-to-Data-Centre-grid-connection-.pdf (2021).

81. Matthew Gooding. The ongoing impact of Amsterdam's data center moratorium; Data Center Dynamics, https://www.datacenterdynamics.com/en/analysis/the-ongoing-impact-of-amsterdams-data-center-moratorium/ (2024).

82. David I. Schneider, Anthony W. DeLorenzo & Briana Becker Stolley. Loudoun County, Virginia, Eliminates By-Right Data Center Development; Holland & Knight LLP, Tampa, Florida, https://www.hklaw.com/en/insights/publications/2025/04/loudoun-county-virginia-eliminates-by-right-data-center-development (2025).

83. Zoning Administration Division. Data Centers – Adopted Zoning Ordinance Amendment; Government of Fairfax County, Fairfax, Virginia, https://www.fairfaxcounty.gov/planning-development/data-centers (Accessed August 2025).

84. Dan Swinhoe. Ireland's energy regulator proposes policy requiring data centers to match load with new power generation; Data Center Dynamics, London, United Kingdom, https://www.

datacenterdynamics.com/en/news/irelands-energy-regulator-proposes-policy-requiring-data-centers-to-match-load-with-new-power-generation/ (2025).

85. Climate Neutral Data Center Pact. Climate Neutral Data Center Pact: Initiative for Climate Neutral Data Centers in Europe; Brussels, Belgium, https://www.climateneutraldatacentre.net/wp-content/uploads/2025/06/CNDCP-Initiative-2025.pdf (Accessed August 2025).

86. Dr. Christoph Papenheim. How Germany's Energy Efficiency Act will impact data center operators; Dentons, Frankfurt, Germany, https://www.dentons.com/en/insights/articles/2023/september/25/energy-efficiency-act-relevance-for-data-centers (2023).

87. National Development and Reform Commission (NDRC). Opinions on Strict Energy Efficiency Constraints to Promote Energy Conservation and Carbon Reduction in Key Areas; NDRC, Beijing, China, https://www.ndrc.gov.cn/xxgk/zcfb/tz/202110/t20211021_1300583_ext.html (2021).

88. Infocomm Media Development Authority (IMDA). Driving a Greener Digital Future: Singapore's Green Data Centre Roadmap; IMDA, Mapletree Business City, Singapore https://www.imda.gov.sg/-/media/imda/files/how-we-can-help/green-dc-roadmap/green-dc-roadmap.pdf (2024).

89. APS Net Zero. Our path to NET ZERO; Australian Government Department of Finance, Canberra, Australia, https://www.sustainability.gov/pdfs/australia-nzgi-roadmap.pdf (Accessed August 2025).

90. Ministère de l'Aménagement du Territoire et de la Décentralisation and Ministère de la Écologique de la Biodiversité de la Forêt de la mer et de la Pêche. Fiscalité des énergies; Paris, France, https://www.ecologie.gouv.fr/politiques-publiques/fiscalite-energies (2020).

91. Environment Agency. Corporate report: Climate Change Agreements: Biennial Progress Report for 2021 and 2022 Gov.UK, London, United Kingdom, https://www.gov.uk/government/publications/climate-change-agreements-cca-biennial-report/climate-change-agreements-biennial-progress-report-for-2021-and-2022. (Accessed October 2025.)

92. Rich Kenny. Review of policy development on energy efficiency of data centres; Interact, Manchester, England, https://interactdc.com/posts/review-of-policy-development-on-energy-efficiency-of-data-centres/ (2024).

93. Covington & Burling LLP. The EU's Energy Efficiency Directive and Its Impact on Datacenters; Washington, DC, https://www.cov.com/-/media/files/corporate/publications/2025/04/the-eus-energy-efficiency-directive-and-its-impact-on-datacenters.pdf (2025).

94. Bundesgesetzblatt. Bundesgesetzblatt Teil I: Gesetz zur Steigerung der Energieeffizienz und zur Änderung des Energiedienstleistungsgesetzes at Abschnitt 4, Sections 11, (2); Food and Agriculture Organization of the United Nations (FAO), Rome, Italy, https://faolex.fao.org/docs/pdf/ger223337.pdf (2023).

95. Inc Energy Trust of Oregon. Data Center Cooling Equipment; Portland, Oregon, https://www.energytrust.org/incentives/commercial-existing-buildings-data-center/ (Accessed August 2025).

96. The General Office of the People's Government of Beijing Municipality. Notice from Beijing Municipal Development and Reform Commission on Issuing Several Provisions on Further Strengthening Energy-Saving Reviews for Data Center Projects; Beijing, China, https://www.beijing.gov.cn/zhengce/zhengcefagui/202307/t20230706_3156356.html (2023).

97. Internet Initiative Japan Inc. (IIJ). Carbon neutrality; IIJ, Tokyo, Japan, https://www.iij.ad.jp/en/datacenter/tech/cn/ (Accessed August 2025).

98. Silicon Valley Power. Energy Efficiency Rebates; City of Santa Clara, Santa Clara, California, https://www.siliconvalleypower.com/businesses/rebates (Accessed August 2025).

99. World Bank Group. 2025 State and Trends of Carbon Pricing; International Bank for Reconstruction and Development / The World Bank, Washington, DC, https://openknowledge.worldbank.org/bitstreams/7fda7708-5f14-4628-941b-a7a7f901b780/download (2025)

100. Net Zero Tracker. Homepage; https://zerotracker.net/ (Accessed August 2025).

101. Shihui Zhang, Wenjia Cai, Xinzhu Zheng, Xuedu Lv, Kangxin An, Yuan Cao, Hou Sang Cheng, Jingyi Dai, Xinyang Dong, Shuting Fan, Yiying Gao, Zaizuo Gong, Yusheng Guan et al. Global readiness for carbon neutrality: From targets to action. Environmental Science and Ecotechnology 25, 100546 (2025). https://doi.org/10.1016/j.ese.2025.100546.

102. Loudoun Climate Project. The Dark Side of Data Centers: Impacts of Loudoun's Data Center Industry; Loudoun County, Virginia, https://loudounclimate.org/data-centers/ (Accessed August 2025).

103. Charlie Paullin. Virginia environmental regulators make info on data center operations more publicly accessible; Virginia Mercury, Richmond, Virginia, https://virginiamercury.com/2024/11/13/virginia-environmental-regulators-make-info-on-data-center-operations-more-publicly-accessible/ (2024).

104. Joint Legislative Audit and Review Commission (JLARC). Data Centers in Virginia; JLARC, Richmond, Virginia, https://jlarc.virginia.gov/landing-2024-data-centers-in-virginia.asp (Accessed August 2025).

105. Sierra Club – Virginia Chapter. The Impacts of Unconstrained Data Center Growth; Richmond, Virginia, https://www.sierraclub.org/virginia/data-centers-harms (Accessed August 2025).

106. Woodstock Power Company (WPC). What Are Generator EPA Tier Ratings?; WPC, Philadelphia, Pennsylvania, https://woodstockpower.com/blog/overview-of-generator-epa-tier-ratings/ (Accessed August 2025).

107. US Environmental Protection Agency (EPA). Fact Sheet and Frequently Asked Questions: Use of Backup Generators to Maintain the Reliability of the Electric Grid; EPA, Washington, DC, https://www.epa.gov/system/files/documents/2025-05/rice-memo-on-duke-energy-regulatory-interpretation-04_17_25.pdf (2025).

108. Kirkland Alert. New EPA Guidance Clarifies When Data Centers and Other Operators May Utilize Emergency Backup Generators to Support Local Power Supply; Kirkland & Ellis LLP, Chicago, Illinois, https://www.kirkland.com/publications/kirkland-alert/2025/05/new-epa-guidance-clarifies-when-data-centers-and-other-operators-may-utilize-emergency-backup (2025).

109. Data Center Dynamics Press Release. Data center emissions and permitting challenges; Data Center Dynamics, London, United Kingdom, https://www.datacenterdynamics.com/en/product-news/data-center-emissions-and-permitting-challenges/ (2024).

110. Jill G. Cooper. Adoption of an Amendment to Chapters 112.1 and 112.2 (Zoning) of the 1976 Code of the County of Fairfax, Virginia; Government of Fairfax County, Fairfax, Virginia, https://www.fairfaxcounty.gov/planning-development/sites/planning-development/files/Assets/Documents/zoning%20ordinance/adopted%20amendments/ZO_112_1-2024-9.pdf (2024).

111. European Commission. CORDIS - EU research results: Fuel cell for prime power in data-centres; European Union, Luxembourg, Luxembourg, https://cordis.europa.eu/programme/id/H2020_FCH-02-9-2020 (Accessed August 2025).

