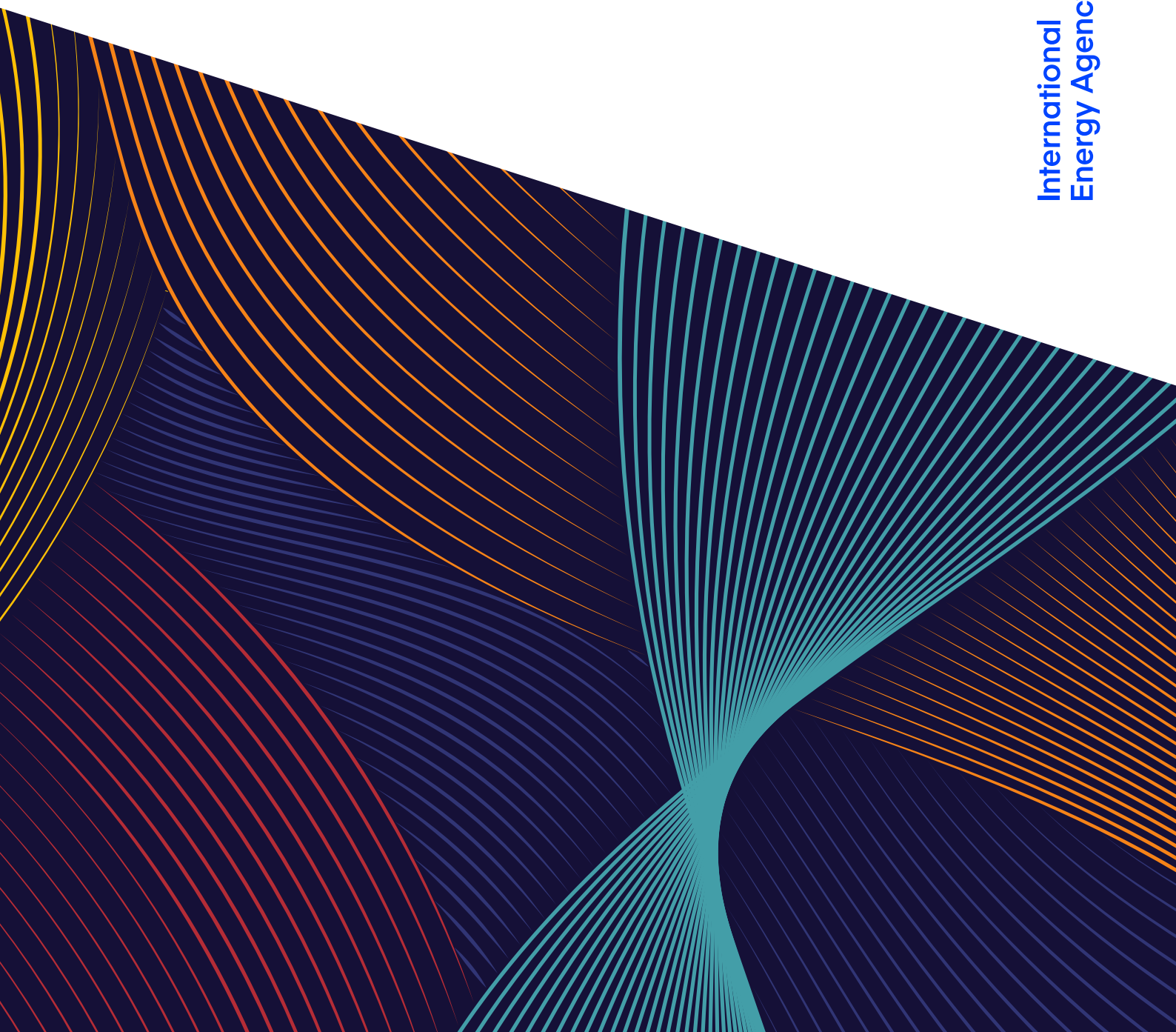
Powering Artificial Intelligence

Strategies for integrating data centres into electricity systems



International Energy Agency

INTERNATIONAL ENERGY AGENCY

The IEA examines the full spectrum of energy issues including oil, gas and coal supply and demand, renewable energy technologies, electricity markets, energy efficiency, access to energy, demand side management and much more. Through its work, the IEA advocates policies that will enhance the reliability, affordability and sustainability of energy in its 32 Member countries, 14 Association countries and beyond.

IEA Member countries:

Australia
Austria
Belgium
Canada
Czech Republic
Denmark
Estonia
Finland
France
Germany
Greece
Hungary
Ireland
Italy
Japan
Korea
Latvia
Lithuania
Luxembourg
Mexico
Netherlands
New Zealand
Norway
Poland
Portugal
Slovak Republic
Spain
Sweden
Switzerland
Republic of Türkiye
United Kingdom
United States

The European Commission also participates in the work of the IEA

IEA Accession countries:

Brazil
Chile
Colombia
Costa Rica
Israel
Romania

IEA Association countries:

Argentina
China
Egypt
India
Indonesia
Kenya
Morocco
Nigeria
Senegal
Singapore
South Africa
Thailand
Ukraine
Viet Nam

This publication, as well as any data and map included herein, are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

Source: IEA.
International Energy Agency
Website: www.iea.org

Abstract

As data centre investments surge globally, there are growing concerns over their implications for local electricity systems, electricity prices, emissions and water use. Prepared at the request of France's G7 Presidency, this report surveys the current state of policies, regulatory activity and corporate strategy in 34 policy jurisdictions covering 90% of global data centre capacity. It identifies key challenges, constraints and opportunities in the policy and regulatory domain pertaining to the integration of data centres in electricity systems. The report identifies sound principles for decision makers that can help ensure energy affordability, sustainability and energy security in the era of artificial intelligence.

IEA. CC BY 4.0.

Acknowledgements, contributors and credits

This study was prepared by the Energy Modelling Office, Energy Demand Outlook (EMO-EDO) Division and the Electricity Systems and Markets (ESM) Division of the International Energy Agency's (IEA) Directorate of Sustainability, Technology and Outlooks and the Directorate of Energy Markets and Security (EMS). The study was designed and directed by Thomas Spencer and Siddharth Singh, under the guidance of Laura Cozzi and Maria Sicilia.

The lead authors of the report were Leonardo Bolstad, Esra Bozkir Broekman, Eren Çam, Javier Jorquera Copier, Hans-Kristian Ringkjob. Other IEA colleagues provided important contributions, including Jay Barlow and Leonidas Kliafas.

Valuable comments were provided by IEA senior management including Dennis Hesseling and Brian Motherway.

Sylvia Beyer and Curtis Jenken provided valuable multilateral support.

The authors would also like to thank Adam Majoe for skilfully editing and copy-editing the manuscript and the IEA Communication and Digital Office, in particular, Jethro Mullen, Julia Horowitz, Astrid Dumond, Poeli Bojorquez and Clara Vallois.

Many experts from outside of the IEA reviewed the report and provided valuable comments. They include: Gilda Amorosi (Microsoft), Harmeet Bawa (Hitachi Energy), Hélène Costa de Beauregard (General Commissariat for Sustainable Development), Lisa Berry (GE Vernova), Pierre-Alexandre Bigel (Directorate-General for Energy and Climate, France), Jean-Philippe Bonnet (RTE), Christina Christopoulou (AWS), Edwin Haesen (ENTSO-E), Rafael Kawecki (Siemens Energy), Eoin Kennedy (EirGrid), Vincent Lesage (Directorate-General for Energy and Climate, France), Julia Matevosyan (Energy Systems Integration Group), Vincent Minier (Schneider Electric), Juliana Reed (DESNZ), Varun Sivaram (EmeraldAI), Stavros Stamatoukos (DG Energy, European Commission), Maud Texier (Octopus Energy), Simon Tweed (EirGrid) and Bruno Yango (Directorate-General for Energy and Climate, France).

This report was developed at the request of the French G7 Presidency, and the IEA would like to thank Pierre-Alexandre Bigel and Hélène Costa de Beauregard for their support. This work reflects the views of the IEA Secretariat but does not necessarily reflect those of the G7 or the IEA's individual Member countries or of any particular funder or collaborator.

IEA. CC BY 4.0.

Table of contents

Executive summary 6
Introduction 12
1. Understanding where and when data centres will consume electricity 17
Load projections 17
Spatial planning 20
2. Enhancing the energy efficiency and sustainability of data centres 25
Where efficiency is measured 25
Reporting and facility performance 28
Wider efficiency priorities 33
Long-term resilience of data centre infrastructure 34
3. Better management of grid connection requests by data centres 36
Permitting 37
Queue management and connection arrangements 41
4. Allocating costs related to grid and infrastructure upgrades for data centres 48
Network cost allocation 49
Policy trade-offs in cost allocation 53
Instruments in use across jurisdictions 54
Comparing tariff terms 55
5. Securing electricity supply for data centres 58
Strategies 58
Policy and regulation 61
6. Integrating data centres securely into the grid 63
Flexibility 63
Stability 66
Annex 71
Data centre and large-load policies by jurisdiction 71
Glossary 77

IEA. CC BY 4.0.

Executive summary

Surging investment in data centres presents opportunities as well as complex trade-offs

Data centre investment is surging as artificial intelligence has become central to economic policies across key markets. Globally, the largest data centre operators – so-called hyperscalers – are set to spend more than USD 700 billion as capital expenditure in 2026, mostly for the development of new data centres. Many jurisdictions want to attract digital infrastructure as the foundation of domestic innovation in artificial intelligence and digital services, while policy makers are making data centres and artificial intelligence key pillars of economic security and digital sovereignty.

Policy makers are increasingly sensitive to growing concerns, particularly among local communities, about data centre impacts on electricity prices, emissions and water. The unprecedented artificial intelligence-driven investment boom into digital infrastructure is straining local grids and stretching the supply chains for key electrical equipment. At the same time, it is driving rapid innovation in energy technologies, procurement strategies, and policy and regulatory frameworks. Many jurisdictions also face constraints in terms of available network capacity, long connection queues and slow permitting timelines.

Requested by France's G7 Presidency, this report takes stock of evolving practices in data centre integration into electricity systems. It surveys the current state of policy and regulatory activity and corporate strategy, looking across the data centre value chain in more than 34 policy jurisdictions covering about 90% of global data centre capacity. It identifies key challenges, constraints and opportunities in the policy and regulatory domain. This report builds on previous IEA analysis, notably the 2025 Energy and AI special report and 2026 Key Questions on Energy and AI report.

Policy makers and data centre operators are responding to a rapidly evolving landscape

Our analysis finds 134 policies and regulations guiding the integration of data centres into electricity systems globally. Nearly 40% of these policies were announced or implemented in the last 12 months, highlighting the dynamic nature of the field.

The most common area of policy activity focuses on improving the visibility of data centre demand in terms of scale, timing and location. This highlights

IEA. CC BY 4.0.

the key challenge that is proactively planning network capacity for data centres and connecting new large data centres to existing networks. Policies aiming to promote the smooth operational integration of data centres into electricity networks are rarer, although notable examples have emerged. Some of these mandate that data centres contribute to generation adequacy (so-called "bring-your-own-power," or BYOP), or to network stability and flexibility.

Data centre operators are also innovating in how they procure electricity and interact with the grid. Traditionally, these companies have been large procurers of renewable electricity; that has continued, with operators signing 45 terawatt-hours (TWh) of power purchase agreements (PPAs) in the first half of 2026, up by nearly 20% on the same period in 2025. Of this, nearly 70% was in the United States. Additional innovations are also emerging, with a small but growing number of power purchase agreements for emerging technologies such as advanced geothermal, small modular nuclear reactors, long-duration energy storage, virtual power plant arrangements, or battery storage. A large number of projects for onsite power have been announced, especially for natural gas in the United States.

Policy efforts focus on improving demand visibility

In several jurisdictions, network planners and operators are seeing connection queues for data centres with cumulative capacity that is larger than today's peak demand. Examples of jurisdictions with large pipelines include the US states of Texas and Virginia, Finland, and the United Kingdom. Connection queues contain duplicative or speculative projects, and policy makers must therefore prepare for potentially very large new loads without clear visibility of how much capacity will ultimately be built. Overestimating demand can risk creating stranded or underutilised grid investments, while underestimating it can lead to bottlenecks and delays.

This "queue inflation" is not unique to data centres, but several of their characteristics sharpen the challenge. Firstly, data centres can be very large projects, and they are often being developed in areas that were not previously major sources of electricity demand; this can substantially shift network topography that had been structured around historical load centres, such as industrial clusters and cities. Secondly, electricity demand for data centres is hard to assess. Even when projects proceed, demand typically ramps up gradually as buildings, servers and customer workloads come online, creating uncertainty over when and how much power consumption will materialise. More fundamentally, AI-focused data centres are huge capital investments for a novel, potentially transformational technology whose development, uptake, and profitability are uncertain.

IEA. CC BY 4.0.

Many jurisdictions are therefore undertaking queue reform with three broad objectives. The first is to ensure that connection queues reflect credible projects, using stronger queue entry requirements, maturity assessments, financial commitments and milestone-based capacity release mechanisms. A second objective is prioritising projects effectively within queues; several jurisdictions are moving from first-come-first served approaches to criteria that combine project maturity with broader considerations, such as strategic value, economic contribution and sustainability. However, ensuring that criteria remain transparent and predictable remains a key challenge. Several jurisdictions are also experimenting with methods to accelerate the connection studies needed to ensure projects can be safely connected to the grid.

Policy and corporate strategies transform some data centres from energy takers to energy makers

Traditionally, data centres have been passive actors in electricity networks; in some cases, that is now changing. In the face of growing pressures and bottlenecks, data centre operators, power project developers and policy makers are redefining how data centres interact with the grid. Importantly, not all data centres are the same, and their ability to interact proactively with the grid depends on factors such as their customer mix; the operator's size, financial position and operational sophistication; and local regulatory and policy conditions.

Policy makers and regulators are increasingly offering non-firm grid connections as a means of connecting data centres faster to electricity networks. Periods of system stress typically occur for relatively few hours per year, and non-firm grid connections could enable an additional 750-900 gigawatts (GW) of grid hosting capacity globally, according to IEA analysis. Jurisdictions ranging from Norway, South Australia, Spain, and the Southwest Power Pool in the United States are experimenting with different models of non-firm connection agreements. These include conditional non-firm or curtailable access; time-dependent, phased or ramped access; and dynamic consumption limits adjusted to real-time or forecast network conditions. A key condition for the greater use of non-firm connections is the need to provide standardised and transparent conditions so data centre operators can understand and mitigate potential risks to their operations. Japan's Watt-Bit Collaboration is an example of a public-private initiative that aims to support more flexibility from compute infrastructure over time.

Data centre operators are increasingly exploring diverse approaches to provide flexibility to the grid. Active management of computing loads remains quite rare and is constrained by the widespread unbundling of compute demand and compute supply. Many data centre operators provide compute services to third-party actors, and therefore have little visibility and control over workloads that take place in their data centres. Compute prices are also astronomical relative to

IEA. CC BY 4.0.

the prices of other flexibility sources: AI-optimised data centre capacity can cost as much as USD 20-30 billion per GW to rent for a year, which is 100 times more than the equivalent capacity cost of a battery. This makes curtailing compute operations very expensive. Nonetheless, the industry is exploring compute flexibility, especially if it enables faster grid connections. Frameworks to standardise the exchange of information across compute demand, compute supply and electricity networks will be critical to bridging a fragmented value chain and unlocking more flexibility. Recent stability-related events – including the abrupt disconnection of a large data centre load (over 3 GW) from the grid and switch to backup supply in the United States – have prompted several system operators to implement or consider technical performance requirements for data centres to support grid stability.

Bring-your-own-power (BYOP) is becoming a more common policy requirement and business strategy but takes different forms. Some jurisdictions, such as Ireland, have stipulated that certain kinds of new data centres need to contribute to system adequacy by building or contracting with new dispatchable capacity. A large number of onsite gas power projects have been proposed in the United States. While many remain at the proposal stage, some are moving forward with credible customers in the technology sector. However, onsite projects are technically complex and depend on installing generation and power equipment with long wait times. For these reasons, many onsite projects are integrating battery storage into their design and are still seeking an eventual grid connection; policy makers will therefore need to understand how they impact the broader grid if and when they connect. In other jurisdictions, including the People's Republic of China (hereafter "China") and Spain, policy makers have placed renewables purchase obligations on new data centres or are considering them. Amid these innovations, the key point for policy makers is that data centre procurement choices offer an opportunity for the sector to contribute to broader policy goals in terms of energy security, adequacy or sustainability.

Data centre sustainability remains in focus

Data centres are currently associated with around 150 million tonnes of CO_2 emissions due to their electricity consumption. Our central scenario sees this growing to around 350 million tonnes by 2035, with recent IEA projections increasing due to the higher use of onsite natural gas. This is set to be around 1.5% of global energy-related emissions in 2035. IEA analysis of data centres dedicated to AI reveals that nearly 40% of identified capacity is in areas with high or extremely high baseline water stress.

Policy makers are increasingly implementing energy and water efficiency standards for data centres, but the nature of the challenge is complex. Today, there are energy efficiency standards for equipment, but not AI models.

IEA. CC BY 4.0.

While there is significant potential to improve the efficiency of models, extremely rapid technical improvements, heterogenous AI services and uncertain baselines make standard-setting at the model level very difficult to implement. Moreover, while market pressures to optimise the use of compute – an expensive production input – are leading to rapid improvements in model efficiency, leading AI labs also face market pressures, and an ample supply of capital, to grow customer numbers and customer use. This can mean offering the best-performing models even if smaller, more efficient ones could satisfactorily complete many tasks. There is also a trade-off between water-use efficiency and energy efficiency in data centres, as cooling equipment that minimises water use tends to consume more energy, and vice versa. Additionally, the largest share of water use in the data centre value chain takes place upstream in power generation.

Data centres concentrate an enormous amount of economic value in a small space geographically, and they are exposed to extreme weather and other physical risks. Industry players are indicating that the high costs of data centres are leading to gaps in insurance coverage. Given the tight linkages between AI, data centres and financial markets, policy makers could consider the broader implications of risks and the extent to which individual data centres, or clusters of them, are exposed.

Policy makers have tools that can address affordability concerns

The effect of data centres on electricity bills is not clear-cut. Where a new large load requires additional network or generation investment, and that cost is recovered from all users, it can raise bills for other consumers. Where it makes fuller use of capacity that already exists and pays its share of the cost, a data centre can spread fixed costs over more consumption and ease pressure on bills. Either way, how the costs of connecting large loads are defined and allocated is a choice for policy makers.

Tariff frameworks allocate risks as well as costs. Network tariffs recover the largely fixed cost of building and maintaining the grid, and regulators design them to balance several principles, among them cost reflectivity (or the idea that each user's charges should reflect the true costs it places on the system). For large loads, this usually means charging for the capacity they reserve rather than the energy they consume. Networks are also planned and built on forecasts of loads that may not materialise, and the size and uncertainty of data centre loads make this risk particularly pronounced. Large-load tariffs can reduce other consumers' exposure to these risks through minimum-payment commitments, long-term contracts matched to the life of the assets, exit fees, collateral and phased connections, while financial commitments at the connection stage screen out projects that are unlikely to proceed. Together this addresses three distinct risks:

IEA. CC BY 4.0.

that a load uses less capacity than it reserved, that it leaves before the assets built for it are paid off, and that it never connects at all.

Effective policy packages involve balancing cost reflectivity and consumer protection against simplicity, investment certainty and competitiveness. Because risk mitigation instruments interact, assessing them together helps avoid both shifting unjustified costs onto existing consumers and charging data centres more than once for the same risk. This depends on transparent coordination among policy makers, regulators, utilities and system operators.

Policy makers should share experiences as they look to shape the data centre buildout

The data centre buildout is increasingly facing local opposition in some jurisdictions. In the face of this, policy makers have multiple tools, and company strategies are also adapting. Data centre grid connections are increasingly being made conditional on being "good grid citizens": contributing to flexibility, efficiency, fairness, stability and capacity adequacy. The industry itself is innovating to give back more to the grid, too, although it is important to recognise that data centres are a diverse asset class, with different impacts, incentives and capacities. As the sector develops rapidly, sharing information about best practices, tested approaches and new innovations will be critical. This report, and the dataset of policies and regulations that accompany it, aims to support these efforts.

6 principles for decision makers for the integration of data centres into electricity systems in the age of AI

1. Strive for clarity and transparency in the geographic and electricity consumption attributes of proposed data centres.

2. Establish transparent and predictable grid connection processes that prioritise credible and mature projects, reflecting strategic priorities where appropriate.

3. Approach energy and water use efficiency holistically, factoring in the complexity of when, where and how energy and water are used.

4. Design transparent and coordinated tariff frameworks that allocate the costs and risks of large loads fairly.

5. Create clear frameworks that enable data centres to contribute actively to grid flexibility, security, stability and sustainability.

6. Make dialogue and experience sharing central to inform decision making in this fast-moving field.

IEA. CC BY 4.0.

Introduction

Data centres are critical infrastructure at the heart of the artificial intelligence (AI) revolution. As AI capabilities and use expand rapidly, technology companies are investing heavily in data centre capacity. In 2025, investment in data centres grew by 17% to exceed USD 550 billion – more than global investment in oil and gas production. This rapid buildout is putting pressure on the energy sector to keep pace, with electricity systems facing the most immediate challenge.

AI is contributing to renewed electricity demand growth in advanced economies. After years of stagnation, electricity demand in advanced economies grew by 2.1% in 2025, its second consecutive year of growth above 2%. Data centres have emerged as a key – although not the only – driver of this growth, accounting for more than 20% of the increase. According to the latest analysis from the International Energy Agency (IEA), electricity consumption by data centres grew by 17% in 2025 and is projected to double over the next five years.

As data centres expand and grow larger on average, they are challenging planning frameworks designed for slower and more predictable demand growth. Data centres also tend to cluster near one another, often close to urban areas, concentrating pressure on electricity networks, land and water resources, and local communities. The IEA's 2026 report Key Questions on Energy and AI highlights physical constraints in energy systems and manufacturing supply chains that are holding back data centre development. These include grid connection delays, limited availability of transformers and other equipment, permitting bottlenecks, access to electricity and the ability of local systems to absorb rapid load growth. Planning is further complicated by uncertain project pipelines: not all announced data centres will materialise, and data centre projects can be developed faster than traditional network planning cycles. Depending on the electricity mix, rising electricity consumption can also increase emissions from power generation, putting pressure on the sustainability strategies of both countries and companies.

These pressures are prompting policy makers, regulators and companies to revise how large loads are assessed, prioritised and connected, with the aim of maintaining security of supply, affordability and sustainability. Where connection requests exceed available grid capacity, dedicated large-load frameworks and queue reforms aim to identify viable projects, allocate capacity and improve system planning. Flexible connections, battery storage, demand response, on-site power generation and grid-enhancing technologies offer additional ways to accommodate grid connections more efficiently. Data centre development also

IEA. CC BY 4.0.

raises questions about who should pay for dedicated grid connections and wider network reinforcement. Once facilities are operational, their scale and operating characteristics can affect system adequacy, stability and reliability, making their continued integration an important consideration for grid operators and policy makers.

Data centre types, workloads and the rise of AI

There are several types of data centres, each with different electricity consumption profiles and other characteristics. How they use electricity and interact with the power system depends on both the type of facility and the workloads it hosts.

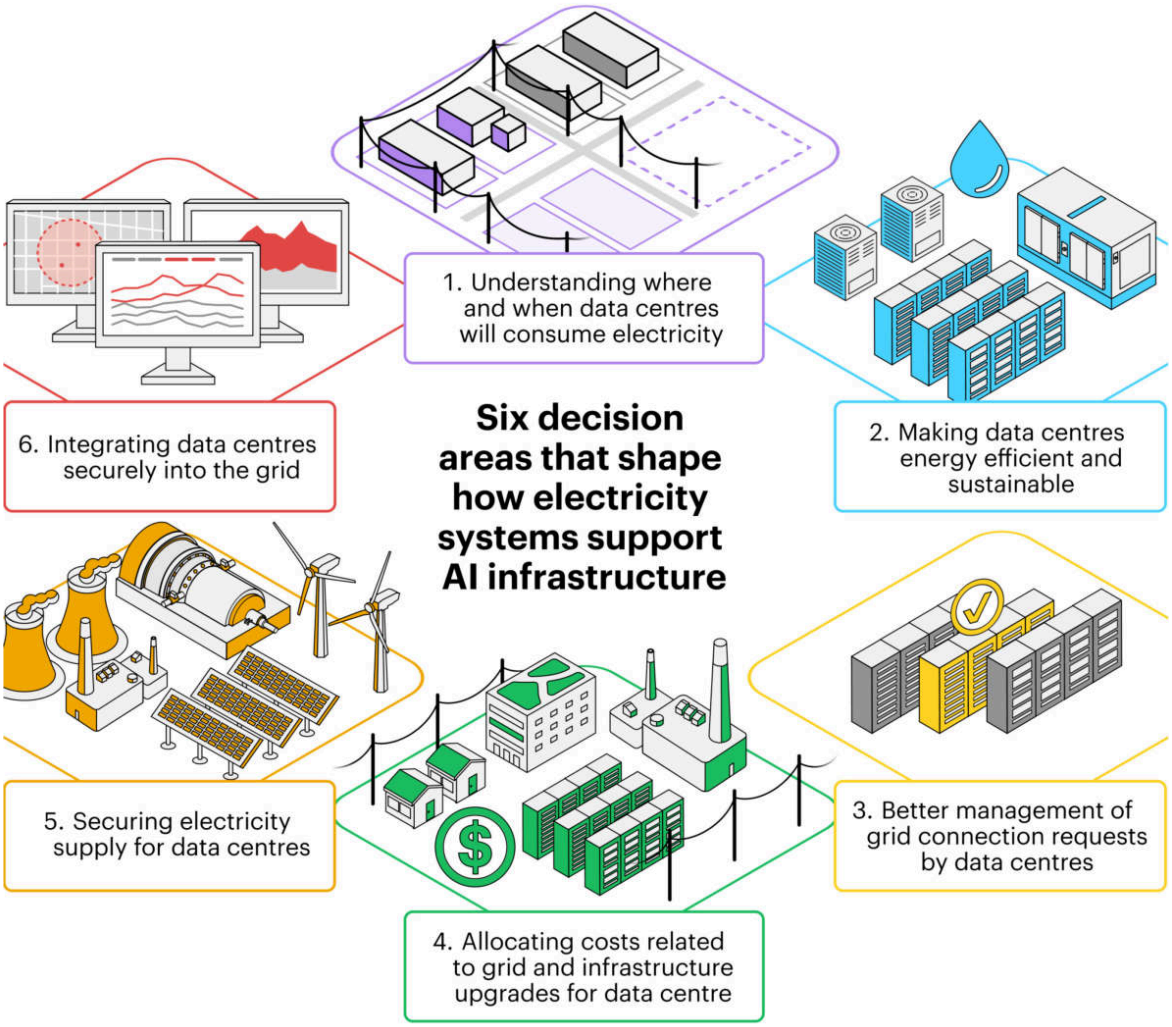
- **Data centre type.** There are three broad data centre archetypes: enterprise data centres, operated primarily on-site for an organisation's own data processing and storage needs; co-location and service provider data centres, which provide customers with space or computing services; and hyperscale data centres, which are massive, purpose-built facilities designed and typically used by a single large consumer, such as a major cloud or digital service provider. These types differ in scale, infrastructure efficiency, and the division of control over facilities and information technology (IT) equipment. The differences affect how electricity demand can be measured, managed and supplied.
- **Workload.** Data centres may host conventional cloud and enterprise computing, such as email or cloud document storage; AI model training, which develops or updates AI models; or AI inference, which uses trained models to generate outputs or make decisions. Inference ranges from latency-sensitive services that require near-real-time responses to less time-sensitive batch processing. These distinctions influence where and when electricity is consumed, equipment utilisation, load profiles and potential flexibility.

Rising demand for AI is driving growth in AI-optimised data centres, which can feature higher rack power and heat densities, specialised cooling and power delivery and, in some cases, more abrupt short-term changes in IT load. This shift is bringing existing data centre planning and integration challenges into sharper focus, but these characteristics are neither universal nor exclusive to AI. AI workloads occur across all three data centre types, although not every facility is equipped to support them, and some host a mix of AI and non-AI workloads. "AI data centre" is therefore used here as shorthand for a facility designed or used primarily for AI workloads, rather than as a separate type of data centre. Its impact on the electricity system depends on its workload mix, scale, design, connection and operation.

IEA. CC BY 4.0.

Taken together, these issues show that integrating data centres into electricity systems involves a series of interdependent decisions rather than a single connection decision. These concern when and where demand emerges; data centre design and efficiency; permitting, queue management and connection arrangements; cost allocation and tariffs; power sourcing; and data centre operation once connected. For clarity, the framework below groups these decisions into the individual chapters of this report. In practice, they are closely linked, and choices in one area can affect how the benefits, costs and risks of development are distributed among data centre developers, electricity consumers and local communities. Governments should therefore consider these policies in relation to one another, recognising that the appropriate mix of measures will depend on local grid conditions, market structures and policy priorities.

Six decision areas that shape how electricity systems support AI infrastructure



IEA. CC BY 4.0.

IEA. CC BY 4.0.

Six decision areas, central policy questions and principal actors and stakeholders that shape how electricity systems support AI infrastructure

Decision area	Central policy questions	Principal actors and stakeholders
1. Understanding where and when data centres will consume electricity	How quickly will electricity demand from data centres grow, what will their load profiles be, where will facilities be located and which approaches can best accommodate their growth?	• Governments • Data centre operators • Planning authorities • System operators • Market operators • Regulators • Generators • Local communities
2. Enhancing the energy efficiency and sustainability of data centres	How can the design of data centre buildings and the technologies within them minimise electricity and water use?	• Data centre operators and tenants • Technology providers • Energy efficiency agencies • Regulators • Water authorities
3. Better management of grid connection requests by data centres	How can credible projects be prioritised, approved and connected quickly when grid capacity is scarce?	• Permitting authorities • System operators • Network operators • Regulators
4. Allocating costs related to grid and infrastructure upgrades for data centres	How should the costs of grid connections and associated infrastructure upgrades be allocated?	• Regulators • Utilities • System operators • Governments • Ratepayers
5. Securing electricity supply for data centres	How can data centres secure the electricity they need and how should it be sourced? Is this electricity supply aligned with the sustainability goals of the entities involved?	• Data centre operators • Utilities • Generators • Regulators • System operators
6. Integrating data centres securely into the grid	Once connected, how can data centres interact with the grid in ways that support its secure and stable operation?	• System operators • Data centre operators • Data centre clients • Regulators • Network code compliance entities

Policy responses across these decision areas are developing rapidly, creating opportunities to learn from international experiences. Many measures are recent, and their effects have not yet been fully established. Rather than providing a definitive assessment, the chapters that follow map emerging approaches across Group of Seven (G7) countries and beyond. They identify

IEA. CC BY 4.0.

common principles and trade-offs that can support timely connections, fair cost allocation and reliable long-term system planning.

This report forms part of the IEA's broader work on energy and AI, which includes the following elements:

- **Publications:** The 2025 IEA special report Energy and AI provided the first global, system-wide analysis of the links between energy and AI, examining both the energy needed to power data centres and the potential for AI to optimise and transform the energy sector. Key Questions on Energy and AI, published in 2026, updated and extended this analysis in light of rapidly evolving technologies, markets and energy sector constraints.
- **Observatory:** The IEA's Energy and AI Observatory provides up-to-date data and analysis on the growing links between the energy sector and AI. This first-of-its-kind platform is developed and maintained by the IEA, with valuable contributions of data and insights from partners in the energy and technology sectors. It complements the IEA's publications on this topic.
- **Convenings:** The IEA has hosted a range of high-level convenings with the technology sector, the energy industry, civil society and governments on energy and AI. These include the 2024 Global Conference on Energy and AI, a series of workshops for Canada's G7 presidency in 2025 that helped inform the development of the G7 Energy and AI Work Plan, and an industry-government dialogue on energy and AI for France's G7 presidency.

Prepared at the request of France's 2026 G7 presidency, this report focuses on strategies for integrating data centres into electricity systems as AI-related electricity demand grows.

IEA. CC BY 4.0.