112. Grant Gross. EU moves toward regulating data center energy and water use; CIO (IDG Communications Inc. dba Foundry), New York, New York, https://www.cio.com/article/2100517/eu-moves-toward-regulating-data-center-energy-and-water-use.html (2024).

113. Bundesgesetzblatt. Bundesgesetzblatt Teil I: Gesetz zur Steigerung der Energieeffizienz und zur Änderung des Energiedienstleistungsgesetzes; Food and Agriculture Organization of the United Nations (FAO), Rome, Italy, 2023).

114. Ministry of Finance of the People's Republic of China, Ministry of Ecology and Environment of the People's Republic of China & Ministry of Industry and Information Technology of the People's Republic of China (MIIT). Green Data Center Government Procurement Requirements Standard; Beijing, China, http://www.ccgp.gov.cn/zcfg/mof/202304/t20230410_19680193.htm (2023).

115. Liu Yake. China sets requirements for government procurement of data center equipment, services; Enviliance ASIA (by EnviX), Tokyo, Japan, https://enviliance.com/regions/east-asia/cn/report_10060 (2023).

116. Brett Walton. Data centers a small, but growing factor in Arizona's water budget; The Water Desk, Boulder, Colorado, https://waterdesk.org/2025/04/data-centers-a-small-but-growing-factor-in-arizonas-water-budget/ (2025).

117. Vanessa Kennedy. A Modern Economy: Transforming Data Center Cooling with Sustainable Water Solutions (Blog); Stack Infrastructure, Denver, Colorado, https://www.stackinfra.com/resources/thought-leadership/a-modern-economy-transforming-data-center-cooling-with-sustainable-water-solutions/ (2024).

118. PUB: Singapore's National Water Agency. NEWater: Overview; Environment Building, Singapore, https://www.pub.gov.sg/Public/WaterLoop/OurWaterStory/NEWater (2024).

119. Milrose Consultants. The Future of Data Center Development: Regulatory Trends Shaping the Industry; New York, New York, https://www.milrose.com/insights/the-future-of-data-center-development (2025).

120. Enviropass. RoHS Compliant Products Worldwide; Enviropass Expertise Inc., Quebec, Canada, https://getenviropass.com/rohs-compliant-products-worldwide/ (Accessed August 2025).

121. Pronexa. What is WEEE? The Waste Electrical and Electronic Equipment Directive; Munich, Germany, https://pronexa.com/blog/what-is-weee-the-waste-electrical-and-electronic-equipment-directive/ (2025).

122. National Environment Agency (NEA). New Nationwide E-Waste Management System Kicks In On 1 July 2021; Government of Singapore, Singapore, https://www.nea.gov.sg/media/news/news/index/new-nationwide-e-waste-management-system-kicks-in-on-1-july-2021 (2021).

123. Electronics Recycling Coordination Clearinghouse (ERCC). Map of States With Legislation; ERCC, Vienna, West Virginia, https://www.ecycleclearinghouse.org/contentpage.aspx?pageid=10 (Accessed August 2025).

124. National Center for Electronics Recycling (NCER). NCER Managed Compliance Calendar Website is One-Stop Location for Electronics Stakeholders; NCER, Vienna, West Virginia, https://www.electronicsrecycling.org/?page_id=37 (Accessed August 2025).

125. National Conference of State Legislatures (NCSL). Electronic Waste Recycling; NCSL, Denver, Colorado, https://www.ncsl.org/environment-and-natural-resources/electronic-waste-recycling (2018).

126. Bitong Li, Dongling Liu, Lina Zhang, Yue Wu, Xianlin Ding & Xiang Zeng. Challenges of E-Waste Dismantling in China. Toxics 12, 867 (2024). https://doi.org/10.3390/toxics12120867.

127. Data Recycling of New England. How China's Waste Import Ban Affects US Electronic Waste Recycling; Berkley, Massachusetts, https://datarecyclingne.com/blog/how-chinas-waste-import-ban-affects-us-electronic-waste-recycling/ (2018).

128. Sundae Ji. https://chemical.chemlinked.com/news/chemical-news/thailand-reinforces-e-waste-import-ban-covering-electrical-or-electronic-components-containing-pcbs (2025).

129. One Union Solutions. Why Some Countries Ban the Import of Used IT Equipment: A Deep Dive into E-Waste Policies; Haarlem, Netherlands, https://oneunionsolutions.com/blog/why-some-countries-ban-the-import-of-used-it-equipment/ (2025).

130. Vivian Lee, Pattabi Seshadri, Clark O'Niell, Archit Choudhary, Braden Holstege & Stefan A. Deutscher. Breaking Barriers to Data Center Growth; Boston Consulting Group (BCG), Boston, Massachusetts, https://www.bcg.com/publications/2025/breaking-barriers-data-center-growth (2025).

131. Bhargs Srivathsan, Marc Sorel, Pankaj Sachdeva, Arjita Bhan, Haripreet Batra, Raman Sharma, Rishi Gupta & Surbhi Choudhary. AI power: Expanding data center capacity to meet growing demand; McKinsey & Company, San Francisco, California, https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/ai-power-expanding-data-center-capacity-to-meet-growing-demand (2024).

132. [reserved]. State Data Center Energy Bills; Center on Global Energy Policy, Columbia University, New York, New York, (Forthcoming September 2025)).

133. Fiona Brocklehurst. International Review of Energy Efficiency in Data Centres for the Australian Department of Industry, Science, Energy and Resources at pg. 30; Australian Department of Industry, Science, Energy and Resources, Canberra, Australia, https://www.dcceew.gov.au/sites/default/files/documents/international-review-energy-efficiency-data-centres.pdf (2021).

134. EDNA Annex (Electronic Devices and Networks Annex) of the Technology Collaboration Programme on Energy Efficient End-Use Equipment (4E TCP). Policy development on energy efficiency of data centres at pg. 4; International Energy Agency (IEA), Paris, France, https://www.iea-4e.org/wp-content/uploads/2024/02/Policy-development-on-energy-efficiency-of-data-centres-draft-final-report-v1.05.pdf (2024).

135. Google Sustainability. Reports and case studies charting our progress and ambitions; Mountain View, California, https://sustainability.google/reports/ (Accessed August 2025).

136. Sustainability at Meta. Meta 2024 Sustainability Report; Menlo Park, California, https://sustainability.atmeta.com/wp-content/uploads/2024/08/Meta-2024-Sustainability-Report.pdf (2024).

137. Microsoft. Environmental Sustainability Report 2025; Redmond, Washington, https://www.microsoft.com/en-us/corporate-responsibility/sustainability/report/ (2025).

138. Kevin Clark. NextEra Energy boss: Embrace “energy realism”; Factor This Power Engineering, Shelton, Connecticut, https://www.power-eng.com/business/nextera-energy-boss-embrace-energy-realism/NextEra Energy boss: Embrace “energy realism” (2025).

139. Morgan Meaker. What Happens When Facebook Heats Your Home; WIRED, Boone, Iowa, https://www.wired.com/story/facebook-energy-heating-homes/ (2023).

140. Sustainability Directory. Data Center Heat Urban District Integration; Shanghai, China, https://prism.sustainability-directory.com/scenario/data-center-heat-urban-district-integration/ (2025).

141. The Green Grid. Welcome to The Green Grid (Homepage); San Ramon, California, https://www.thegreengrid.org/ (Accessed October 2025).

142. Peter Judge. Climate Neutral Data Centre Pact publishes list of European service providers signed up to net zero by 2030; Data Center Dynamics, London, United Kingdom, https://www.datacenterdynamics.com/en/news/climate-neutral-data-centre-pact-publishes-list-of-european-service-providers-signed-up-to-net-zero-by-2030/ (2023).

143. Google Data Centers. Innovating sustainable ideas. Growing renewable solutions. ; Google, Mountain View, California, https://datacenters.google/operating-sustainably/ (Accessed August 2025).

144. Microsoft. Our Microsoft Sustainability Journey; Redmond, Washington, https://www.microsoft.com/en-us/corporate-responsibility/sustainability-journey (Accessed August 2025).

145. Jeremy Rodriguez. Dial It In: Data Centers Need New Metric for Energy Efficiency; NVIDIA, Santa Clara, California, https://blogs.nvidia.com/blog/datacenter-efficiency-metrics-isc/ (2024).

146. Alexandra Sasha Luccioni, Emma Strubell & Kate Crawford. From Efficiency Gains to Rebound Effects: The Problem of Jevons' Paradox in AI's Polarized Environmental Debate. arXiv:2501.16548 (2025). https://ui.adsabs.harvard.edu/abs/2025arXiv250116548S.

147. Yuan Ma & Ziran Zhang. ICT and energy rebound effect: Empirical analysis based on data from Chinese cities. Journal of Environmental Management 370, 122651 (2024). https://doi.org/10.1016/j.jenvman.2024.122651.