1. Understanding where and when data centres will consume electricity



Given the wide range in projections of growth in electricity demand from data centres, a central challenge for power system planners is determining how much load will materialise, when it will emerge, what its daily and seasonal profile will be and where it will be located. As data centre demand grows and facility sizes increase, traditional planning frameworks designed for gradual and more predictable load growth are being challenged. Managing these large, uncertain loads will require closer co-ordination across stakeholders and clearer information on both demand and location.

Load projections

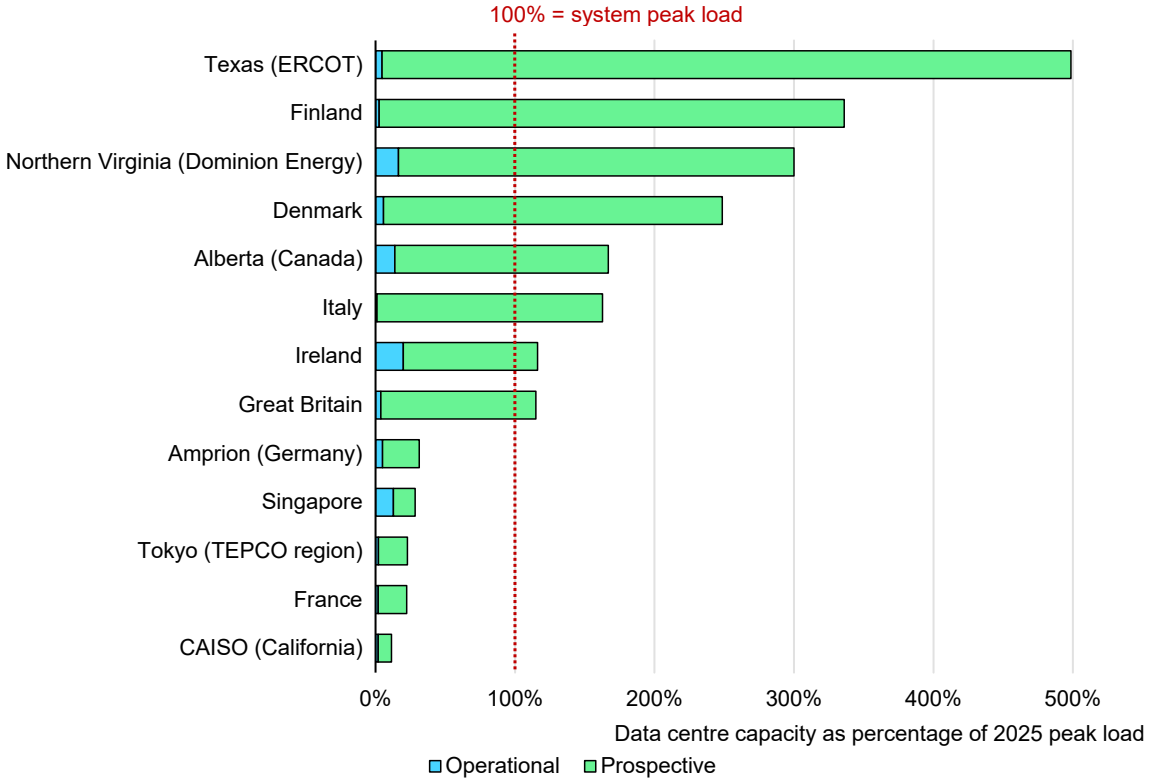
Electricity demand from data centres is expanding rapidly, but its future scale and timing remain highly uncertain. System operators are increasingly facing connection requests that, in aggregate, far exceed current system peak load in some regions. Some requests may be speculative, in particular where developers have not yet secured customers or contracted workloads. Such requests can involve greater uncertainty than projects submitted by prospective end users or backed by existing or prospective customers. A central challenge is therefore to assess the credibility of project applications.

The figure below illustrates the contrast between connection queues and system peak load in key regions. While operational and committed capacity remains relatively limited in most systems, the combined capacity of projects in connection queues or application pipelines is often several times larger. In some jurisdictions, these pipelines amount to multiple times the current peak load, reflecting both strong expected growth and significant uncertainty. As a result, not all projects in these pipelines are expected to be built.

IEA. CC BY 4.0.

This creates a fundamental planning challenge: networks must prepare for potentially very large new loads without clear visibility of how much capacity will be realised. Overestimating demand can risk creating stranded or underutilised grid investments, while underestimating it can lead to bottlenecks and delays.

Operational and prospective data centre capacity relative to system peak load in selected jurisdictions



IEA. CC BY 4.0.

Notes: CAISO = California Independent System Operator; ERCOT = Electric Reliability Council of Texas; TEPCO = Tokyo Electric Power Company. Data are indicative and not directly comparable across markets. Sources vary in whether they report IT load, total electrical load, grid connection requests, committed projects or policy targets. Queue maturity and reporting dates also differ.

Source: IEA analysis based on data from CBRE, Dominion Energy, ERCOT, the Alberta Electric System Operator, Fingrid, Terna, the Office of Gas and Electricity Markets, Cushman & Wakefield, TEPCO, Réseau de Transport d'Électricité, CAISO, Energinet, EirGrid and the Commission for Regulation of Utilities.

Several features of data centre development contribute to this uncertainty. Even when projects proceed, actual electricity demand does not generally reach contracted capacity immediately after initial energisation. Instead, it ramps up as buildings, server capacity and customer workloads are brought online progressively, adding uncertainty about when and how much demand will materialise. Developers often seek to secure multiple connection options early in project development, particularly in tight markets where access to power has become a key determinant of project viability. This creates a degree of project optionality, where some requests reflect firm investment plans, while others function as placeholders or strategic options. Rapid growth in the sector has also

IEA. CC BY 4.0.

attracted new entrants seeking to secure land, permits and grid access before making final decisions on scale, technology or location. Uncertainty is further compounded by a mismatch in timelines, where data centres can move from announcement to operation in two to three years, while network upgrades can take much longer.

The current scale of speculative and duplicate connection requests can create challenges for both developers and system operators. For developers, these requests can slow access to scarce grid capacity. For system operators and planners, they can obscure underlying demand growth and complicate efficient network and system planning.

Policy responses have therefore begun to focus on making connection pipelines more credible. Chapter 3 examines this issue in its discussion of queue management and connection arrangements.

Alongside efforts to improve the credibility of connection pipelines, approaches to load forecasting are increasingly being adapted to account for the characteristics of data centre demand. They use indicators such as the stage of project development, formal connection applications, financial commitments and signed power supply or network service agreements. In Australia, the Australian Energy Market Operator screened 44 GW of data centre connection requests for its demand forecast, and assessed 7.9 GW as likely to proceed, based on project and grid connection status, supporting evidence and possible duplication. In the United States, the Electric Reliability Council of Texas (ERCOT) publishes an unadjusted forecast based on large-load projections from transmission service providers and an adjusted forecast that reflects observed project behaviour, including average delays, load profiles by type and average project realisation rates. More broadly, forecasts are increasingly incorporating assumptions on how demand materialises in practice, including ramp-up profiles from initial energisation to full load, typical utilisation rates relative to contracted capacity and the likelihood of delays or cancellations.

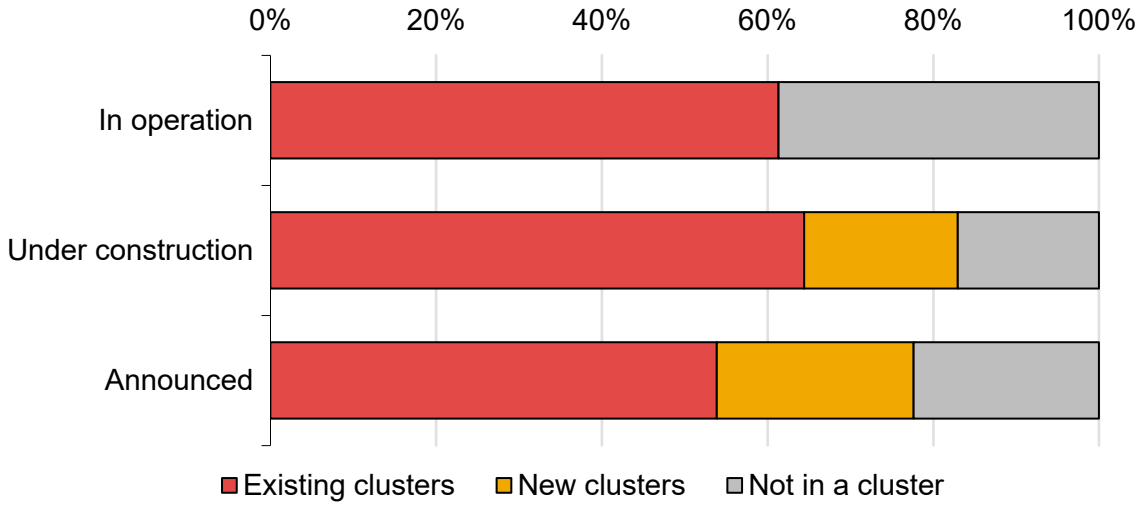
These dynamics have important implications for policy and co-ordination. Improving visibility of the scale and timing of data centre developments can reduce system costs by avoiding premature or unnecessary investment in generation capacity or grid infrastructure to support future loads that may not ultimately materialise. Early co-ordination among developers, system operators and regulators is therefore increasingly important. Measures such as stronger financial commitments for new large loads, milestone-based screening, phased connection arrangements and greater transparency about expected load ramp-up can help align network and capacity planning more closely with the actual outlook for electricity demand from data centres.

IEA. CC BY 4.0.

Spatial planning

In most jurisdictions, data centre siting remains largely applicant-led. Developers usually select sites based on land availability, access to fibre-optic networks, customer proximity, tax incentives, permitting conditions, power and water availability and commercial strategy. The impacts on the electricity system are then assessed through a project-specific connection study. This approach can work where infrastructure and resources are readily available, but clustering in capacity-constrained locations can delay connections and concentrate pressure on electricity networks, water resources, land use and surrounding communities. These pressures intensify as data centre development becomes more geographically concentrated. IEA analysis shows that around 60% of global data centre capacity was located in clusters in 2025. On a cumulative basis, this share would exceed 65% if projects under construction are completed and 70% if all announced projects are realised. The latter estimates are conditional on project pipelines and should not be interpreted as forecasts.

Geographical concentration of data centres by development stage, 2025



IEA. CC BY 4.0.

Notes: Data centre clusters refer to groups of facilities with a combined design capacity of at least 500 MW within a 100 km radius, identified using the density-based spatial clustering of applications with noise (DBSCAN) clustering algorithm. Announced projects may not all be realised.
Source: IEA analysis based on BloombergNEF (2025), as presented in IEA (2026), Key Questions on Energy and AI.

Locational signals can steer projects towards places where large loads can be accommodated with fewer delays, reinforcements and cumulative local impacts. The objective is not necessarily to prevent clustering but to direct development towards locations better able to accommodate the associated infrastructure needs and local effects. Emerging approaches generally seek to guide rather than prohibit development by improving visibility of available capacity, enabling suitable

IEA. CC BY 4.0.

sites or incorporating location into connection decisions. Together, these tools can help align data centre growth with grid, land, water and local conditions.

One approach is to improve the information available before formal connection applications are submitted. Capacity maps, indicators of available grid capacity and data on capacity available to new demand can show both current conditions and how conditions may change as a result of planned network development, existing reservations and connection timelines. Elia, Belgium's electricity system operator, provides a location-based view of how much additional load, generation or storage could be accommodated in different target years and under different flexibility assumptions. Spain's transmission system operator, Red Eléctrica, publishes the capacity available at each transmission node for new demand to access the network. These tools do not replace detailed connection studies, but they can reduce uncertainty before developers commit to a site.

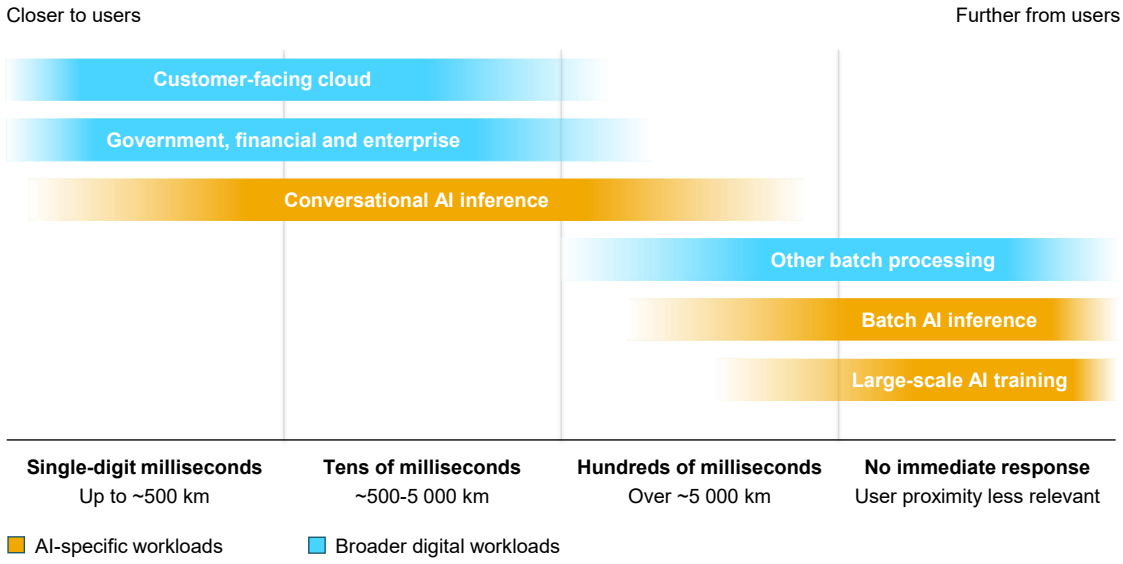
Greater visibility of available capacity can steer projects towards feasible locations, but it may also concentrate speculative requests at apparently favourable nodes. It should therefore be combined with application screening, project maturity criteria and effective queue management (see Chapter 3) to help ensure that scarce capacity is allocated to projects that are likely to proceed.

Governments, utilities or network operators can also identify in advance locations where grid infrastructure, land, permitting conditions and other supporting resources are more favourable. AI or data centre growth zones are an emerging form of this approach, clustering projects in locations assessed as able to accommodate them. The United Kingdom has designated five AI Growth Zones, selected using criteria including power availability, water supply and wastewater infrastructure, digital connectivity, local support and community benefits. France uses a broader, grid-focused model that is not specific to data centres. The French state and the transmission system operator, Réseau de Transport d'Électricité (RTE), have pre-reserved capacity at four large, buildable land plots close to high-voltage transmission infrastructure to accelerate connections for very-high-power consumption facilities. EDF is also offering former power generation sites for data centre projects with capacities of several hundred megawatts.

Grid suitability alone, however, does not make a location equally suitable for every workload. Customer-facing cloud services and transaction-oriented government, financial and enterprise services are relatively sensitive to latency, i.e. the delay in transmitting and processing data. Large-scale AI training, batch AI inference and other batch processing are generally more latency-flexible and can be located further from users or data sources. Conversational AI inference falls between these cases, depending on the application and service architecture. Locational price signals and designated zones therefore have greater scope to influence the siting of latency-flexible workloads than services tied closely to users or data sources.

IEA. CC BY 4.0.

Indicative latency requirements of selected data centre workloads and implications for data centre proximity to final customers



IEA. CC BY 4.0.

Notes: Positions and distances are indicative. Time bands are categorical and not to scale; faded ends show variation and overlap. Distance equivalents are based on Google Cloud's approximation of 1 millisecond of round-trip time per 100 km of network travel, assuming routes twice the straight-line distance. These do not represent siting limits. Actual latency varies with routing, access networks, processing and architecture. Latency figures exclude communication within the data centre.
Source: IEA analysis based on Google Cloud (2026) and Amazon Web Services (2025).

These differences should inform where zones are designated. Early consultation with data centre operators can establish whether grid-ready sites also meet the latency, connectivity and proximity-to-user requirements of their intended workloads. Otherwise, designated zones may attract limited industry uptake and may not achieve their intended objectives for efficient grid use and economic development.

In more constrained systems, location can become part of the formal connection decision. Under Ireland's Large Energy Users Connection Policy, system operators must consider whether a proposed data centre connection is in a constrained or unconstrained part of the electricity network when determining whether it can be accommodated. Location therefore does more than guide developers towards suitable sites; it can directly influence the assessment of a connection application.

For comparison, the table below distinguishes three forms of spatial planning:

- **Capacity visibility:** Publishing maps or data on current and expected grid capacity, constraints and connection prospects before a formal application is submitted.
- **Site enabling:** Identifying or preparing locations where land, permitting conditions and grid infrastructure make large loads easier or faster to accommodate.

IEA. CC BY 4.0.

- **Locational screening:** Formally considering a project's location and local network conditions as part of the connection assessment. Location may consequently affect whether, when and on what terms the project can connect.

Spatial planning measures in selected jurisdictions

Jurisdiction	Capacity visibility	Site enabling	Locational screening
France		●	●
Japan	●		
United Kingdom	●	○	
United States	●	●	
European Union	○		
Ireland	●	○	●
Spain	●		●
Singapore		●	
China		●	

Key: ● measure in force ● subnational, utility-led or not specific to data centres ○ announced or under development
Note: Status as of July 2026.
Source: IEA analysis based on official government, regulator and system operator publications; see the annex.

Many of these approaches have only been recently introduced, and their effectiveness will depend on implementation. Identifying preferred locations does not by itself ensure that electricity supply, water and other supporting infrastructure will be available when required. Nor does it remove the need to assess the cumulative effects of several large projects locating in the same area. Spatial planning should therefore consider not only existing grid headroom but also whether planned network capacity can be delivered when needed. It must also take account of access to electricity, constraints on land and water availability, system resilience and effects on host communities. Preferred-site status should complement, rather than replace, detailed connection, planning and environmental assessments.

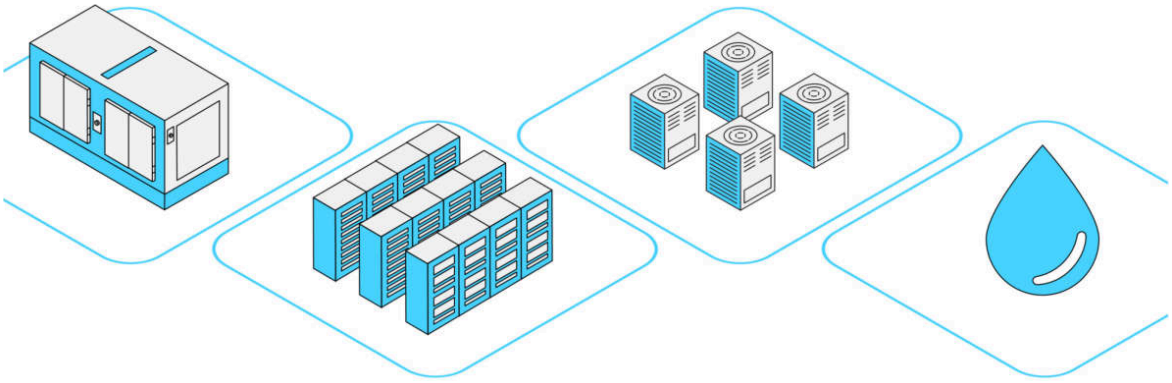
Spatial planning should move from reactive, project-by-project assessment towards dynamic, plan-led frameworks for major loads – including data centres. This requires information to flow in both directions. Governments and system operators can model alternative demand and network development scenarios, publish regularly updated capacity information and identify a portfolio of suitable locations based on deliverable grid capacity, clean power availability, land, water,

IEA. CC BY 4.0.

resilience and local impacts. Data centre developers and operators should, in turn, provide credible, sufficiently granular and regularly updated information on the location, scale, phasing and expected load profiles of planned developments so that they can be reflected in network planning. Site designations and accelerated processes should be reviewed as infrastructure plans and project pipelines evolve and should remain linked to credible project delivery and the efficient use of scarce capacity.

IEA. CC BY 4.0.

2. Enhancing the energy efficiency and sustainability of data centres



Energy efficiency is often described as the "first fuel": energy that does not need to be produced, transported or consumed. Efficiency gains have already been rapid. IEA analysis shows that the performance per watt of leading AI processors and accelerators improved by around 30-40% per year on average between 2016 and 2025.[1] Yet total electricity demand from AI is still expected to rise as its use expands. For policy, the central question is how to sustain and spread these gains across facility design, power equipment, IT equipment and workloads so that growth in computing demand does not translate into equivalent growth in electricity and water use. Every megawatt of demand avoided can reduce requirements for generation, grid infrastructure, transformers, land, water and connection capacity.

Water use is also an important dimension of data centre resource efficiency. Cooling choices can shift resource use between electricity and water: systems that reduce the electricity used for cooling may increase water consumption, while measures that minimise water use may increase electricity requirements under some conditions. These outcomes depend on facility design, climate and operating conditions, so energy and water performance should be assessed together and in relation to local resource constraints.

Where efficiency is measured

Data centre efficiency is determined across three connected layers. Facility design and power equipment affect the energy and water needed for cooling and power delivery; IT equipment and utilisation determine how much of the electricity reaching servers produces useful work; and software and workload design determine how much computing is required to perform a task. Improvements in

[1] The estimate is based on floating-point performance per watt across successive generations of leading graphics processing units and tensor processing units. Numerical precision and architecture vary across the series; the estimate therefore does not represent a directly comparable measure of energy savings for a fixed workload.

IEA. CC BY 4.0.

one layer can also produce savings elsewhere. For instance, more efficient or better-utilised servers can reduce both IT electricity consumption and the cooling capacity required to remove the resulting heat.

At the facility level, decisions made during planning and construction determine much of a data centre's long-term resource use and can be costly to reverse. Location and climate affect the potential to use outside air to reduce the energy needed for cooling, while building layout, materials, insulation, airflow, cooling systems and power distribution determine the energy needed to support IT equipment. In warmer climates in particular, insulation, airtightness and solar protection can reduce external heat gains. General building energy codes cover these aspects of the building envelope but do not always address the continuous internal heat loads and electrical systems specific to data centres. Standard 90.4 of the American Society of Heating, Refrigerating, and Air-Conditioning Engineers provides dedicated mechanical and electrical efficiency requirements and has been incorporated into codes such as the Washington State Energy Code. Facility planning also determines whether waste heat can be recovered, depending on its temperature and the presence of a suitable user nearby.

At the IT level, efficient servers, storage and networking equipment can increase the amount of computing delivered per unit of electricity consumed. Power management and workload consolidation can increase the utilisation of installed equipment by reducing the number of lightly loaded servers and their associated cooling needs. The amount of energy that can be saved depends on how the data centre is operated and who controls its servers. Older enterprise systems may contain lightly used servers that can be consolidated. Where suitable, moving these workloads to more efficient, shared infrastructure, including cloud facilities, can reduce energy use per unit of computing through higher utilisation. Hyperscale cloud operators generally make greater use of active fleet and power management, while control over IT equipment in co-location facilities may rest with tenants.

Although examining utilisation can help identify underused capacity, it does not show how efficiently equipment delivers useful work and should not be used as a standalone performance threshold. Policy should instead focus on measured computing or storage performance per unit of power for comparable equipment and workloads. Policy tools for conventional servers are relatively well established. The voluntary US ENERGY STAR specification uses the Server Efficiency Rating Tool (SERT), developed by the Standard Performance Evaluation Corporation (SPEC), to assess active-state performance per watt and idle-state power consumption. EU Ecodesign rules incorporate SERT-based measures of active-state efficiency and idle power, alongside mandatory power supply efficiency requirements. SPEC is also developing a successor benchmark for the power use and performance of next-generation servers. Metrics remain

IEA. CC BY 4.0.

less mature for AI accelerators, whose energy efficiency cannot be reliably inferred from theoretical computing capacity alone. Procurement and equipment upgrade policies should weigh expected operational savings against remaining equipment life and opportunities for reuse or refurbishment.

Software and workload efficiency concerns how much computing is needed to complete a task. Current policy frameworks do not generally incentivise matching model capabilities and computing requirements to specific tasks, for example by selecting less energy-intensive models where suitable. Most focus on facility performance or disclosure rather than whether a less energy-intensive model could deliver the required result. Commercial incentives may, however, encourage this approach because routing simpler tasks to smaller models can reduce computing costs. Workload scheduling can also improve efficiency where it enables greater batching or equipment utilisation. Shifting otherwise unchanged electricity use to a different time primarily provides grid flexibility and is addressed in Chapter 6. Estimating the electricity used by AI applications nevertheless remains challenging because training sessions may be shared across models, and computational tasks may be carried out in different locations and at different times. Research estimates that the computing resources required to achieve a given level of language model performance have halved roughly every eight months, but energy requirements differ greatly among text, image, video, reasoning and agentic applications. Energy-per-task indicators are therefore useful only for comparing similar tasks and performance levels. Policy is at an earlier stage in this area. The EU AI Act requires general-purpose AI providers to document known or estimated model energy consumption, while the European Commission is developing standardised methods covering training and inference. Harmonised measurement approaches, benchmarks and disclosures can improve transparency and provide a more consistent basis for comparing the energy requirements of similar tasks at comparable performance levels.

The table below summarises the three efficiency layers and links each to the principal levers, policy tools, indicators and actors. Operational control and access to data may be shared, so reporting obligations should be assigned to actors able to influence and measure the relevant outcome.

IEA. CC BY 4.0.

Principal data centre efficiency levers and policy approaches

Efficiency layer	Principal levers	Main policy tools and indicators	Principal actors
Facilities, buildings and power systems	Siting and building design; cooling and power delivery; heat reuse	Building, mechanical and electrical standards; PUE, WUE and ERF	Data centre developers; facility operators
IT equipment	Efficient hardware; power management; consolidation and utilisation	Product standards and procurement; server efficiency and utilisation	Equipment manufacturers; data centre operators and tenants
Software and workloads	Efficient models; selection of models appropriate to the task; workload scheduling	Measurement and disclosure; energy per task	Model developers for training; cloud and application providers for inference

Notes: PUE = power usage effectiveness; WUE = water usage effectiveness; ERF = energy reuse factor. The framework covers energy and water use during operation. It does not assess embodied energy, emissions or material use associated with constructing facilities or manufacturing and disposing of equipment.

Reporting and facility performance

Existing data centre efficiency policy focuses mainly on facility performance. Reporting is the most prevalent intervention across the jurisdictions reviewed because site energy use, water use and waste heat can be measured without defining the value or performance of the computing delivered. Binding requirements for power usage effectiveness (PUE), water usage effectiveness (WUE) and heat reuse remain less common.

Reporting can also provide the basis for progressively stronger policy. In the European Union, data centres with at least 500 kW of installed IT power demand must report their energy and water use, waste heat and indicators including PUE and WUE. The European Commission has assessed the first reporting cycle, consulted on a draft EU rating scheme and launched a separate preparatory study on minimum performance standards. This illustrates a possible progression from reporting to comparable benchmarks, and ultimately to requirements where indicators are sufficiently mature.

Where policy goes further, it generally addresses three facility-level outcomes. PUE measures total facility energy relative to IT energy and therefore captures the electricity used for cooling, power distribution and other supporting systems. WUE relates on-site water use to IT energy and, when considered alongside PUE, can reveal whether lower electricity use is being achieved at the cost of higher water consumption. Heat reuse requirements encourage the use of waste heat to replace energy used elsewhere. These measures are complementary: none provides a complete picture of facility performance or assesses the efficiency and computational productivity of the IT equipment itself.

IEA. CC BY 4.0.

Power and water usage effectiveness, and heat reuse provisions in selected jurisdictions

Jurisdiction	PUE	WUE	Heat reuse	Reporting
Canada	●	●	●	●
France	●	●	●	●
Germany	1.3	●	10-20% ERF	●
Italy	●	●	●	●
Japan	1.3	●	●	●
United Kingdom	1.3	0.4 L/kWh	●	●
United States	●	●	●	●
European Union	●	●	●	●
Ireland	1.3	●	●	●
Spain	●	●	●	●
Singapore	1.25	2.0 L/kWh	●	●
China	1.2	0.8 L/kWh	●	●

Key: ● national requirement or government target ● voluntary measure only ● no national measure specific to data centres identified

Notes: PUE = power usage effectiveness; WUE = water usage effectiveness; ERF = energy reuse factor. Singapore's 1.25 PUE standard applies to new data centres only.

Source: IEA analysis based on official legislation and policy documents; see the policy collection in the annex.

Binding requirements remain much less common than reporting. Germany provides one of the most comprehensive examples. Data centres beginning operation from July 2026 must achieve an annual average PUE of 1.2 and an energy reuse factor of at least 10%, rising to 20% for facilities beginning operation from July 2028. Electricity used solely to raise recovered heat to a useful temperature is excluded from the PUE calculation. This recognises that heat recovery can increase energy use within the facility while reducing energy consumption across the wider energy system.

IEA. CC BY 4.0.

Facility indicators are comparable only when they are calculated consistently. PUE is also sensitive to IT loading because cooling, power conversion losses and other facility overheads do not all vary in proportion to IT electricity use. A lower PUE at higher load does not, by itself, indicate that more useful computing is being delivered per unit of electricity. Germany assesses the annual average PUE, Singapore specifies performance at full IT load and Japan uses an operator-average benchmark. Reporting rules should therefore specify the period, loading condition, measurement point and facilities included, rather than relying on the PUE value alone.

Comparability is more difficult for WUE because the numerator – the quantity of water used in the calculation – can represent either water withdrawal or water consumption. Water withdrawal counts water taken from external sources, including water subsequently discharged, whereas water consumption counts the portion not returned, principally because it evaporates. Withdrawal indicates gross demand on water supplies and infrastructure; consumption more closely represents water removed from immediate availability. Operators currently use both approaches.

Water measures in water usage effectiveness calculations by selected hyperscalers

Company	Water measure	How it is calculated
Google	Freshwater consumption	Freshwater input minus water returned; Gemini inference methodology using ISO Category 2
Microsoft	Water withdrawal	Water entering the data centre boundary, including water later discharged
Amazon Web Services	Water withdrawal	Water entering the data centre boundary, including water later discharged
Meta	Water withdrawal	Water entering the data centre boundary, including water later discharged

Note: The table describes the numerator used in published WUE reporting, rather than the companies' wider water accounting.
Sources: Google (2025), Measuring the Environmental Impact of Delivering AI at Google Scale; Microsoft (2026), 2026 Environmental Sustainability Report; Amazon Web Services, AWS Cloud Sustainability; Meta (2025), Environmental Data Index.

The choice of numerator can materially affect the reported indicator. Amazon Web Services, for example, converts an industry consumption-based WUE of 0.56 L/kWh to a withdrawal-based value of 0.84 L/kWh under its stated cooling system assumption, representing a 50% increase for the same modelled system. WUE reporting should therefore define the numerator, facility boundary, reporting period and IT energy denominator. The indicator should also be accompanied by data on absolute water withdrawal and consumption, water sources and local water stress.

IEA. CC BY 4.0.

Taken together, PUE and WUE provide important but incomplete measures of efficiency. PUE captures aggregate cooling, power conversion and distribution overhead at the facility level without distinguishing their respective contributions, while WUE relates facility water use to IT energy. Neither shows how efficiently electricity is delivered to and used by IT equipment or how much useful computing is delivered by software and models. Low values also do not necessarily imply low absolute electricity or water use. Responsibilities differ across the value chain: facility operators can report site performance, while tenants, cloud providers and model developers may be better placed to report equipment utilisation and workload-level indicators. The following section considers how measurement can extend across the different efficiency layers.

When does water become a constraint for data centres?