148. D. Patterson, J. Gonzalez, U. Hölzle, Q. Le, C. Liang, L. M. Munguia, D. Rothchild, D. R. So, M. Texier & J. Dean. The Carbon Footprint of Machine Learning Training Will Plateau, Then Shrink. Computer 55, 18-28 (2022). https://doi.org/10.1109/MC.2022.3148714.

149. Julia Borgini. Navigate Energy Star data center standard and certification; Informa TechTarget, Newton, Massachusetts, https://www.techtarget.com/searchdatacenter/tip/Navigate-Energy-Star-data-center-standard-and-certification (2023).

150. Japan National Body/Green IT Promotion Council. DPPE: Holistic Framework for Data Center Energy Efficiency - KPIs for Infrastructure, IT Equipment, Operation (and Renewable Energy); Tokyo, Japan, https://home.jeita.or.jp/greenit-pc/topics/release/pdf/dppe_e_20120824.pdf (2012).

151. Infrarati. And yet another Data Center Performance Energy Metric; https://infrarati.wordpress.com/2010/04/20/and-yet-another-data-center-performance-energy-metric/ (2010).

152. Zack Royster. Zack Royster's Post: The Greengrid Datacenter AI Activity; LinkedIn, Sunnyvale, California, https://www.linkedin.com/posts/zack-royster-a525b1150_thegreengrid-datacenter-ai-activity-7348440920779812865-L5rU/ (2025).

153. V. D. Reddy, B. Setz, G. S. V. R. K. Rao, G. R. Gangadharan & M. Aiello. Metrics for Sustainable Data Centers. IEEE Transactions on Sustainable Computing 2, 290-303 (2017). https://doi.org/10.1109/TSUSC.2017.2701883.

154. Virginia's Legislative Information System (LIS). Virginia Acts of Assembly — 2020 Session Chapter 1193; Virginia State Government, Richmond, Virginia, https://legacylis.virginia.gov/cgi-bin/legp604.exe?201+ful+CHAP1193+pdf (2020).

155. Ram Iyer. Google inks $3B deal to buy hydropower from Brookfield; Tech Crunch, San Francisco, California, https://techcrunch.com/2025/07/15/google-inks-3b-deal-to-buy-hydropower-from-brookfield/ (2025).

156. Green Finance in Tech, Datacenters.com, https://www.datacenters.com/news/green-finance-in-tech-what-singtel-s-loan-means-for-the-industry#:~:text=Understanding%20Green%20Finance%20in%20Infrastructure,demonstrate%20long%2Dterm%20climate%20resilience (7 April 2025).

157. Katrina Rymill. What Are Green Bonds? How Are They Used in the Data Center Industry?; Equinix, Redwood City, California, https://blog.equinix.com/blog/2024/10/03/what-are-green-bonds-how-are-they-used-in-the-data-center-industry/ (2024).

158. ESG Today. Areim Raises Over $970 Million for Sustainable Data Center Fund; Modiin, Israel, https://www.esgtoday.com/areim-raises-over-970-million-for-sustainable-data-center-fund/ (2025).

159. Tan Ai Leng. DayOne raises RM15b in green financing to develop Johor data centres; ShareInvestor Pte Ltd, Kuala Lumpur, Malaysia, https://www.shareinvestor.com/news/550680?source=stbt (2025).

160. A. Barton Hinkle. Good Intentions, Bad Outcomes: The Story of Government; Reason Foundation, Los Angeles, California, https://reason.com/2017/07/26/good-intentions-bad-outcomes-the-story-o/ (2017).

161. Russell King. Are well-intentioned social policies causing problems?; Medium, San Francisco, California, https://medium.com/@policyreformer/are-well-intentioned-social-policies-causing-problems-828f8591e422 (2023).

TEXT BOX

Industry Initiatives

Minjue Wu and Gareth Jones

Many large data center owners and operators have longstanding commitments to sustainability, including ambitious pledges to reduce greenhouse gas emissions and water use. Parallel to these corporate commitments, industry associations, such as The Green Grid, have advocated for and supported sustainable practices for many years.

Leading industry sustainability initiatives are summarized in Table 1.

A. Pledges and disclosures of major data center operators

Table 1 shows the pledges and disclosures of several leading hyperscalers and other major data center operators.

Water pledges are increasingly common, with most major hyperscalers targeting "water positive" status by 2030. Some report water usage effectiveness (WUE) at a fleetwide level, while others report at a facility level. Coverage is incomplete and inconsistent (three-fifths of data center operators surveyed in 2024 do not actively track water usage metrics), creating problems for estimating absolute water consumption, especially in water-stressed regions.[19,20]

E-waste receives less attention than greenhouse gas emissions or water use. Microsoft has a pledge to reuse or recycle 90% of servers and components by 2025 and to implement zero-waste data centers by 2030; Google pledges zero waste to landfill; and AWS notes sourcing hardware from reuse inventory where possible. As of 2025, none of the six major hyperscalers above have published publicly available up-to-date e-waste metrics at the facility-level.[1,2,9,16,18,21] Across the sector, fleetwide quantitative data on reuse, recycling and landfill diversion is missing for more than two-thirds of surveyed data center operators.[20]

The ambitious pledges from several hyperscalers with respect to renewable energy procurement, efficiency standards and water use have created ripple effects across the broader data center ecosystem. Most reporting is fleetwide rather than facility specific. Despite ambitious top-line pledges, progress on water use and e-waste reduction can be difficult to evaluate externally.

Table 1. Pledges and disclosures of leading data center owners and operators

■ Pledges (Successful) ■ Pledges (Ongoing) ■ Disclosure

	Greenhouse Gas	Energy	Water	E-Waste
Amazon/ Amazon Web Services (AWS)	■ Match all electricity w/100% renewable energy by 2025 (Achieved in 2023 and 2024)[1]	■ Disclosed fleetwide average PUE (1.15 in 2024)[1]	■ Water positive by 2030[1] ■ WUE: 0.15 L/kWh in 2024[1]	■ Source hardware from its own reuse inventory[1]
Microsoft	■ Carbon negative by 2030[2] ■ Eliminate historical emissions since 1975 by 2050[3]	■ Disclosed fleetwide average PUE (1.12 in 2024)[3]	■ Water positive by 2030 for direct operations[2] ■ Reduce data center WUE by 40% from its 2022 baseline by 2030[3] ■ WUE: 0.30 L/kWh in 2024[4]	■ Reuse or recycle 90% of servers and components by 2025[2] ■ Implement zero-waste data centers by 2030[2]
Google	■ 24/7 carbon-free energy by 2030[5] ■ Reduce absolute emissions (combining scope 1, 2, and 3) by 50% from its 2019 baseline[6]	■ Disclosed fleetwide average PUE (1.09 in 2024)[7]	■ Use air cooling in highly stressed watershed areas, which fail the responsible use threshold for water cooling[8] ■ Achieve 120% water replenishment by 2030[8] ■ Adopted Water Risk Framework in 2023[8]	■ Achieve zero waste to landfill from data center operations[8]
Meta	■ 75% reduction in operational greenhouse gases from 2017 baseline[9] ■ 42% reduction in Scope 1 and 2 emissions from 2021 baseline by 2031[9] ■ 100% renewable energy for all owned data centers by 2020[9,10] ■ Net zero emissions in Scopes 1-3 by 2030[11]	■ Disclosed fleetwide average PUE (1.08 in 2023)[12] ■ Facility-level reporting of electricity use[13]	■ Water positive by 2030[12] ■ Discloses facility-level water data; WUE: 0.18 L/kWh in 2023[12]	■ Recycled 91% of data center construction waste in 2023[12]
Equinix	■ 100% renewable energy by 2030[14] ■ Captures Scope 1-3 emissions[15]	■ Disclosed fleetwide average PUE (1.39 in 2024)[16] ■ Targets an average PUE of 1.33 by 2030[14]	■ WUE: 0.95 L/kWh in 2024[16] ■ Began disclosing fleetwide WUE in 2023[15]	■ Certify all centers to ISO 14001 e-waste standards by 2027[17] ■ Redeployed 86% of excess server hardware in 2024[17]
Digital Realty	■ Carbon neutrality in the European Union by 2030[18] ■ Reduce Scope 1 and 2 emissions by 68% from 2018 baseline per square foot[18] ■ Reduce Scope 3 emissions by 24% from 2018 baseline by 2030[18]	■ Improve PUE by 5% in North America and 3% in Europe annually (■ in 2023)[18]	■ Improve WUE by 5% in North America from 2023[18]	■ Disclosed 80% waste diversion in data center construction in 2024[18]

B. Industry associations and other initiatives

Industry consortia and international bodies have developed standards and targets to align data centers with global climate goals and other environmental objectives. Some of these have since been integrated into legally binding regulations.