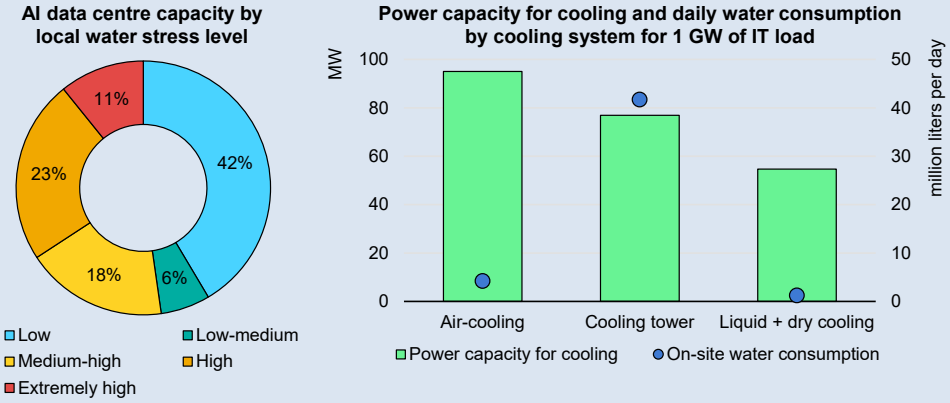
Water consumption associated with data centres arises from three main processes: around two-thirds is associated with primary energy supply and electricity generation, around one-quarter with the direct cooling needs of data centres, and the remainder with semiconductor and microchip manufacturing. Water use associated with electricity generation generally occurs at the location of generation and is often outside the direct control of the data centre operators. By contrast, water consumed for cooling within a data centre can be directly relevant for spatial planning decisions. Cooling also involves a trade-off between energy and water use, as technologies that consume less water may be more energy intensive and vice versa. Assessments of data centre sustainability should therefore consider both resources.

Water consumption by data centres has attracted scrutiny in regions experiencing water stress. Concerns about pressure on local supplies have contributed to community opposition and even moratoria on new developments (see the box on moratoria in Chapter 3). Assessing this risk requires consideration of both where facilities are located and how they are designed and operated.

An analysis of data centres dedicated to AI revealed that around 34% of identified capacity was located in areas with high or extremely high baseline water stress in September 2026 (figure left panel). However, the underlying datasets do not reveal the cooling technologies used in these regions, so the resulting on-site water consumption cannot be determined.

IEA. CC BY 4.0.

AI data centre capacity by local water stress and cooling system capacity and daily water consumption



IEA. CC BY 4.0.

Notes: The right panel assumes an IT load of 1 GW with a utilisation rate of 80%. Cooling capacity is derived from PUE and power supply equipment efficiency (Lei & Masanet, 2022); on-site water consumption is derived from site WUE.
Sources: Left panel: IEA analysis based on Epoch AI, AI Data Centers, and World Resources Institute Aqueduct 4.0. Right panel: IEA analysis based on Lei et al. (2025).

Cooling technologies are associated with a wide range of water and energy requirements (figure right panel). Cooling towers that use evaporation – and are therefore water intensive – are much less energy intensive. They can lower PUE relative to conventional air-cooled chillers but raise on-site WUE. Dry coolers limit water use but require more electricity, especially in warmer conditions. This trade-off is not fixed. Liquid cooling operating at higher coolant temperatures can reduce electricity demand and, when paired with a dry cooler, limit on-site water use. If the closed server loop instead transfers heat to an evaporative cooling tower, water is still consumed.

Aggregate water use figures also require local context. Virginia's 2024 legislative review found that data centres accounted for less than 0.5% of water withdrawals across the state, but their share of water use, excluding reclaimed water, reached 21% at one of the six utilities examined. State-wide totals can therefore obscure pressure on the water supplies serving individual sites.

These considerations can materially affect project development. In Chile, an environmental court partially annulled approval of Google's proposed Cerrillos data centre in 2024 and required the assessment to account for the effects of climate change on water availability. The company had separately proposed replacing water-based cooling towers with air-cooled chillers. The case illustrates the importance of assessing cooling design and future water conditions together.

Planning and efficiency policies should therefore assess PUE and WUE alongside absolute and peak water requirements, water sources, seasonal availability and local water stress. WUE thresholds and comparisons across jurisdictions should account for climatic and local conditions. Publishing assessments and proposed mitigation measures can support informed permitting and public discussion.

IEA. CC BY 4.0.

Wider efficiency priorities

Efficiency policy can also recognise and incentivise benefits beyond the data centre boundary. Data centre design and siting strongly influence whether waste heat can serve nearby users. At Hamina in Finland, recovered heat is expected to meet around 70-75% of local district heating demand. This benefit is captured by the energy reuse factor rather than PUE or task-level indicators and depends on the temperature of the recovered heat, network proximity and local demand. Planning and incentives can therefore encourage heat-ready designs where a viable user exists.

Data centre design also determines how efficiently electricity is converted and distributed between the grid connection and IT equipment. The IEA identifies higher-voltage direct-current systems and advanced power electronics as potential ways to reduce conversion stages and losses. Published estimates place power transformation and distribution losses at 0-2% in large-scale facilities and 2-5% in mid-size facilities. Solid-state transformers could offer additional control and material savings, but current designs may have lower peak efficiency and higher thermal requirements than conventional high-efficiency transformers, while large-scale deployment remains limited. Rapid deployment timelines and constrained equipment supply chains do not always favour the most efficient available designs.

Regularly updated standards and procurement requirements can raise minimum performance and encourage leading technologies, but measurement methods must also keep pace with changing compute architectures. Current EU ecodesign requirements for active state efficiency and idle state power, for example, do not apply to high-performance computing servers or servers with integrated accelerators, underscoring the need for revisions involving standards bodies, operators and manufacturers. Efficiency requirements should nevertheless account for reliability, equipment availability and part-load performance as demand ramps up, while pilots and research support can assist technologies that remain at earlier stages of development.

Near-term policy tools include consistent facility reporting, absolute energy and water data, building and electrical standards and equipment procurement requirements. Where doing so reduces total energy use, policy can also facilitate the consolidation of suitable workloads from underutilised on-premises systems into efficient shared infrastructure, including cloud facilities. Disclosure of IT utilisation and task-level metrics requires further standardisation. Efficiency can provide a given computing service with less electricity and infrastructure, but rapid growth in demand for computing services means that efficiency alone is unlikely to reduce aggregate data centre electricity demand in the near term. Its immediate value to power systems is to moderate demand growth and reduce or defer the

IEA. CC BY 4.0.

generation, grid and on-site infrastructure required to serve it. Efficiency should complement, rather than substitute for, the forecasting and siting measures discussed in Chapter 1 and the operational integration measures discussed in Chapter 6.

Long-term resilience of data centre infrastructure

A significant share of global data centre capacity is exposed to physical hazards that could affect both operations and the integrity of assets. One analysis finds that over half of global data centre capacity is chronically exposed to drought or heat stress, and that almost 80% is exposed to elevated wildfire, wind or flood risk. Another analysis finds that more than a quarter of data centre capacity in the United States could be exposed to high hail risk, and around 40% could be significantly exposed to tornadoes.

These risks could be exacerbated if data centres are built in new locations exposed to more severe risks. Constraints on land, water and electricity are already driving data centres beyond conventional market locations.

Geographical concentration could also amplify the risks posed by extreme events. Data centres have tended to form clusters because of factors such as the availability of IT and energy infrastructure and access to incentives. IEA analysis finds that data centres continue to be built in both existing and new clusters. In Europe, the share of data centre capacity located in clusters is set to exceed 50%, while the global share is expected to exceed 70%. This concentration could increase the impact of a single extreme event on the regional infrastructure supporting data centres.

Data centre facilities also represent highly concentrated investments. A large data centre campus could include multiple server buildings and power infrastructure components, representing a total capital investment of several tens of billions of dollars. This degree of capital concentration appears exceptional compared with conventional infrastructure investments, even when considering large and complex infrastructure projects. Community opposition and moratoria on the construction of data centres could also shift development to locations exposed to greater physical risks.

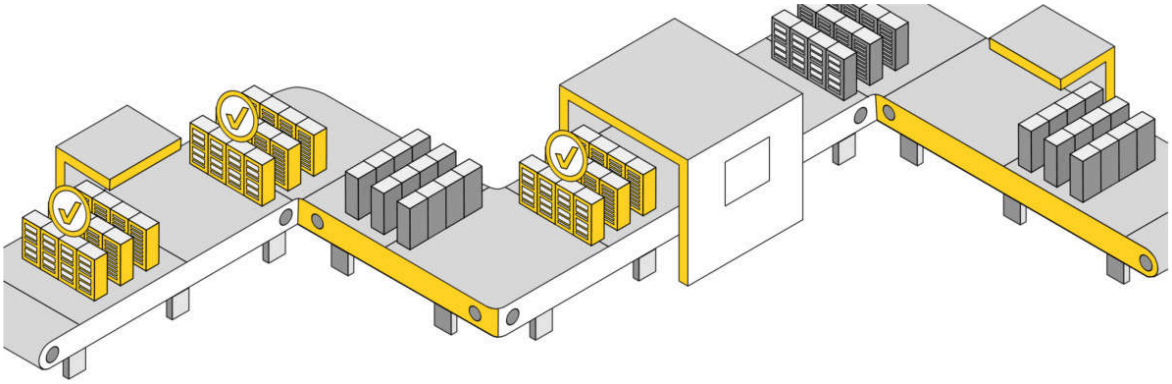
As a result, exposure to extreme events and increasing capital investment are driving changes in insurance and capital markets for data centres. Developers face rising insurance costs and may need to use alternative insurance instruments or multiple layers of insurance that do not necessarily provide full coverage.

IEA. CC BY 4.0.

Ultimately, reducing physical risks could entail several measures, including modelling climate and natural hazard risks at the earliest stages of site selection, strengthening infrastructure resilience, providing redundancy for power equipment and electricity and water supplies and minimising energy and water needs from the outset. Broader efforts – especially those that enhance the resilience of grid infrastructure – can also benefit the communities in which data centres are located. Redeveloping former industrial sites, which often have favourable attributes such as on-site infrastructure and available land, can also ease constraints on new land and infrastructure and reduce impacts on communities. A notable example is the Start Campus data centre hub in Sines, Portugal, which is being developed on the site of a former coal-fired power plant and is intended to use seawater cooling and renewable energy.

IEA. CC BY 4.0.

3. Better management of grid connection requests by data centres



Connection queues have grown rapidly in major data centre markets around the world. As discussed in Chapter 1, connection requests in jurisdictions such as Northern Virginia, Texas (ERCOT) and Finland would exceed current system peak load if they were all to materialise. Project development can also advance faster than grid connection processes, connection works and the wider network reinforcements needed to serve large new loads.

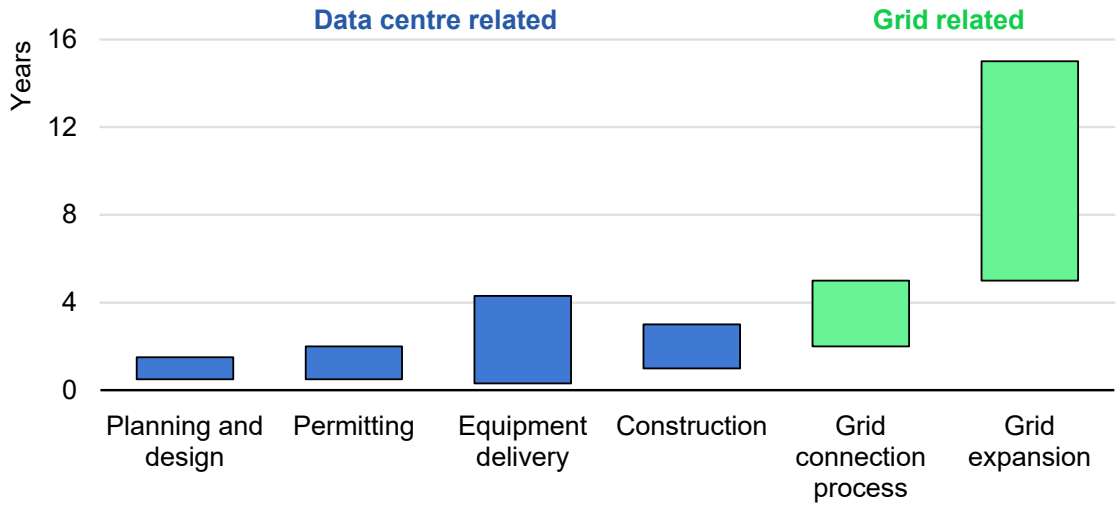
This mismatch in lead times is especially significant where grid capacity is already constrained. In such cases, effective queue management frameworks are critical to ensure that projects are credible, prioritise projects according to transparent and non-discriminatory criteria and streamline the necessary studies and approval processes. These frameworks can turn crowded application pipelines into a smaller set of viable projects that progress efficiently towards connection.

The design of the connection offer is also important. Once a project is considered credible, the key decision for the system operator and project developer is whether the project must wait for fully firm access or whether phased capacity, conditional access or non-firm arrangements could enable earlier connection while any necessary reinforcements are being completed. A firm connection allows a data centre to draw its contracted capacity without routine restrictions, while a non-firm connection allows the system operator to limit consumption under predefined conditions. A phased connection provides capacity progressively as the project develops or grid capacity becomes available and may include firm or non-firm access at different stages.

As data centre demand grows, predictable permitting frameworks, credible and efficient connection queues and clear access options are becoming increasingly important in determining where projects locate and how quickly they can proceed.

IEA. CC BY 4.0.

Indicative timelines for key processes in data centre development



IEA. CC BY 4.0.

Permitting

Permitting comprises the regulatory approvals that a data centre must obtain before construction can begin and, in some cases, during operation. It enables authorities to assess a project's impacts on the surrounding community and environment, attach binding conditions and provide the public with a formal opportunity to participate in the decision-making process. In principle, it is the process through which a data centre can earn its "social licence" to operate. A central tension in data centre policy is therefore how to preserve this protective function while responding to pressure to build faster. Required permits vary among jurisdictions but commonly include the following categories.

The speed, scale, novelty and complexity of the current data centre buildout are challenging permitting processes, especially where the permitting authority is split across several bodies or requirements vary significantly across jurisdictions. For developers, this can make project timelines less predictable and potential sites more difficult to compare.

Local concerns are often raised through permitting processes, including zoning or planning hearings, environmental reviews and appeals. These processes provide formal channels for communities to scrutinise proposed developments and raise concerns about land use, water, noise, emissions and local infrastructure. Where concerns are not resolved, these processes can delay or reshape projects or result in additional conditions.

IEA. CC BY 4.0.

Common permitting requirements for data centre projects

Permit type	Coverage	Main concern
Land use and zoning	Zoning, site approval and building permits	Site compatibility and local planning
Environmental review	Assessment of environmental impacts	Mitigation requirements and appeals
Water permits	Water withdrawals and discharge	Water availability and discharge quality
Air permits	Emissions from on-site power generation during construction and operation	Local air quality
Noise permits and conditions	Cooling equipment, transformers and generators	Noise limits and community impacts
Fire, hazardous materials and storage permits	Fuel, batteries and coolant chemicals	Safety and emergency response

Note: Permitting requirements vary by jurisdiction, site characteristics and facility design. The categories shown are indicative and may apply at different stages of project development or operation.

Permitting risks can also arise from rules that are not specific to data centres. In some jurisdictions, data centre approvals may be affected by broader environmental constraints, including rules on protected areas, local air quality or water availability. Improving permitting may therefore require better co-ordination among environmental, land use and infrastructure authorities.

To improve predictability and reduce delays, some governments are seeking to streamline approval processes or centralise responsibility, especially for projects that are considered to be strategically important. In the United Kingdom, data centres are being brought within the Nationally Significant Infrastructure Projects regime, allowing qualifying projects to be considered through a national rather than solely local permitting route. The aim is to provide a clearer and faster pathway for strategically important projects while retaining formal assessment and consultation requirements. France and Italy have taken similar steps to improve permitting timelines. France is streamlining and accelerating approvals for large strategic projects (PINM), while Italy has introduced a target of ten months for the single authorisation process for data centres. In Spain, the region of Aragon is implementing a two-step fast-track permitting procedure for strategically important investments to enable the rapid authorisation of major data centre programmes. For this procedure, a project must first be declared to be of “general regional interest”. It then enters the fast-track process, in which the regional government co-ordinates planning, environmental and building permits in a single file, provides a single point of contact and holds one round of public consultation.

Beyond streamlining, some authorities are also using permitting processes to attach additional conditions to data centre development. Germany’s Energy Efficiency Act adds requirements for data centre energy performance and waste

IEA. CC BY 4.0.

heat utilisation, while excess heat recovery is increasingly being considered in parts of the Nordic region. Singapore provides a more selective model. After temporarily pausing new data centre approvals, the government reopened development under more selective conditions, with a focus on energy efficiency, sustainability and efficient land use. The European Commission's proposed Cloud and AI Development Act would link compliance with European Union-wide data centre sustainability requirements to eligibility for designation within Data Centre Acceleration Zones. All relevant permitting procedures in these zones would have to be completed within 12 months.

Beyond reducing delays, a core policy challenge is how to make permitting more predictable, better co-ordinated and sufficiently transparent. Effective permitting frameworks should combine clear timelines, co-ordination across authorities, transparent public engagement and enforceable conditions addressing local environmental impacts. Streamlining can reduce unnecessary delays, but if it is perceived as weakening scrutiny or limiting local participation, it may intensify the opposition it is intended to address.

What data centre moratoria reveal about permitting and public confidence

As data centre projects grow and cluster, authorities have sometimes paused approvals or limited access to land and electricity networks. Such measures often arise when project pipelines outpace the ability of permitting, grid connection and wider infrastructure planning processes to assess and accommodate their cumulative impacts. They can be a response to growing public opposition and may signal weak public confidence in how projects are assessed and how their costs, risks and benefits are distributed.

In the United States, subnational authorities adopted nearly five times as many recorded moratoria in the first seven months of 2026 as in the whole of 2025. Planning and land use concerns were most prominent, alongside concerns about community impacts, electricity prices, water and the environment, although the increase may partly reflect greater public attention and more comprehensive tracking.

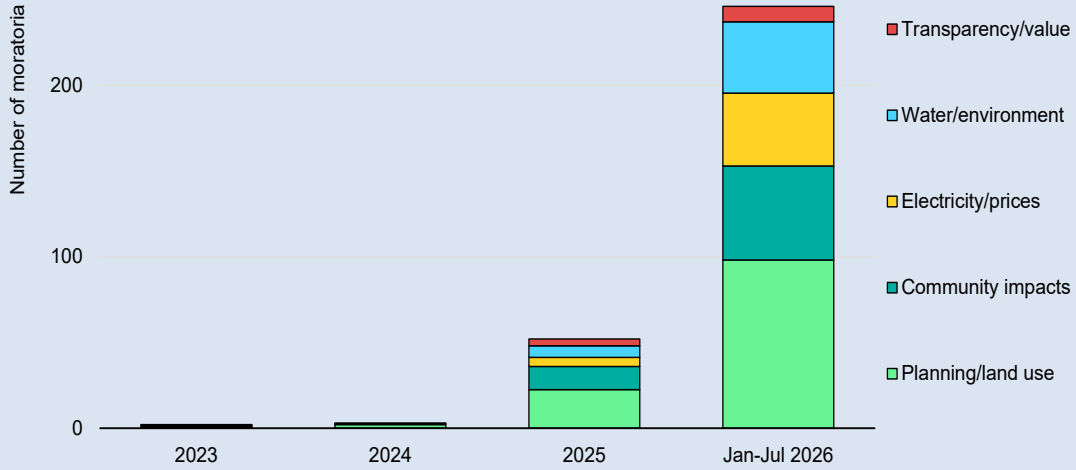
Recent state-level measures include New York's July 2026 pause on certain environmental permits for data centres of 50 MW or more. After the period covered by the figure, Texas also paused projects advancing through ERCOT's interconnection process in August 2026, pending an audit.

International experience shows how such pauses can evolve into managed development frameworks. Singapore's 2019 pause was followed by the competitive allocation of capacity based on energy efficiency, sustainability and economic contribution criteria. Amsterdam adopted tighter siting rules, while

IEA. CC BY 4.0.

Ireland linked new connections to project location, system conditions and additional electricity supply. In September 2026, Thailand paused 49 projects under construction and 117 awaiting approval while developing rules covering electricity, water, siting, environmental safeguards and community impacts. These cases show how regulatory gaps identified after projects have entered development can expose them to delays, redesign and development costs that may not be recovered.

Data centre moratoria in the United States by cited concern, 2023-July 2026



IEA. CC BY 4.0.

Notes: The analysis covers moratoria adopted by US states, counties, municipalities and townships. Each moratorium is counted once. Where multiple concerns are cited, its value is divided equally among them. Categories are IEA classifications.
Source: IEA analysis based on Tracking American AI Data Center Buildout (2026).

Early transparency and community engagement (rather than communication only after an application is filed) are therefore becoming forms of project risk management. Relevant information includes the developer and intended operator, project scale and timetable, electricity and water needs, local impacts, infrastructure costs and expected benefits. A Carbon Direct review found that concerns about process and transparency were the most consistent factors across seven detailed cases. It also identified 46 US AI data centre projects, representing more than USD 170 billion in announced investment, that were blocked, withdrawn or stalled following community opposition between January 2024 and mid-May 2026. This figure represents announced investment associated with affected projects, not capital necessarily spent or lost.

Moratoria can delay credible projects, increase uncertainty and shift development elsewhere without resolving the underlying constraints. These risks can be reduced by establishing predictable siting, permitting and connection rules, clarifying responsibility for infrastructure costs and local impacts, and providing early disclosure and community engagement before projects advance.

IEA. CC BY 4.0.

Queue management and connection arrangements

Alongside permitting, connecting data centres in constrained grids requires decisions on queue management (which projects enter and how they progress) and connection arrangements (the type of grid access they receive). As connection requests grow, countries are modernising queue processes around three objectives: ensuring that projects are credible, allocating scarce grid capacity efficiently and streamlining connection studies and offers. At the same time, alternative connection arrangements, including non-firm and phased connections, are emerging as ways to provide earlier access where full firm capacity is not immediately available. This section describes emerging practices and identifies key considerations for policy design. Chapter 4 considers the allocation of costs associated with data centre grid integration, while Chapter 5 addresses bring-your-own-supply arrangements.

Connection queue management frameworks

Growing connection queues show that grid capacity is a scarce resource in several countries. The IEA estimates that in 2025, at least 1 700 GW of renewable generation and more than 600 GW of battery storage at advanced stages of development[2] were waiting in connection queues globally. As of May 2026, at least 180 GW of data centre projects at advanced stages of development were also in connection queues globally. A developer seeking to connect a data centre in a constrained grid may therefore face a lead time of several years, particularly if the developer is seeking a fully firm connection.

Large connection queues across several countries show that traditional queue management practices are under pressure. Legacy frameworks, often based on first-come, first-served principles, were not designed for very large, fast-moving and sometimes speculative projects seeking capacity in constrained grids. Data centres are therefore being considered within wider queue reforms that aim to determine which projects connect, when they connect and on what basis. Emerging practices do not point to a single best-practice model. Instead, countries are using different measures to address different policy objectives. The table below presents an overview of these objectives, the associated approaches used to pursue them and key design considerations. It also includes examples of approaches focused on generation to illustrate where a principle could potentially also be applied to loads.

[2] Projects at advanced stages include those at late stages in their connection requests and those under review. They exclude projects at early stages of the process as well as those that are unlikely to materialise.

IEA. CC BY 4.0.

Policy objectives and associated approaches to connection management for load and supply

Policy objective	Examples of approaches	Key design considerations
Ensuring a credible queue	Entry screening criteria: Denmark, France Maturity reviews: Norway, Great Britain (proposed) Financial commitments and withdrawal penalties: Spain, FERC Order 2023 (for generation) Legacy queue clean-up: Sweden Use-it-or-lose-it capacity rules: France, Spain (proposed), Japan (proposed)	Avoiding unnecessary barriers for credible projects
Prioritising projects effectively	Criteria-based prioritisation: Great Britain (readiness and strategic need for generation and storage), Germany (maturity and system benefit), EU Guidance Competitive allocation of demand-side capacity: Singapore, Spain	Keeping criteria transparent, predictable and based on clearly defined objectives
Streamlining network study and connection offer processes	Cluster studies: Germany, ERCOT Dedicated large-load processes: SPP, France, China Study automation: MISO (for generation) Stricter time limits: Australia, United States	Balancing faster processing with assessment quality; flexibility in case of project dropouts and operator resource availability

Notes: ERCOT = Electric Reliability Council of Texas; EU = European Union; FERC = Federal Energy Regulatory Commission; MISO = Midcontinent Independent System Operator; SPP = Southwest Power Pool. While some examples of approaches are currently only applicable to generation or storage, they are included for illustrative purposes.

A first objective is to ensure that the queue reflects credible projects. This is particularly relevant for large loads because capacity requests can reserve scarce grid capacity before projects are completed. Several jurisdictions are using stronger queue entry requirements, maturity assessments, financial commitments and milestone-based capacity release mechanisms to prevent projects that do not progress from retaining reserved capacity. In Great Britain, the Office of Gas and Electricity Markets (Ofgem) is considering stronger queue entry and permanence criteria as part of its reform of demand connections, with a particular focus on data centres. Possible requirements include evidence of financial backing, planning permission and refundable deposits linked to milestones. Sweden and Norway combine a first-come, first-served approach with strict maturity requirements, with capacity reservations subject to cancellation if projects do not meet key milestones. Since 2026, demand projects in Spain with connection permits at voltage levels of at least 1 kV have been required to pay monthly capacity reservation fees, which lead to discounts on network tolls after activity begins. In the United States, the Federal Energy Regulatory Commission (FERC) called on regional grid operators in June 2026 to justify or reform their application processes for large loads, including through application fees and readiness requirements. Overall, a key design consideration is ensuring that measures intended to deter speculative projects do not create excessive barriers for viable projects, particularly those of smaller developers.

IEA. CC BY 4.0.

A second objective is to prioritise projects effectively within the queue. Moving beyond legacy first-come, first-served approaches is not straightforward, as prioritisation criteria can reflect project maturity, grid constraints, broader policy objectives and the legal framework. In practice, several jurisdictions combine technical and maturity criteria with wider strategic considerations to determine which projects advance. Great Britain has moved towards a "first ready and needed" approach, under which progress depends on readiness and, for generation and storage, alignment with the country's energy strategy. France is expected to implement a first-ready, first-served principle by the end of 2026. German transmission system operators introduced a primarily maturity-based process in 2026 to determine which load and storage projects advance to the network study phase. The Netherlands prioritises projects that relieve congestion or support safety, security and basic social needs. Singapore uses competitive calls to allocate scarce grid capacity, with its second call making at least 200 MW available for data centres based on criteria covering strategic value, economic contribution and sustainability. Overall, the key design consideration is to ensure that prioritisation criteria remain transparent, predictable and based on clearly defined objectives.

Whether used for queue entry, continued participation or prioritisation, maturity requirements involve a further trade-off between ensuring project credibility and maintaining early visibility of prospective demand. Requiring advanced planning or permitting before projects progress through the connection process can help establish their credibility, but may shift bottlenecks to those processes and reduce network operators' early visibility of prospective demand. Staged approaches can mitigate this risk by allowing early grid assessment to advance before all planning and permitting milestones are completed, while reserving firm capacity or priority for projects that subsequently meet defined maturity milestones. The European Commission's proposed Cloud and AI Development Act would complement this approach by creating a single point of contact for designated data centre acceleration zones, helping to facilitate co-ordination across planning, environmental permitting and grid connection processes.

A third objective is to streamline network studies and connection processes to shorten application timelines where feasible. As large-load requests grow, project-by-project assessments can become slower and less able to capture aggregate grid impacts. Cluster or cyclical study processes can assess multiple requests together, using common assumptions and identifying shared constraints or upgrade needs. Dedicated large-load processes can also shorten timelines. France has announced a dedicated fast-track connection pathway for five major data centre projects on state-identified sites, targeting connection in 2028-2029. In the United States, ERCOT's Batch Zero process will group qualified large-load projects into a single system-wide study. Southwest Power Pool's High Impact Large Load framework creates a dedicated fast-track study pathway for large

IEA. CC BY 4.0.

loads, with a route to interconnection agreements within 90 days. More broadly, FERC has required six US regional grid operators to justify or reform their rules for integrating large loads, including their study processes. German transmission system operators use cyclical processing and cluster studies for mature applications.

Study automation can complement these process reforms by standardising repetitive modelling, screening and reporting tasks while retaining detailed review for complex cases. The Midcontinent Independent System Operator, which operates the grid in the central United States, has developed an automated process for generation interconnection studies that can reduce early processing times from months to days, illustrating the potential benefits of automation. Overall, the key design consideration is to balance speed with assessment quality, provide flexibility for management if projects change or drop out after being assessed and ensure that system operators have sufficient resources.

Network measures to make better use of grid capacity

Connection reforms are not the only way to accommodate growing demand for grid access. Network operators can also make better use of existing infrastructure by recovering reserved but underused connection capacity and deploying technologies that optimise grid use. The Netherlands and France have recently introduced mechanisms that allow network operators to recover underused capacity from existing grid users and make it available for other connections. Grid-enhancing technologies and other network-side measures can further increase the capacity of existing networks. These include dynamic line rating, advanced power flow control, topology optimisation, storage used as a transmission asset, reconductoring and voltage uprating. IEA analysis in Electricity 2026 estimates that deployment of these technology upgrades could unlock enough grid hosting capacity globally to connect around 450 GW to 700 GW of projects currently at advanced stages of connection queues. The IEA report Modernising Grids in the Age of Electricity examines how digital technologies, including AI, can help make better use of existing grids.

Connection arrangements

Once technical assessments are complete, the connection offer defines the conditions under which a data centre can consume electricity from the grid after the project is completed and energised. Two key features of grid access are firmness and timing. Firmness determines whether power consumption can be limited under specified conditions, while timing determines whether consumption

IEA. CC BY 4.0.

capacity is available in full from the outset or increases gradually as the project develops or grid reinforcements are completed.

Data centre developers have traditionally preferred fully firm connections. In constrained grids, however, waiting for a fully firm connection at full project capacity may delay access for several years until the necessary network or supply upgrades are completed. Non-firm connection arrangements can provide earlier access by allowing consumption to be limited under predefined conditions and may offer additional benefits such as lower grid fees. Phased connections can align reserved capacity with the gradual ramp-up of data centre demand, as projects with multiple stages of development typically take years to reach full capacity. In Great Britain, for example, 81% of demand projects in the connection queue indicated that they would be open to signing an alternative connection agreement in exchange for faster grid access.

These approaches can help make better use of existing grid infrastructure and reduce the inefficient reservation of scarce capacity. The IEA estimates that non-firm connection agreements could enable around 750 GW to 900 GW of grid hosting capacity globally, highlighting the potential scale of this approach. In constrained power systems, non-firm arrangements may be the principal route through which data centres can connect in the near term. For instance, the European Union's Strategic Roadmap for Digitalisation and AI in Energy promotes flexible connection arrangements for data centres where these can facilitate grid access. The European Commission is working with national regulatory authorities on a framework for flexible connection arrangements, for which data centres may be well suited.

We highlight four types of emerging alternative connection arrangements, the first three of which are non-firm models. Under these arrangements, data centres may respond directly by adjusting their own consumption from the grid or by paying for others in the network to adjust their consumption or supply:

- **Conditional non-firm or curtailable access**: Capacity can be reduced when specified electricity system conditions arise, such as congestion or disturbances.
- **Time-dependent access**: Capacity is guaranteed for a defined share of hours or during specified periods, with restrictions during constrained periods.
- **Dynamic operating envelope or active management**: Consumption capacity limits change dynamically according to real-time or forecast network conditions.
- **Phased or ramped access**: Contracted capacity increases progressively as demand or grid capacity develops.

Although these arrangements are not yet widespread, several models are emerging. In the United States, the Southwest Power Pool's Conditional High Impact Large Load service was approved in June 2026, providing large loads with

IEA. CC BY 4.0.

conditional non-firm transmission access for up to seven years while they await firm service. Since 2025, the Netherlands has offered a time-dependent product that guarantees firm capacity for at least 85% of annual hours, enabling earlier grid access and lower grid charges. In July 2025, South Australia introduced dynamic grid use limits for commercial and industrial customers in exchange for lower grid charges. Spain approved a set of four types of flexible demand access permits in July 2026, covering fixed periods, disconnection during contingencies and active management, to be introduced gradually. In July 2025, France's regulator approved a procedure allowing transmission-connected loads to request a gradual increase in connection capacity over ten years, with intermediate capacity milestones and provisions to reduce connection capacity if it is underused.

Non-firm and phased connection arrangements can help unlock grid access, but their design raises important trade-offs, particularly for data centres. Greater standardisation can improve transparency, comparability and non-discriminatory access and reduce transaction costs, while tailored agreements may better reflect local network constraints and the ramp-up profile of individual facilities. Policy makers could therefore establish a limited set of standardised options for their jurisdiction, with scope to tailor the terms where necessary to meet specific project requirements.