- The Green Grid: Formed in 2007 as the first industry body to advocate for and support sustainable data center practices, The Green Grid developed and introduced standards to support data centers to improve efficiency, including power usage effectiveness (PUE), WUE and carbon usage effectiveness (CUE).[22-25] In 2016, these standards and others developed by The Green Grid were adopted by the International Organization for Standardization (ISO) as part of ISO/IEC (International Electrotechnical Commission) 30134.[26]

- International Telecommunications Union (ITU) - Best Practices for Green Data centers: The ITU, with its mandate from member states as the specialized telecommunications agency of the United Nations, has developed a set of non-binding recommendations for developing green data centers.[27]

- EU Code of Conduct (CoC) for Data center Energy Efficiency: Launched in 2008 by the European Commission's Joint Research center, the CoC addresses rising data center energy use by targeting information technology (IT) load (equipment power) and facilities load (supporting systems).[28] It helps operators assess performance and set time-bound improvement targets. The goal of the CoC was to support data centers to achieve voluntary EU-set targets on efficiency. Companies join as Participants (implementing the CoC) or Endorsers (promoting it).

- Climate Neutral Data Center Pact (CNDCP): The CNDCP is a European industry-led agreement aligning with the EU's Green Deal, representing more than 85% of the EU's data center capacity.[29] Originally set up as a way to support data centers to become more energy efficient by setting targets and standards, since 2025 the CNDCP has abolished its own standards and instead now focuses on supporting regulators to set effective targets to reach the goal of climate-neutral data centers by 2030.[30] Members are now expected to report against EU standards.

- iMasons Climate Accord (ICA): The ICA is a voluntary industry coalition that develops standards and guidelines to support the industry in reporting carbon in data center power, materials and equipment.[31] The ICA has also developed a maturity model to support participating companies in improving sustainability. In particular, the ICA has provided a standardized carbon accounting framework, including Scope 3 and embodied emissions in data center materials and IT equipment, through spend-based embodied carbon accounting.

- LEED (US)/BREEAM (UK) certifications: LEED (US) and BREEAM (UK) are private organizations that have developed green building standards applied to data centers worldwide.[32] In a 2019 survey, data center professionals had much lower awareness of LEED and BREEAM than they did of the EU CoC.[33]

- Institute of Electrical and Electronics Engineers (IEEE) P7100 Environmental Impacts of Artificial Intelligence (AI) Working Group: According to the IEEE website, the P7100 standard "defines a measurement framework for reporting environmental indicators for training models and deriving inference on AI systems. This includes harmonized measurements of compute intensity (e.g., energy use) with associated environmental impacts (e.g., carbon dioxide (CO_2) emissions or water consumption).[34]

Recommendations

1. *Industry consortia and standards bodies should* ***provide technical support to small and medium operators*** *to assist with adopting and implementing sustainability pledges.*

2. *Data center owners and operators should* ***adopt third-party auditing of sustainability pledges as standard protocol*** *to enhance stakeholder confidence in reporting on sustainability pledges.*

3. *Financial institutions should* ***tie financial support to sustainability performance and disclosures****.*

References

1. Amazon. 2024 Amazon Sustainability Report; Seattle, Washington, https://sustainability.aboutamazon.com/2024-amazon-sustainability-report.pdf (2025).

2. Microsoft. 2025 Environmental Sustainability Report; Redmond, Washington, https://cdn-dynmedia-1.microsoft.com/is/content/microsoftcorp/microsoft/msc/documents/presentations/CSR/2025-Microsoft-Environmental-Sustainability-Report.pdf (2025).

3. Microsoft. 2024 Environmental Sustainability Report; Redmond, Washington, https://cdn-dynmedia-1.microsoft.com/is/content/microsoftcorp/microsoft/msc/documents/presentations/CSR/Microsoft-2024-Environmental-Sustainability-Report.pdf (2024).

4. Microsoft. Measuring datacenter energy and water use to improve Microsoft Cloud sustainability; Redmond, Washington, https://datacenters.microsoft.com/sustainability/efficiency/ (Accessed August 2025).

5. Google. Carbon-free energy moonshot in Environmental Report 2025 (Mountain View, California, 2025, https://www.gstatic.com/gumdrop/sustainability/google-2025-environmental-report.pdf#page=79).

6. Google Data Centers. Innovating sustainable ideas. Growing renewable solutions.; Google, Mountain View, California, https://datacenters.google/operating-sustainably/ (Accessed August 2025).

7. Google. Energy efficiency in Environmental Report 2025 (Mountain View, California, 2025, https://www.gstatic.com/gumdrop/sustainability/google-2025-environmental-report.pdf#page=21).

8. Google. Environmental Report 2024; Mountain View, California, https://www.gstatic.com/gumdrop/sustainability/google-2024-environmental-report.pdf (2024).

9. Meta. 2025 Sustainability Report; Menlo Park, California, https://sustainability.atmeta.com/wp-content/uploads/2025/08/Meta_2025-Sustainability-Report.pdf (2025).

10. Edward Palmieri. The next decade: How Facebook is stepping up the fight against climate change; Engineering at Meta, Menlo Park, California, https://engineering.fb.com/2020/09/14/data-center-engineering/net-zero-carbon/ (2020).

11. Urvi Parekh. Achieving our goal: 100% renewable energy for our global operations; Tech at Meta, Menlo Park, California, https://tech.facebook.com/engineering/2021/4/renewable-energy/ (2021).

12. Meta. 2024 Sustainability Report; Menlo Park, California, https://sustainability.atmeta.com/wp-content/uploads/2024/08/Meta-2024-Sustainability-Report.pdf (2024).

13. Meta. 2021 Sustainability Report; Menlo Park, California, https://sustainability.fb.com/wp-content/uploads/2022/06/Meta-2021-Sustainability-Report.pdf (2022).

14. Equinix. Environment: Efficiency and Clean Power; Redwood City, California, https://sustainability.equinix.com/environment/efficiency-and-clean-power/ (Accessed August 2025).

15. Christopher Wellise. How We Worked to Build a Better Future in 2023 (Blog); Equinix, Redwood City, California, https://blog.equinix.com/blog/2024/04/29/how-we-worked-to-build-a-better-future-in-2023/ (2024).

16. Equinix. Sustainability Report FY2024; Redwood City, California, https://www.equinix.com/content/dam/eqxcorp/en_us/documents/resources/infopapers/ip_2024_sustainability_en.pdf (2025).

17. Equinix. Environment: Circularity; Redwood City, California, https://sustainability.equinix.com/environment/circularity/ (Accessed August 2025).

18. Digital Realty. 2024 Impact Report; Digital Realty Trust, Inc., Austin, Texas, https://go2.digitalrealty.com/rs/087-YZJ-646/images/Report_Digital_Realty_2024_Impact_Report.pdf (2025).

19. Impax Asset Management. Addressing the hidden water footprint of data; Banque Nationale de Paris (BNP) Paribas Asset Management Europe, Paris, France, https://viewpoint.bnpparibas-am.com/addressing-the-hidden-water-footprint-of-data/ (2025).

20. Douglas Donnellan, Andy Lawrence, Daniel Bizo, Peter Judge, John O'Brien, Jacqueline Davis, Max Smolaks, Jabari Williams-George & Rose Weinschenk. Uptime Institute Global Data Center Survey Results 2024; Uptime Intelligence (Uptime Institute), New York, New York, https://datacenter.uptimeinstitute.com/rs/711-RIA-145/images/2024.GlobalDataCenterSurvey.Report.pdf?version=0 (2024).

21. Google. Waste reduction in Environmental Report 2025 (Mountain View, California, 2025, https://www.gstatic.com/gumdrop/sustainability/google-2025-environmental-report.pdf#page=44.

22. The Green Grid. About Us; Washington, DC, https://www.thegreengrid.org/about-us (Accessed August 2025).

23. The Green Grid. PUE: A Comprehensive Examination of the Metric; Washington, DC, https://www.thegreengrid.org/en/resources/library-and-tools/20-PUE%3A-A-Comprehensive-Examination-of-the-Metric (2014).

24. Michael Patterson, Dan Azevedo, Christian Belady & Jack Pouchet. WP#35 - Water Usage Effectiveness (WUE): A Green Grid Data Center Sustainability Metric; The Green Grid, Washington, DC, https://www.thegreengrid.org/en/resources/library-and-tools/238-WP%2335—-Water-Usage-Effectiveness-%28WUE%29%3A-A-Green-Grid-Data-Center-Sustainability-Metric (2011).