In either case, the conditions of grid access need to be clear. These include when limitations may apply, which users are curtailed first, the expected frequency, duration and notice periods, whether limitations are capped, whether there is a pathway to firm access and which monitoring and communication systems are needed. Non-firm arrangements may also require sufficiently detailed and timely information on network constraints and available capacity, particularly where access limits vary dynamically. For data centres, phased arrangements can help align reserved capacity with demand growth, reducing the risk that large capacity requests block other users before projects are fully built out. Without this clarity, these arrangements risk creating uncertainty for investors, concerns about operators meeting their uptime commitments to customers, unequal treatment of users and inefficient reservation of scarce grid capacity.

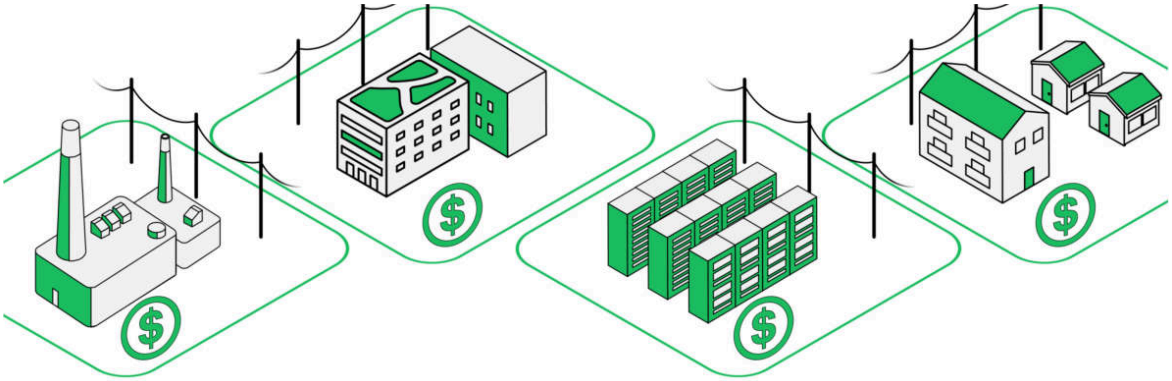
IEA. CC BY 4.0.

Emerging examples of connection arrangements that apply to large loads

Arrangement type	Examples	Key design considerations
Conditional non-firm or curtailable access	**Norway**: Voluntary "connection under special conditions" involves uncompensated curtailment during power system disturbances **Denmark**: Permanent scheme offers faster connection and reduced tariffs in exchange for possible interruptions **Southwest Power Pool (United States)**: Conditional High Impact Large Load service offers grid access for large loads, subject to curtailment under system stress and offered for up to seven years	Clearly defining curtailment triggers, priority, notice, frequency and duration, and any pathway to firm access
Time-dependent access	**Netherlands**: Voluntary time-dependent transmission rights for storage and load guarantee contracted capacity for at least 85% of annual hours and offer lower tariffs; restrictions require day-ahead notice	Ensuring that restricted periods are sufficiently predictable and commercially workable for large loads
Dynamic operating envelope or active management	**South Australia**: Commercial and industrial consumers keep grid use within dynamic limits in exchange for lower charges **Spain**: Actively managed transmission and distribution access models will be introduced progressively from September 2026	Requiring adequate automation and clear rules for compliance
Phased/ramped access	**France**: may increase according to predefined milestones and capacity levels **Great Britain:** Phased agreements are being explored as part of demand connections reform	Linking capacity increases to credible milestones, avoiding over-reservation and clarifying whether later capacity is guaranteed

IEA. CC BY 4.0.

4. Allocating costs related to grid and infrastructure upgrades for data centres



Connecting a large data centre to the grid can require investment in dedicated connection assets, wider network reinforcement and, in some systems, additional generation. Whether and when this affects the charges paid by other consumers depends on the spare capacity at the connection point, the size and operating profile of the facility and the flexibility it can offer when the network is constrained. The effect also depends on how the resulting costs are allocated between the connecting load and other network users.

How these costs are allocated is a matter for policy makers. The allocation can affect electricity affordability by changing network charges and, ultimately, the electricity bills faced by different customer groups. However, the effects are highly context-specific, and the pace and location of growth, prevailing system conditions, market design and regulatory frameworks all shape the outcome.

An important distinction underpins this analysis. Network costs are the costs incurred in developing, operating and maintaining the network. Network charges are the regulated means through which these costs are recovered from the network users, determining who pays, how much and when. The two are closely related but can move independently. A change in the network charges a consumer pays may reflect a change in the underlying network costs, a reallocation of those costs among consumer classes or between consumers and generators, or a transfer to public budgets financed through taxation, borrowing or other earmarked public revenues. Unit charges can also change when the volume of consumption over which fixed network costs are recovered grows or shrinks, without any change in the total costs.

IEA. CC BY 4.0.

Network cost allocation

The instruments used to allocate network costs are long established. What has changed is the scale, speed and novel locations (i.e. outside of traditional demand centres) of the loads to which they are now being applied. A single data centre facility can request a connection comparable to a substantial share of the peak demand in the area where it proposes to locate. Data centres can also be built in one to two years, much faster than the network lead times needed to serve them, which may be several years or longer. As such, network investments may need to be committed against anticipated rather than demonstrated demand. Meanwhile, network charging frameworks have typically been designed for demand that grows in smaller increments and at predictable locations. Data centre demand does not always follow these patterns and may therefore require new framework approaches.

A large load can trigger system costs at several levels. First, connection assets that serve only the new load, such as a dedicated line and substation linking the site to the network, can usually be assigned fully to the load that caused them. Second, reinforcement of the wider network serves both the new load and other network users, so the appropriate cost allocation depends on whether the reinforcement would have been needed without the new load, which can only be assessed against a current network plan. Third, capacity and adequacy costs depend on how far the load coincides with system peak rather than on its size. Finally, system and balancing services form a separate cost category in which any flexible loads may reduce costs rather than add to them.

Connection charging generally follows one of two broad approaches. Under shallow connection charging, the connecting customer pays for assets that serve it alone, while the cost of reinforcing the wider network is recovered from all users of the network. Under deep connection charging, the customer also pays some or all of the wider reinforcement costs triggered by its connection. Both approaches are used in the jurisdictions examined in this report, but the boundary between shallow and deep charging is drawn differently in each, so the share of the costs available for allocation through a large-load tariff differs among jurisdictions.

The allocation of wider network reinforcement costs is usually the most disputed because the costs and benefits do not necessarily accrue to the same consumers. If there is no deliberate change in tariff design, part of the cost of the grid infrastructure built to serve a single data centre may be recovered from households and other existing users, even though the benefits from digital infrastructure and investment for tax revenue and clean energy procurement may accrue more widely. As a consequence, cost allocation in such instances may become a disputed issue that jurisdictions without a strategic network plan cannot resolve other than through negotiation.

IEA. CC BY 4.0.

Two structural features shape these disputes. The first is common to all electricity networks. Electricity systems are built to meet peak demand and are highly capital-intensive, with capital recovery accounting for a large share of total system costs and remaining largely unaffected by how much electricity flows during any given period. Connection charges linked to the immediate cost of connecting to the network recover only part of the investment needed to build and maintain it. Congestion revenues signal where network capacity is scarce, but they generally decline when grid expansion eases the constraint. They therefore cannot fully fund a network designed to provide sufficient capacity.

Most network costs are recovered through regulated charges based on indicators such as energy use, peak demand and contracted capacity. These indicators cannot attribute all shared network costs precisely to individual users. This reflects the high fixed and shared costs of electricity networks, while tariff design determines how efficiently and fairly those costs are distributed. Large loads do not change the underlying economics, but their scale and concentration make the choices and the amounts involved more consequential.

The second feature is particularly relevant to data centres, although it is not unique to them. Data centres typically operate at high load factors, which can support efficient use of generation and network assets. However, this does not automatically translate into lower costs for other consumers, since a flat load may still add to system peak demand. A data centre's use of contracted grid capacity may also remain low for years after connection as computing equipment is installed progressively, causing demand to ramp up slowly. Uncertainty about future requirements, combined with competition for scarce capacity, may encourage operators to contract for more capacity than they need in the near term. Networks may consequently be built and remunerated before the reserved capacity is fully used, with the difference potentially falling on other users through charges until utilisation catches up. This effect is greater when charges are primarily volumetric, because loads with low consumption pay less despite occupying already financed capacity. In congested areas, reserved but unused capacity may also deny access by other applicants, creating a cost that is not directly reflected in tariffs.

Not all cost effects reach consumers through the network charges. Additional demand can raise energy and capacity costs before any network investment is made, as more expensive plants are dispatched more frequently and systems procure additional firm capacity to maintain adequacy. These effects reach consumers through wholesale electricity prices and, where capacity mechanisms operate, through capacity payments. They are distinct from the network costs allocated through tariff design but ultimately affect the same electricity bills.

IEA. CC BY 4.0.

Costs for consumers are more pronounced when demand is falling, as the fixed costs of the existing network must be recovered from a shrinking customer base. Charges per remaining consumer may therefore rise even without new investment. Whether additional load spreads fixed costs across a broader base depends on where it connects: spare capacity nationally does not imply spare capacity at a constrained node. The outcome therefore depends on local system conditions and the load's location.

Fiscal measures can operate alongside network charges and influence who ultimately pays for grid investments. Tax exemptions can offset the costs recovered through connection charges, while levies can increase the overall amount paid by a network user beyond the charges included in the tariff. At least 37 US states offer tax incentives targeting data centres, while several jurisdictions have introduced charges that accrue to public budgets rather than to network revenue. On 1 July 2026, Finland moved data centre electricity from its lower electricity tax category to the general category, raising the rate from EUR 0.05 cents to EUR 2.24 cents per kilowatt-hour. In Canada, Alberta plans to introduce a levy on computing hardware at grid-connected sites of 75 MW and above from the end of 2026. The levy will be creditable against provincial corporate income tax and set lower for facilities that supply their own power. Transfers run the other way as well, as in Germany, where a federal subsidy of EUR 6.5 billion financed from the Climate and Transformation Fund cut average transmission network charges by 57% for all consumers in 2026. These measures do not change the underlying network costs, but they change who bears them.

Allocating risks

Cost allocation is ultimately a question of risk sharing. Networks are planned and paid for on the basis of forecasts of load that may not be realised. The main risks concern whether a connecting load will use the capacity it reserves, remain connected for the life of the assets built for it, meet its obligations, require service at all times or even arrive at all. There are different instruments that can be introduced to mitigate these risks. Demand and utilisation risk arises when a load uses less capacity than it reserved. Stranded-asset risk arises when a project shrinks or leaves before the assets built for it have been paid off. Counterparty and default risk concerns whether the load can meet its financial obligations. Reliability risk concerns whether the load must be served at all times, including when the system is constrained. Speculative-request risk arises when a load never arrives after investment has been committed. This risk is generally addressed upstream of the tariff, through financial commitments and milestone deposits at the connection stage.

IEA. CC BY 4.0.

There are different risk mitigation measures matching each risk category. These instruments are increasingly applied together, making it important to consider their cumulative effect.

Risks associated with connecting large loads and instruments to mitigate them

Arrangement type	Examples
Demand and utilisation	Minimum demand or take-or-pay charges, under which the load pays for a high share of contracted capacity regardless of use; capacity-based rather than volumetric charging; phased connection agreements
Stranded assets	Contract terms matched to the depreciation life of the relevant assets; exit or stranded-cost charges; depreciation policy
Counterparty and default	Collateral, creditworthiness requirements and parent company guarantees
Reliability	Choice between firm and non-firm service; flexibility-linked tariffs offering lower charges or faster connection in return for curtailable load
Speculative requests	Handling during the connection request phase, i.e. upstream of the tariff decision, through financial commitments and milestone deposits at the connection stage, queue management and use-it-or-lose-it provisions

In the United States, the Virginia rate class for large loads, approved in November 2025 and taking effect in 2027, combines minimum charges based on 85% of contracted transmission and distribution capacity and 60% of contracted generation capacity with a 14-year contract term, exit fees and collateral of around USD 1.5 million per MW. Ohio pairs a minimum payment for contracted capacity with a long minimum contract term to discourage data centre facilities from reserving large amounts of grid capacity and leaving it underutilised. Oregon pairs a dedicated rate class with a long-term supply commitment. On the connection side rather than the tariff side, Great Britain is consulting on commitment fees alongside queue milestones and phased agreements, and Ireland combines conditional access with the possible loss of firm status where associated generation underperforms. Combining several risk mitigation instruments is therefore common across jurisdictions.

Each element addresses a distinct risk, but the underlying exposures may overlap. Instruments calibrated separately may recover more than the costs placed at risk. For example, if a load is already paying for capacity it is not using, this may be sufficient to fund the asset, meaning a minimum charge might cover much of what an exit fee is intended to cover. Collateral may price the same shortfall an additional time. The appropriate test is whether the policy package as a whole is calibrated against the costs that would actually be stranded, rather than each element being set on its own terms. Just as existing consumers should

IEA. CC BY 4.0.

not bear costs they did not cause, a large load should not pay substantially more than the costs and risks attributable to it.

Flexibility differs from the other risk mitigation instruments mentioned above because, where it is dependable and operationally useful, it can mitigate several risks simultaneously. A load that can be curtailed when instructed may reduce or defer the reinforcement otherwise required to provide firm capacity. This can lower reliability exposure, reduce the risk of assets becoming underutilised if projected demand does not materialise and limit the amount of firm capacity reserved at the point of connection. These benefits, however, depend on flexibility being available at the required location and time for a sufficient duration, and performing reliably when called upon. This supports recognising verified flexibility through standardised tariff or connection arrangements where it creates demonstrable avoided system costs, rather than through concessions negotiated case by case.

Policy trade-offs in cost allocation

The trade-offs between different cost allocation options require further attention based on policy goals. Cost reflectivity is one of several principles that regulators must consider in tariff design. European regulators note that fully cost-reflective tariffs can be complex and produce volatile electricity prices, which may conflict with the simplicity, stability and predictability required by consumers and long-term investors. Risk allocation involves similar judgement, since socialising demand or stranded-asset risk may be simpler but can create moral hazard, whereas assigning the risk fully can stall investment. The relevant responsibilities may also sit at different levels of governance. In the United States, for example, wholesale charges fall within the responsibilities of system operators and federal regulators, while retail rate design generally rests with state commissions. Coherent allocation therefore requires co-ordination and transparency. This depends on information about connection costs and the scale and flexibility of the load, which is held by utilities and developers and is not disclosed by default.

What makes these trade-offs a policy choice, rather than a technical calculation, is that each sets legitimate, competing goals against each other so that they cannot be objectively ranked. Weighing cost reflectivity against stability, consumer protection against investment and national competitiveness against local affordability requires distributional judgement about who should bear the costs and risks. Regulators and system operators can frame the options and quantify their effects, but the choice of where to strike each balance ultimately reflects national priorities and must be determined by policy makers.

IEA. CC BY 4.0.

Instruments in use across jurisdictions

Countries facing rapid data centre growth start from different positions in terms of grid readiness and relevant market and regulatory frameworks. A common objective is to require large loads to make credible financial commitments, contribute to the costs they cause and, where possible, secure their own clean electricity supply while protecting existing consumers. Jurisdictions differ in the instruments they use and in their objectives, which may include protecting affordability, accelerating connections, attracting investment or steering siting and the fuel mix.

Changes have been most rapid in the United States. As of August 2026, 24 states had approved or were considering at least one large-load tariff, with proposals pending in a further six states. At the federal level, the Federal Energy Regulatory Commission issued orders to the six regional grid operators in June 2026, directing them to justify or reform their rules governing large load connections.

In the European Union, the European Commission issued a proposal in July 2026 to replace Article 18 of Regulation (EU) 2019/943 with a new set of provisions on network charges. The proposal seeks to redesign EU network charging rules so that system operators and grid users face stronger incentives to reduce overall system costs. These include incentives for operators to improve cost-effectiveness and timely access, signals for users to shift consumption, greater transparency and mandatory performance indicators set by regulators to encourage efficient grid use. The proposal also suggests that regulators create special cost-reflective tariff regimes for user categories with distinct consumption profiles, such as energy-intensive industries and data centres, paired with safeguards to protect the affordability of households and small and medium-sized enterprises.