25. Christian Belady, Dan Azevedo, Michael Patterson, Jack Pouchet & Roger Tipley. WP#32 - Carbon Usage Effectiveness (CUE): A Green Grid Data Center Sustainability Metric; The Green Grid, Washington, DC, https://www.thegreengrid.org/en/resources/library-and-tools/241-WP%2332—-Carbon-Usage-Effectiveness-%28CUE%29%3A-A-Green-Grid-Data-Center-Sustainability-Metric (2010).

26. International Organization for Standardization (ISO). ISO/IEC 30134-1:2016; ISO, Geneva, Switzerland, https://www.iso.org/standard/63450.html (2016).

27. International Telecommunications Union (ITU). L.1300 : Best practices for green data centres; ITU, Geneva, Switzerland, https://www.itu.int/rec/T-REC-L.1300-201406-I/en (2014).

28. The Joint Research Centre: EU Science Hub. The EU Code of Conduct for Data Centres - towards more innovative, sustainable and secure data centre facilities; Brussels, Belgium, https://joint-research-centre.ec.europa.eu/jrc-news-and-updates/eu-code-conduct-data-centres-towards-more-innovative-sustainable-and-secure-data-centre-facilities-2023-09-05_en (2023).

29. Ben Maynard. Hasty Imposition of Minimum Performance Standards May Undermine Environmental Progress; Climate Neutral Data Center Pact (CNDCP), Brussels, Belgium, https://www.climateneutraldatacentre.net/2025/07/29/hasty-imposition-of-minimum-performance-standards-may-undermine-environmental-progress/ (2025).

30. Climate Neutral Data Center Pact (CNDCP). New Era For Climate Neutral Data Center Pact; CNDCP, Brussels, Belgium, https://www.climateneutraldatacentre.net/2025/06/11/new-era-for-climate-neutral-data-center-pact/ (2025).

31. iMasons Climate Accord. About Us; Los Angeles, California, https://climateaccord.org/about/ (Accessed August 2025).

32. Novva Data Centers Content Team. Building Standards for a Green Data Center; Novva Data Centers, West Jordan, Utah, https://www.novva.com/media-center/green-building-standards-data-centers/ (2025).

33. BREEAM. Data Centres: The Sustainability Gap (Survey Report); BRE, Watford, United Kingdom, https://files.bregroup.com/breeam/surveyreport/breeam-survey-report-data-centres-sustainability-gap.pdf (2019).

34. Institute of Electrical and Electronics Engineers (IEEE). IEEE P7100 Environmental Impacts of Artificial Intelligence Working Group; IEEE, New York, New York, https://sagroups.ieee.org/eiai/ (Accessed August 2025).

TEXT BOX

Local Opposition

David Sandalow, Minjue Wu and Angela Yuan

Strong local opposition to new data centers is emerging in many places. Such opposition is not new. Community organizations in Ireland, the Netherlands, Chile and other locations have been mobilizing to halt new data center construction for years.[1-3] However, recent surges in the construction of hyperscale data centers have sparked a significant increase in local pushback in the United States and around the world.[4,5]

In the United States, opposition to data centers has focused on grid strain, electricity costs, land use, residential quality of life and environmental impacts.[6,7] Data Center Watch, an activist group, estimates that over $64 billion in data center projects were blocked or delayed in the United States between 2023 and 2025 due to local pushback.[8] Examples of recent local opposition include the following:

- In Arizona, the Tucson City Council rejected Amazon's Project Blue AI data center campus due to concerns about substantial water consumption.[9]
- A new North Carolina data center proposal was withdrawn amidst local protests against rising electricity rates and potential grid strain.[10]
- In Indiana, at least five data center proposals were rejected in 2024 and 2025.[11]
- Virginia residents blocked a new data center in Chesapeake, citing concerns about noise, environmental degradation and the project's proximity to residential areas.[12]

Communities in other countries are objecting to data center development as well:

- In Santiago, Chile, local protests successfully halted authorization of a $200 million Google data center project, which has been put on hold until the plan is revised to comply with stringent sustainability standards.[13]
- In the United Kingdom, local opposition has delayed several data center projects, with communities raising concerns about how data center construction on greenbelt-designated land would impact the environment and strain infrastructure.[14]

- Local governments in Dublin, Amsterdam and Singapore have imposed single and multi-year moratoriums to reduce grid congestion and local concerns.[15-17]

This local opposition is sometimes at odds with policies of national, state and provincial governments, which often adopt policies to encourage data center development for a range of economic and security reasons.[18-19]

Recommendations

1. *Data center owners and operators should* ***engage collaboratively with local communities throughout the lifecycle of a project****, from site selection to post-construction operations. This engagement should* ***include communication of both the expected benefits to the community from the data center and potential risks*** *(including those related to grid strain, water resources and local air pollution).*

2. *Data center owners and operators should* ***work collaboratively with local communities*** *in areas near data centers to* ***implement measures that protect residential quality of life****.*

3. *Governments should* ***require data center owners and operators to provide certificates of collaboration with host communities*** *and, where appropriate,* ***enter into community benefit agreements****.*

References

1. DW. Protesters halt new data center (video); Microsoft Network (MSN), Redmond, Washington, https://www.msn.com/en-us/news/technology/protesters-halt-new-data-center/vi-AA1xOn7n?ocid=EMMX#details (2025).

2. Omroep Flevoland. Burgemeester en wethouders wisten niet van provinciale lobby datacenters; Lelystad, Netherlands, https://www.omroepflevoland.nl/nieuws/269133/burgemeester-en-wethouders-wisten-niet-van-provinciale-lobby-datacenters (2022).

3. Not Here Not Anywhere. Data Centres; Dublin, Ireland, https://notherenotanywhere.com/campaigns/data-centres/ (Accessed August 2025).

4. Carey L. Biron. As Trump promotes data centres, communities push back; Context, Thomson Reuters Foundation, London, United Kingdom, https://www.context.news/ai/as-trump-promotes-data-centres-communities-push-back (2025).

5. Julia Rone. Pushback against energy-guzzling data centres grows; Context, Thomson Reuters Foundation, London, United Kingdom, https://www.context.news/big-tech/opinion/pushback-against-energy-guzzling-data-centres-grows (2023).

6. Sean Patrick Cooper. Noisy, Hungry Data Centers Are Catching Communities by Surprise; The New York Times, New York, New York, https://www.nytimes.com/2024/09/15/opinion/data-centers-ai-amazon-google-microsoft.html (2024).

7. Christopher Tozzi. Why Communities Are Protesting Data Centers – And How the Industry Can Respond; Data Center Knowledge (Informa TechTarget), Newton, Massachusetts, https://www.datacenterknowledge.com/data-center-construction/why-communities-are-protesting-data-centers-and-how-the-industry-can-respond (2024).

8. Data Center Watch. $64 billion of data center projects have been blocked or delayed amid local opposition; United States, https://www.datacenterwatch.org/report (2025).

9. Sebastian Moss. Residents cheer as Tucson rejects Amazon's massive Project Blue data center campus in Arizona; Data Center Dynamics (DCD), London, United Kingdom, https://www.datacenterdynamics.com/en/news/residents-cheer-as-tucson-rejects-amazons-massive-project-blue-data-center-campus-in-arizona/ (2025).

10. Bracey Harris. Data center developer backs off $30 billion plans for Dale Earnhardt widow's property; NBC News, New York, New York, https://www.nbcnews.com/news/us-news/dale-earnhardt-widow-data-center-backs-off-mooresville-north-carolina-rcna224827 (2025).

11. Matthew Zeitlin. Indiana Faces a Data Center Backlash; Heatmap News, New York, New York, https://heatmap.news/energy/indiana-data-center-backlash (2025).

12. Ryan Murphy & Emily Feng. Why more residents are saying 'No' to AI data centers in their backyard; National Public Radio (NPR), Washington, DC, https://www.npr.org/2025/07/17/nx-s1-5469933/virginia-data-centers-residents-saying-no (2025).

13. Associated Press (AP). Google says it will rethink its plans for a big data center in Chile over water worries; New York, New York, https://apnews.com/article/chile-google-data-center-water-drought-environment-d1c6a7a8e8e6e45257ac84fb750b2162 (2024).

14. Kelvin Chan. As the data center industry booms, an English village becomes a battleground; Associated Press, New York, New York, https://apnews.com/article/britain-data-center-fdb196e2dec8bdf18eab6b8a6a672cbd (2024).

15. Jack Quann. ‘We’re facing down the barrel of blackouts’ – Protests calls for halt to new data centres; Newstalk, Dublin, Ireland, https://www.newstalk.com/news/were-facing-down-the-barrel-of-blackouts-protest-calls-for-halt-to-new-data-centres-1277717 (2021).

16. Matthew Gooding. The ongoing impact of Amsterdam's data center moratorium; Data Center Dynamics (DCD), London, United Kingdom, https://www.datacenterdynamics.com/en/analysis/the-ongoing-impact-of-amsterdams-data-center-moratorium/ (2024).

17. Janice Lim. Singapore unlikely to draw large-scale data centre investments despite added capacity: Report; The Straits Times, SPH News Centre, Singapore, https://www.straitstimes.com/business/singapore-unlikely-to-draw-large-scale-data-centre-investments-despite-added-capacity-report (2024).

18. The White House. Winning the Race: America’s AI Action Plan; Washington, DC, https://www.whitehouse.gov/wp-content/uploads/2025/07/Americas-AI-Action-Plan.pdf (2025).

19. European Commission. AI Continent – new cloud and AI development act – “Call for evidence for an impact assessment – Ares(2025)2878100”; Brussels, Belgium, https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/14628-AI-Continent-new-cloud-and-AI-development-act_en (Accessed August 2025).

APPENDIX A Recommendations From All Chapters

Chapter 1: Data Center Energy Use

1. *Governments and regulatory agencies should* ***develop public data repositories on the energy use and characteristics of data centers*** *at national and sub-national scales, including via mandatory data collection initiatives from data center operators with strict data quality and measurement and verification protocols.*

2. *Data center operators should* ***improve reporting and transparency on the energy use, peak power demand, operating characteristics and other environmental attributes (e.g., water consumption) of data centers*** *to improve the empirical knowledge base for data center energy analysts.*

3. *Governments, philanthropies and research institutions should* ***organize and convene forums to establish best practice analysis methods and data sharing initiatives*** *to rapidly improve the state of science for estimating and projecting data center energy use.*

4. *Researchers should* ***conduct regular inter-model comparisons of data center energy models and scenarios*** *to understand model differences, identify potential improvements, and establish and coalesce on best practices.*

5. *Governments, research institutions and international organizations should* ***convene forums specifically aimed at developing and disseminating scenario narratives for exploring future growth pathways for AI and data centers****, including associated frameworks and analysis approaches; similar to the Shared Socioeconomic Pathways,47 such scenarios would enable inclusion of data center futures in climate change mitigation scenarios.*

6. *Governments and research institutions should* ***develop and disseminate models and datasets of data center energy use in developing and***

***emerging economies**, which have historically been overlooked by the research community but whose data center electricity use may grow in the future.*

7. *Research institutions should **reinforce the need for best practices in analyses adopted by policymakers**, given wide variance in results associated with low-quality studies.*

8. *Governments and research institutions should **improve approaches to identifying existing and planned data centers**, such as through in-country sources and open shared location datasets such as the IEA AI Observatory. Improved understanding of the spatial patterns of data centers is important for assessing potential local impacts and proactively designing policies that avoid or minimize those impacts.*

9. *Governments and companies should **support research to better understand the CO_2 emissions of AI data center power sources**, focusing on more granular, grid-scale modeling of emissions. This includes closely tracking announced investments in cleaner power technologies, as well as tracking those investments when they ultimately come online, for more accurate forward-looking scenarios.*

Chapter 2: Data Center Energy Efficiency

2.1 It Equipment

1. *Companies, industry standard setters and engineering consortia should **align on common metrics for calculating and reporting the energy efficiency of IT equipment**. Data center operators should support such efforts since energy efficiency directly impacts their operating costs.*

2. *Governments and educational institutions should **develop and distribute resources to assist non-technical audiences in understanding and analyzing the energy requirements of IT equipment**.*

3. *Data center operators, utilities and government agencies should **consider the nature of AI computation workloads in designing, provisioning, operating and regulating data centers**. Differences between AI training and inference should be paramount during decision making.*

4. *Governments, utilities and industry consortia should **advance knowledge sharing platforms, case study data, diagnostic tools and training***

materials related to improving the energy efficiency of IT hardware *from procurement, operations and management perspectives.*

5. *Data center operators should* ***conduct, support and publish AI inference demand forecasts****. Trends between centralized data center computations and edge applications should be emphasized.*

6. *Data center operators should* ***redouble efforts to maximize the energy efficiencies of existing IT hardware****, including the adoption of efficient equipment, virtualization, zombie server identification initiatives, and refresh cycles optimized for energy efficiency.*

2.2 Software

1. *Educational institutions should* ***prepare the next generation of computer scientists and policymakers with the conceptual tools to understand and advance software efficiency****.*

    a. ***Create cross-disciplinary curricula*** *on algorithmic efficiency, carbon-aware computing and AI systems engineering.*
    b. ***Include energy literacy in computer science and AI degree programs****, covering both micro-level optimizations (e.g., quantization) and macro-level system design (e.g., flexible compute).*
    c. ***Develop policy bootcamps or executive courses for non-technical audiences****, such as civil servants, journalists and business leaders, on emerging AI compute trends and their implications for sustainability.*
    d. ***Fund open-access software efficiency toolkits and benchmarks****, especially those focusing on inference-time efficiency and emissions transparency.*

5. *Data center operators should* ***incorporate software-aware workload management and emissions-aware operations*** *into core planning.*

    a. ***Partner with utility companies and cloud platforms*** *to offer real-time grid carbon intensity data and renewable energy forecasts for intelligent job scheduling.*
    b. ***Offer time-delayed training products and pricing schemes*** *that incentivize shifts to lower-carbon AI training workflows.*
    c. ***Provide application programming interfaces (APIs)*** *that expose real-time energy and emissions data for AI workloads, enabling software developers to optimize code based on environmental impact.*

d. ***Adopt software-aware procurement criteria*** *that favor AI models and systems with verifiable efficiency gains or emissions-conscious design.*

5. *Software development companies should **build efficiency into the core of AI model design, training and deployment**.*

    a. ***Develop APIs and platforms that report model energy use**, offering customers transparency and options for greener usage.*
    b. ***Collaborate with academia to publish standardized benchmarks and best practices for evaluating software energy use**, not just performance or accuracy.*

3. *Regulatory bodies should **establish policies that ensure AI progress aligns with the public interest, energy constraints and climate goals**.*

    a. ***Incorporate software efficiency and emissions data disclosure requirements into AI governance frameworks**, especially for high-volume models or widely deployed systems.*
    b. ***Mandate transparent compute and energy reporting** for AI systems procured with public funding or deployed in sensitive sectors (e.g., health, education).*
    c. ***Develop incentives for energy-efficient AI systems**, such as research and development tax credits or public procurement preferences.*

4. *Educational institutions should **prepare the next generation of computer scientists and policymakers with the conceptual tools to understand and advance software efficiency**.*

    a. ***Create cross-disciplinary curricula** on algorithmic efficiency, carbon-aware computing and AI systems engineering.*

2.3 Cooling Technologies

1. ***Local governments should:***

    a. ***Work with utilities to expand access to non-potable or recycled water sources.***
    b. ***Offer data centers incentive structures**—such as tax abatements or expedited permitting—**tied to clear sustainability performance benchmarks**, such as targets for PUE, WUE and heat recovery ratios.*

2. ***Local governments in regions with naturally favorable climates for cooling should:***

    a. ***Actively promote this advantage to data center operators*** *and develop targeted incentives to attract new facilities, positioning their areas as energy— and cost-efficient locations for sustainable data center development.*
    b. *Accelerate deployment of advanced cooling by* ***streamlining permitting*** *for projects that integrate sustainable thermal management strategies, such as free cooling, heat reuse or closed-loop systems.*
    c. ***Update zoning regulations*** *to enable co-location of data centers with facilities that can use waste heat, such as greenhouses or municipal buildings.*

3. ***Local governments in regions with favorable conditions for thermal integration into district energy systems should work directly with data center operators*** *who understand regional opportunities on such projects.*

4. ***National policymakers should:***

    a. ***Establish the market conditions and regulatory frameworks*** *necessary for the broad adoption of energy— and water-efficient cooling technologies.*
    b. ***Create and enforce minimum energy performance standards for data centers****, along with voluntary or mandatory reporting requirements for PUE and WUE; expand programs, such as the US EPA's ENERGY STAR for Data Centers or the EU Code of Conduct for Data Centres, which can help set performance baselines and identify leaders.*
    c. ***Use government funding to support research and pilot programs*** *for promising but commercially immature technologies, such as two-phase immersion cooling, modular systems, thermal batteries and zero-water liquid cooling.*
    d. ***Provide tax credits, green bonds and procurement incentives*** *to help de-risk early adoption and support widespread deployment of sustainable cooling systems.*

5. ***Universities and research institutions should:***

    a. ***Prioritize studies of novel cooling techniques****, including many of the innovations described above, such as desiccant-based systems, thermal batteries, AI-optimized thermal control platforms and climate-specific hybrid systems.*

b. ***Host experimental testbeds*** *or collaborate with industry to evaluate the performance of emerging solutions in field conditions.*
c. ***Create open access datasets, simulation tools and digital twins*** *to allow broader communities to model, compare and benchmark advanced cooling approaches; standardizing such tools will improve planning accuracy and reduce design risk.*
d. ***Create or expand curricula on thermal systems, green data infrastructure and resilient design*** *to train the next generation of engineers and planners.*

6. ***Standards organizations*** *such as ASHRAE, International Standards Organization (ISO) and OCP should:*

   a. ***Facilitate innovation and interoperability*** *to evolve their guidance.*
   b. ***Establish uniform testing protocols and certification pathways*** *to validate performance of new technologies—especially liquid cooling, rear-door heat exchangers and high-efficiency refrigerants.*
   c. ***Push for global alignment on definitions and performance thresholds*** *to ower costs, reduce vendor lock-in and allow data center operators to deploy advanced cooling with greater confidence across international markets.*

7. ***Cooling equipment manufacturers****, bridging the gap between research and widespread implementation, should:*

   a. ***Invest in research and development (R&D)*** *focused on compact cold plates, advanced heat exchanger surfaces and system-integrated controls with predictive maintenance and AI optimization capabilities.*
   b. ***Offer comprehensive, modular solutions*** *that include sensors, telemetry and leak detection to reduce operational complexity.*
   c. ***Prioritize low-GWP and natural refrigerant alternatives****, consistent with the climate goals outlined in the Montreal Protocol Kigali Amendment, particularly ss regulations around refrigerants evolve.*

8. ***Data center developers and operators should:***

   a. ***Integrate design early in the project development process****, especially in siting decisions.*
   b. ***Select locations*** *that enable the use of free cooling, heat reuse or access to non-potable water and renewable energy.*
   c. ***Install climate-appropriate cooling systems****—such as evaporative cooling in dry regions or air-side economization in temperate zones—in tandem with IT deployment strategies.*

   d. ***Establish facility-level energy and water performance targets and publish sustainability metrics annually.***
   e. ***Evaluate and design for heat reuse opportunities****, either through district heating connections or local use cases like agricultural greenhouses, building heating or industrial preheating.*
   f. ***Set aside dedicated infrastructure for pilot deployments*** *of emerging cooling systems, allowing testing without disrupting core operations.*

9. ***Utilities should:***

   a. ***Partner with data centers to support load shifting*** *(to align significant workload periods with cooler times of day).*
   b. ***Integrate waste heat into community heating systems.***
   c. ***Offer incentive structures for grid-responsive cooling*** *in which data center cooling systems adjust their operation in response to signals from the electric grid.*

10. ***Environmental organizations should advocate for:***

    a. ***Low-carbon and water use***
    b. ***Transparent reporting***
    c. ***Waste heat reuse***
    d. ***Responsible siting*** *to align with climate and sustainability goals*

11. ***Investors and financiers should require disclosures*** *on WUE, PUE and refrigerant use in project finance deals because these metrics directly impact a data center's operational efficiency, climate risk exposure, regulatory compliance and long-term sustainability performance—all of which influence financial returns, reputational risk and alignment with Environmental, Social, and Governance (ESG) commitments.*

12. ***Insurance providers should reduce risk premiums for data center operators that adopt redundant and fault-tolerant cooling systems****—especially those with active leak detection and real-time monitoring—because these technologies significantly lower the likelihood of costly outages, equipment damage, and water or refrigerant leaks, thereby reducing the insurer's exposure to operational and environmental claims. For most companies, this is a shift from current practice, where premiums often do not fully account for the added risk mitigation these systems provide.*

13. ***End-use customers****, such as large cloud clients,* ***should shape demand by requiring data centers to meet high-efficiency and low-emissions***

***cooling benchmarks** in service agreements, such as maintaining a low PUE (ideally below 1.3) and minimizing greenhouse gas emissions by using low-GWP refrigerants, carbon-free electricity for cooling, and water-efficient or closed-loop systems.*

2.4 Heat Reuse

1. *Data center operators should **adopt high-temperature liquid-cooling systems**—such as direct-chip or immersion cooling—that achieve exit temperatures of 45-70 °C, enabling effective heat reuse in applications, such as district heating.*

2. *National and subnational governments **should require feasibility studies for heat reuse in permitting large new data-center projects and offer incentives, such as fast-track permitting and subsidies, to deploy such systems**. National and subnational governments should consider requiring 10-20% heat reuse mandates for new data centers (such as in Germany).*

3. *District heating utilities and municipal planners should **proactively partner with data-center developers to map potential synergies and create or extend thermal infrastructure** that connects data centers to buildings, industrial users and aquaculture facilities.*

4. *Heat host industries (e.g., hospitals, laundries, greenhouses and industrial processes) should **actively engage with data-center operators to explore using waste heat for 24/7 applications**, including agriculture, drying, aquaculture and wastewater treatment.*

5. *Technology developers and standards organizations should **produce guidelines, matchmaking tools and technoeconomic frameworks that facilitate collaboration between data-center operators and prospective heat hosts**, building on the work of the Open Compute Project (OCP) and others.*

6. *Research institutions, utilities and innovative companies should **pilot alternative uses and technologies—such as data-center-powered DAC systems**—evaluating performance and return on investment to increase reuse pathways.*

Chapter 3: Data Center Greenhouse Gas Emissions

3.1 On-site Greenhouse Gas Emissions (Scope 1)

1. *Data center operators should **examine alternatives to the continued use of diesel for on-site backup generation**. This could include alternative drop-in fuels with low-carbon intensity where available or the adoption of low-carbon backup generation, such as hydrogen fuel cells and the use of on-site energy storage.*

2. *Data center operators should **ensure that HVAC and fire suppression equipment leak detection protocols are modernized and carefully implemented to reduce F-gas leakage**. They should also closely follow regulatory developments around adopting advanced, low-GWP refrigerants and fire-suppression equipment.*

3. *Governments should **review current limitations on maximum operating limits for diesel backup generators** to ensure that air quality impacts and greenhouse gas emissions are minimized.*

4. *Utilities should **continue to meet high grid reliability performance targets**, reducing the need for on-site backup generation at data centers.*

3.2 Power Supply Greenhouse Gas Emissions (Scope 2)

1. *Data center operators should **maximize energy efficiency**, including using advanced cooling and other highly efficient equipment and implementing algorithmic efficiency whenever possible.*

2. *Data center operators should consider implementing **load flexibility and on-site storage**, particularly for grids with highly variable emissions intensity.*

3. *Data center operators should **include grid carbon intensity as a key siting consideration** and seek to site data centers in the lowest-emitting grid regions as much as possible.*

4. *Data center operators and utilities should work together to **identify the optimal mix of new low-carbon power generation technologies to add to the grid to meet rising data center load**. This should include consideration of data center load flexibility when determining the amount of new generation required.*

5. *Data center operators considering on-site/BTM power generation solutions should seek to* ***minimize emissions when selecting generation technologies****.*

6. *Data center operators should continue to* ***support emerging/developing low-carbon power generation technology****.*

7. *In addition to the above strategies, data center operators should continue to* ***procure renewable energy through PPAs****. They should also anticipate potential changes to the Greenhouse Gas Protocol Scope 2 guidance and plan accordingly when determining the necessary amount and type of procurement.*

8. *Grid operators should work closely with data centers to* ***understand the appropriate amount of new capacity to add to meet rising load and should seek to maximize low-carbon generation technologies for new capacity additions as much as possible****.*

9. *Grid operators should continue to* ***reform and accelerate the interconnection process for intermittent renewable generation*** *in order to provide new low-carbon capacity to meet data center and other demand.*

3.3 Embodied Greenhouse Gas Emissions (Scope 3)

1. *Governments should* ***assemble and share data related to direct, indirect and embodied greenhouse gas emissions from data center construction and operation****. Data center owners and operators should* ***volunteer to share site-specific estimated Scope 3 emissions data proactively and invite third-party review****. If necessary, governments should* ***require disclosure of this information****.*

2. *All stakeholders should* ***gain familiarity with the embodied emissions of data centers. They should recognize that abatement options today are real but limited and potentially expensive****.*

3. *Before designing and siting data centers, data center owners and operators should* ***identify and assess potential options to reduce Scope 3 emissions through material reduction and substitution****. Companies should use existing scientific criteria for high-quality, low— carbon goods and should consider developing their own criteria.*

4. *During procurement and construction phases, data center owners and operators should* ***assess the availability of low-carbon strategies and***

***materials**, including IT materials and building materials and use those low-carbon strategies and materials wherever possible. They should consider EACs to speed emissions reduction and support low— carbon manufacturing facilities, such as biocoke in blast furnaces, carbon-free steel production and cement with CCUS. They should also consider adhering to low embodied-carbon procurement standards for electronics developed by industry consortia.*

5. *Governments should support **comprehensive, transparent studies on optimizing overall embodied emissions reductions** across the full spectrum of data center IT equipment. These studies should be conducted by independent, third-party researchers, with relevant data shared voluntarily by data center operators.*

6. *During the operational phase, data center operators should **minimize IT equipment refresh rates** and seek to procure low embodied-emissions servers, networking equipment, memory and related equipment.*

7. *Governments should **assess the current supplies of low-carbon building materials and consider adding production capacity through policy measures**, including direct grants, government-backed procurement, contracts for differences, etc. They should also consider regulating production of IT hardware to reduce emissions, in particular focusing on F-gas use, leakage and destruction.*

8. *Governments should **support development of advanced technologies that limit the greenhouse gas footprint** associated with data center construction. They should explore and support applied research into alternative production approaches to chip-making that use less F-gases and manage their leakage better. They should explore alternative pathways to manufacturing cement, concrete and steel.*

Chapter 4: Accelerating Low-carbon Power with AI Data Centers

1. *Utilities and independent power producers (IPP) should:*

    a. ***Deploy advanced control tools** to accelerate interconnection and grid studies and to operate flexible portfolios. These tools include model-predictive control, enhanced forecasting and, where appropriate, AI.*

b. ***Adopt staged or ramped interconnections for large loads*** *(within standard planning cycles) and require telemetry and fast power—capping from data centers and consider on-site storage to provide demand response and regulation while maintaining service-level objectives.*
c. *Use these tools to* ***prioritize non-wires alternatives*** *and to reduce curtailment in renewable-rich zones.*

2. *Electricity regulators should:*

    a. ***Establish clear 24/7 carbon-free energy procurement pathways*** *that treat storage and clean-firm resources as first-class options alongside renewables.*
    b. *Enable* ***advanced market commitments (AMCs)*** *that allow multi—buyer participation, recognize hourly matching, and credit verifiable flexible-load performance.*

3. *National governments, regulators, and utilities should:*

    a. ***Expand targeted public-private risk-sharing*** *to lower the cost of firm, low-carbon supply while keeping rates affordable amid rising public concern about electricity bill impacts from data center-driven capacity additions.*
    b. ***Pair corporate offtake with loan guarantees****, liability and fuel frameworks (where relevant), and long-duration storage demonstrations.*
    c. ***Adapt CfD-style mechanisms to clean-firm resources and storage*** *so FOAK projects are followed by repeat builds of the same design.*

4. *Large data center operators with load flexibility, hyperscalers and procurement authorities should:*

    a. ***Commit to portfolio-based, 24/7 carbon-free procurement*** *that include renewables, storage and clean-firm resources where available.*
    b. ***Publish transparent hourly performance*** *and adopt grid-supportive operating modes, such as fast power caps and brief curtail on-signal, to* ***unlock faster interconnection and lower system costs****.*
    c. ***Prioritize deliverable power*** *to the public grid (when siting in resource-rich regions), rather than exclusively behind-the-fence supply.*

5. *National and local governments should:*

    a. ***Link siting incentives for new AI campuses to system and community value.***
    b. *Require additional deliverable **clean capacity, storage co-procurement and community benefit plans** that include workforce pipelines, water stewardship and shared transmission upgrades.*

6. *National governments and utilities, including public power and transmission owners, **should invest in grid modernization**, including advanced transmission, system visibility and congestion management, so resource-rich zones can host large and flexible loads without unnecessary overbuild. Regulators should authorize these investments, set incentives and ensure timely cost recovery.*

7. *Academic experts and system operators should **advance operations-ready forecasting for clean grids**. Priorities include post-processing and downscaling of weather models (including via AI) for wind and solar, probabilistic products that feed unit commitment and storage scheduling, and open benchmarks that **connect forecast improvements to avoided reserves, reduced curtailment, and emissions reductions**.*

Chapter 5: Data Center Water Use

1. *Governments should **assemble and share data** related to direct, indirect and embodied water consumption from data center construction and operation. Data center owners and operators should volunteer to **share site-specific water use and consumption data** proactively and invite third-party review. If necessary, governments should require disclosure of this information.*

2. *All stakeholders should **recognize that data center water use is tiny in relation to water use by other sectors globally but can be very significant in water-scarce regions**.*

3. ***Before siting data centers**, data center owners and operators should **assess likely water impacts**, including in particular by consulting with local stakeholders. In water-scarce regions, companies should consider several steps to reduce likely water impacts:*

    a. ***Apply advanced cooling approaches** to reduce direct water use, with potential additional expense.*

*b. **Assess the cost and viability of water reclamation** and reuse and of increasing water supply (e.g., through desalination).*
*c. **Maximize non-thermal power supplies**, including solar, wind and batteries, including potential overbuilding of variable renewable resources and hybrid load balancing using a mix of thermal and non-thermal generation.*
*d. **Develop procurement standards** for building materials and chips with low-water footprints. Where possible, procure low-water footprint materials, including the cost of a modest green premium.*

4. Governments should support the development of advanced technologies that limit the water footprint associated with AI use.

*a. Most importantly, governments should **support replacing fossil generation** with non-thermal generation and should encourage use of air cooling in existing facilities, both of which would dramatically reduce indirect water use.*
*b. Similarly, governments should **support novel cement, concrete and steel technology** that would reduce water consumption, as well as CO_2 emissions. Where possible, companies should accelerate adoption and procurement of low-water pathways.*
*c. Governments should **undertake a cost-benefit analysis** based on the lifetime of operation and seek support to reduce risk and cost.*

*5. Governments **should undertake initial and then systematic analysis to understand the technology options, costs and trade-offs between water-conserving options**. These analyses can serve as the basis for policy, including regulation or incentives. Data center builders and operators should share their data with government agencies to help identify low-cost, large-volume options for water footprint reduction.*

Text Box: Data Center Electronic Waste

1. Governments should:

*a. **Adopt and harmonize global standards for reuse, refurbishment and recycling of e-waste**, including for sanitization of data-bearing information technology (IT) equipment*
*b. **Adopt and strengthen extended producer responsibility rules.***

2. Data center operators should:

*a. **Refurbish, resell or donate retired equipment**; process that* *equipment through certified recyclers; and publicly report end-of-life outcomes*

b. ***Prioritize measures to reduce equipment turnover****, such as regular preventative maintenance*

c. ***Reduce equipment purchases where possible by using tools such as virtualization****, cloud computing and shared infrastructure*

3. *Manufacturers should:*

    a. ***Design modular equipment that is easily repaired and disassembled****, provide spare-part support, include clear recyclability labeling, and establish validated pathways for reuse, resale or refurbishment*

    b. ***Prioritize use of recycled or easily recyclable materials***

    c. ***Reduce or eliminate use of hazardous materials***

Chapter 6: Government Policy

1. *Governments should* ***collect and share data*** *on data centers' energy use and environmental impacts.*

2. *Governments should* ***build capacity to better understand fast-moving trends*** *with respect to data centers' energy use and environmental impacts.*

3. *Governments should* ***use a broad set of metrics when regulating data centers' energy use and environmental impacts****, including not just PUE.*

4. *Governments should* ***assist the rapid buildout of clean power capacity*** *to help meet growing data center power demand.*

5. *When governments procure data center services, they should* ***require vendors of data center services to disclose their energy use, water use and greenhouse gas emissions****.*

6. *IEA Member governments should* ***expand the IEA's Energy and AI Observatory****, devoting additional resources to monitoring and reporting on data centers' energy use and environmental impacts, as well as policy trends with respect to data centers around the world. The Clean Energy Ministerial should* ***expand CEM's work on data centers*** *under its power sector and artificial intelligence initiatives.*

Text Box: Industry Initiatives

1. *Industry consortia and standards bodies should* ***provide technical support to small and medium operators*** *to assist with adopting and implementing sustainability pledges.*

2. *Data center owners and operators should* ***adopt third-party auditing of sustainability pledges as standard protocol*** *to enhance stakeholder confidence in reporting on sustainability pledges.*

3. *Financial institutions should* ***tie financial support to sustainability performance and disclosures****.*

Text Box: Local Opposition

1. *Data center owners and operators should* ***engage collaboratively with local communities throughout the lifecycle of a project****, from site selection to post-construction operations. This engagement should* ***include communication of both the expected benefits to the community from the data center and potential risks*** *(including those related to grid strain, water resources and local air pollution).*

2. *Data center owners and operators should* ***work collaboratively with local communities*** *in areas near data centers to* ***implement measures that protect residential quality of life****.*

3. *Governments should* ***require data center owners and operators to provide certificates of collaboration with host communities*** *and, where appropriate,* ***enter into community benefit agreements****.*