IEA. CC BY 4.0.

Instruments used to allocate data centre costs by jurisdiction

	Large-load tariff or minimum take	Financial security or deposits	Cost-reflective connection charges	Flexible or non-firm connection	Locational or time signals	Bring-your-own capacity requirements
United States - federal (FERC)	●	●	●	●	●	●
United States - states	●	●	●	●	●	●
Great Britain		●	●	●	●	
Ireland			●	●	●	●
Netherlands			●	●	●	
Nordic countries (Finland, Sweden and Norway)			●			
Singapore						●
Malaysia	●		●			
Australia		●		●		
Japan				●	●	

Key: ● established or primary instrument ● emerging or partial instrument
Notes: FERC = Federal Energy Regulatory Commission. Blanks indicate that the instrument is not a primary feature in the jurisdiction.
Source: See the policy collection in the annex.

Comparing tariff terms

Separate tariff classes for large loads have spread rapidly within the United States but remain uncommon elsewhere. Other jurisdictions have responded mainly through the connection process instead, and cost-reflective connection charging and flexible or non-firm arrangements are each recorded in seven of the ten jurisdictions covered. Globally, no common definition of a large load has emerged yet. Thresholds range from a minimum of 1 MVA in Ireland, with the full regime applying at 10 MVA, to 25 MW in Ohio and Virginia and 75 MW in Texas. Virginia also adds a load factor criterion. Comparisons across jurisdictions therefore can be misleading if they combine different consumer groups.

Two broad groups of instruments are visible. In the United States, commitments are generally quantified within tariffs through minimum-take percentages, contract terms and collateral. Elsewhere commitments are more often attached to grid access, such as through clean-supply conditions in Ireland, a charge on granted

IEA. CC BY 4.0.

but unused capacity in Spain and readiness gates in Great Britain. These approaches address overlapping cost and risk allocation objectives, but jurisdictions use different instruments to pursue them.

Where instruments have been introduced, they have generally accumulated rather than replaced one another, and the material reviewed does not show how the individual elements are calibrated in relation to one another. Fiscal measures have also developed separately from network charging and sit with finance ministries rather than with energy regulators.

Who pays and who gains when a data centre connects is a distributional choice for policy makers. Granting access can require large, long-lived investments in generation capacity and connection and network infrastructure. Deciding who bears the cost does not change its total size, but it determines how much falls on the connecting load, other consumers, taxpayers or future consumers.

Tariff design is where much of this choice is made. The depth of connection charging sets how much of the reinforcement cost is in question. Demand- and capacity-based charges set how much a high-load-factor customer contributes to fixed costs. Locational and time signals determine whether siting is influenced at all. Contract terms, minimum-take obligations, collateral and exit charges determine who holds the risk that a load arrives late, arrives smaller than requested or does not arrive at all.

The measures examined are too recent for their outcomes to be assessed. What can be shared now is the reasoning behind them, covering which costs are at stake, which approaches exist for recovering them and what each choice implies for the allocation of costs and risks. Transparency and predictability in how this reasoning is applied matter as much as the instrument selected, both for investors deciding whether to build and for consumers asked to accept the result.

IEA. CC BY 4.0.

Comparison of leading large-load tariff regimes

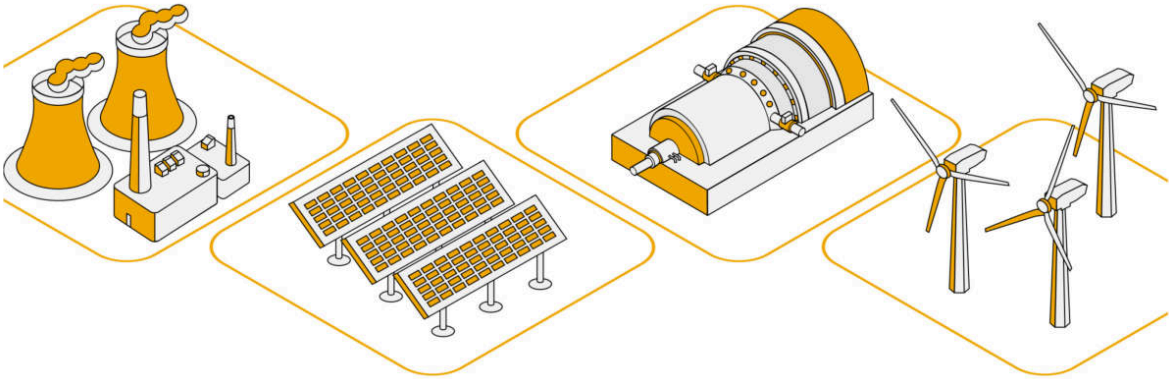
Jurisdiction	Scope	Minimum take	Contract term	Stranded-cost protection	Collateral or deposit	Flexibility or self-supply requirement
AEP Ohio (United States)	Data centres >25 MW	≥85% of contracted capacity	~12 years (4-year ramp + 8-year minimum)	Exit fee of 3 years of minimum charges	Load study fee ≤ USD 100 000	None identified
Virginia GS-5 (United States)	≥25 MW and load factor ≥ 75%, from 1 January 2027	85% of transmission and distribution; 60% of generation	14 years	Exit fees; 3 years' notice to reduce contracted demand	~USD 1.5 million per MW (cash or letter of credit)	None identified
Texas Senate Bill 6 (United States)	>75 MW*	Interconnection cost share	Per interconnection agreement	Hold-harmless requirement: stranded costs kept out of rates	USD 100 000 transmission screening fee	Load must be curtailable; disclosure of backup generation
Ireland (CRU)	Data centres ≥1 MVA; tiers below and above > 10 MVA	None	Per connection agreement	Import capacity reduced or moved to non-firm on non-compliance	None identified	Dispatchable capacity matched to import capacity; 80% of demand matched by renewables
Great Britain (Ofgem)	Transmission demand queue	None	Per connection agreement	Milestone and readiness requirements	Refundable deposits and commitment fees	None identified
Spain	Demand access permits ≥1 kV	None	Per permit	Capacity reservation charge on granted access capacity	Monthly capacity reservation charge on granted capacity	Flexible access permits at nodes without firm capacity
Malaysia	Ultra-high-voltage class	85% of declared maximum demand	Per supply agreement	Reinforcement cost borne by the data centre ("user pays")	None identified	None identified

Notes: AEP = American Electric Power; CRU = Commission for Regulation of Utilities; Ofgem = Office of Gas and Electricity Markets. *Texas curtailment and remote-disconnect requirements apply to loads above 75 MW.

Sources: IEA analysis based on Public Utilities Commission of Ohio (2025), Virginia State Corporation Commission (2025, 2026), Texas Senate Bill 6 (2025), CRU (2025) and Ofgem (2025, 2026).

IEA. CC BY 4.0.

5. Securing electricity supply for data centres



The challenges of powering data centres quickly, securely and sustainably are driving diversification and innovation in corporate energy procurement strategies, policies and regulation. Ensuring that data centres have reliable access to grid power and can maintain continuous operations is a key concern for data centre operators. In addition, technology companies seek to align their electricity consumption with their sustainability goals. They continue to enter into power purchase agreements (PPAs) for clean energy to supply data centres, with the long-standing use of renewable energy PPAs being combined with growing interest in nuclear energy. While most data centres continue to be powered by the grid, data centre operators are increasingly deploying on-site power generation from natural gas where grid connections of the required capacity cannot be provided within the required timescales, especially in the United States. Data centres are also procuring energy storage, which could present new opportunities for how they interact with the grid. Policies and regulations are addressing “bring-your-own” generation models and clean energy sourcing requirements.

Strategies

Corporate PPAs continue to be a prominent tool for data centre power procurement. A PPA is generally a contract between a customer and a generator that allows the customer to secure a fixed, long-term power price and establish a link with a specific generation source. They are often used to support corporate clean energy and sustainability targets. While a PPA provides a contractual connection between a customer and a generator, the customer is almost always physically connected to the wider electricity grid, which supplies electricity from a mix of generators. As discussed further below, data centres have also recently expressed significant interest in off-grid arrangements, in which a data centre and a power generator may form both contractual and physical links while remaining isolated from the wider electricity grid.

IEA. CC BY 4.0.

Data centre demand has driven significant volumes of PPAs for both renewable and nuclear energy. In 2025, data centres signed agreements for almost 45 TWh of electricity from renewables, roughly equivalent to the total electricity consumption of Nigeria that year. Data centres are also looking to procure significant volumes of nuclear energy from both conventional reactors and small modular reactors that are currently under development. Several technology companies have entered into gigawatt-scale agreements for power from existing nuclear power plants. As of January 2026, data centre operators had agreed expressions of interest (memoranda of understanding and letters of intent) covering about 45 GW of small modular reactor capacity, signalling the demand momentum behind these technologies. Commercial operation of these types of reactors is not expected to begin until the 2030s.

Data centre PPAs for clean energy are particularly prevalent in the United States. About 65% of renewable PPAs signed in the United States over the past five years were connected specifically with data centres. In the first quarter of 2026, corporate clean energy contracts in the United States reached their highest quarterly level in at least five years, driven in large part by data centre demand. For example, in February 2026, Google and Xcel Energy entered into an agreement for 1.9 GW of clean energy, comprising wind, solar and energy storage.

Power sellers and buyers can generally shape PPA terms to reflect their needs, which can in turn affect the wider electricity system. Several technology companies have pursued targets to match their electricity consumption with clean energy sources on an hourly – rather than annual – basis to minimise reliance on emitting resources. Earlier IEA analysis estimated that around 80% hourly matching with low-emissions sources could be achieved in the United States at costs comparable with those of projects matching annual electricity consumption volumes. Hourly matching with clean energy sources can help drive significant system-level emissions reductions at a competitive cost, but its value to the electricity system depends on the grid conditions. If applied narrowly, it can introduce additional system-level challenges. For example, it can lead generators to respond to the load profile of an individual customer rather than real-time system-wide conditions. System policies and PPAs should therefore both be tailored to support positive system-level outcomes.

Demand for rapid power provision is also reshaping how data centres relate to the grid. As highlighted above, connecting a large load to the electrical grid may take several years. This is driving growing interest in on-site power models, which can take several forms, including but not limited to the following:

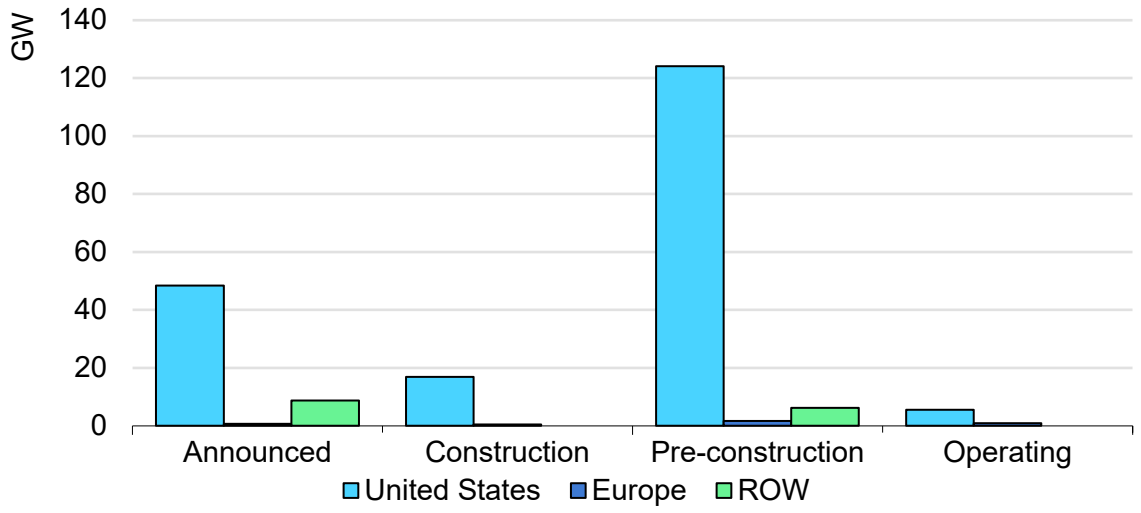
- One arrangement is purely **off-grid** or “islanded”, with on-site generation and storage meeting all of a data centre’s demand and no connection to the wider electrical grid. Such projects tend to be exceptionally rare.

IEA. CC BY 4.0.

- A second arrangement is a **mixed** model, in which the data centre is powered by significant on-site generation, while a grid connection supplements this generation and provides additional reliability.
- A third arrangement is a "**bridge**" model, in which the data centre begins operating off-grid and later secures a grid connection.

While data centre operators continue to prefer reliable grid connections, project announcements and development pipelines indicate significant interest in on-site power arrangements where grid connections cannot be secured in a timely manner. Natural gas-based on-site generation is increasingly the technology of choice. The global pipeline of natural gas generation intended for on-site use by data centres is around 210 GW. Of this, around 190 GW is at the proposal stage, while 17 GW is under construction. Projects in the United States represent over 95% of this capacity. A notable example is the June 2026 announcement by Chevron and Microsoft of a 20-year PPA for a 2.67 GW natural gas project to provide on-site power for a data centre in Texas. However, there are still relatively few examples of completed data centres with significant on-site generation. There is also considerable uncertainty about project timelines because of backlogs in the supply of gas turbines and uncertainty over whether all announced projects will proceed.

Planned on-site natural gas generation capacity for data centres, 2026



IEA. CC BY 4.0.

Note: ROW = Rest of world.
Sources: IEA (2026), Key Questions on Energy and AI; IEA analysis of Global Energy Monitor data, August 2026.

On-site generation raises several considerations and potential challenges. To meet the high reliability requirements of data centres, generation capacity in off-grid arrangements may need to be significantly overbuilt relative to the capacity of the data centre. For example, powering a continuous 1 GW data centre load at

IEA. CC BY 4.0.

the necessary reliability level through natural gas turbines would require about 1.3 GW to 1.7 GW or more of capacity. Additionally, off-grid or bridge arrangements may not be covered by comprehensive planning or regulation that other grid resources often go through. This could lead to suboptimal outcomes if on-site resources later connect to the wider electrical grid. While on-site generation is often put forth as an approach to reduce cost impacts on other electricity ratepayers, increased demand for fuel and power equipment can still cause indirect impacts. By contrast, grid-connected flexible operations can facilitate faster grid connections, reducing the need for on-site generation.

Energy storage is also increasingly being procured by data centres. Its roles include smoothing the short-term power fluctuations that AI data centres can produce and providing longer-term reliability. Energy storage can also help fulfil voltage ride-through requirements that are being imposed on data centres in some regions. Data centre procurement is driving the productization of large-scale battery storage designed specifically for IT loads. Data centres are also procuring long-duration energy storage. For example, Google's February 2026 agreement with Xcel Energy includes 300 MW of energy storage with a discharge duration of 100 hours. In Australia, Amazon entered into a deal to control the operation of 50 MW of battery capacity as part of its broader commitment to match its energy use with carbon-free power in the country.

Policy and regulation

Policy and regulation are both reflecting and shaping data centre energy procurement strategies. Several countries have introduced clean energy requirements for data centre energy consumption. For example, Ireland's policy requires data centres to meet at least 80% of their annual demand from additional renewable energy generated within the country. A six-year "glide path" towards compliance begins when the data centre is energised.

Summary highlights of data centre clean energy sourcing requirements

Country	Clean energy sourcing requirements for data centres
China	• 80% green electricity target for new data centres in national computing hubs by 2025-2026
Germany	• 50% renewable electricity from 2024 • 100% renewable electricity from 2030
Ireland	• At least 80% of annual demand from additional domestic renewables (certain new data centres)
Singapore	• At least 50% of proposed new capacity must be powered by eligible green energy pathways

Source and note: See the annex for additional examples and details.

IEA. CC BY 4.0.

The concept of “bring your own” – under which data centres are incentivised or required to procure additional electricity generation or capacity to help meet their consumption – is also appearing in policy. A prominent example is Ireland, where certain data centres connecting to the grid are required to provide dispatchable generation to cover their demand. This generation must be located on-site or nearby and must participate in the wholesale electricity market.

In the United States, the “bring your own” concept is reflected in different ways. At the federal level, FERC has required regional grid operators, in a series of orders issued in June 2026, to address co-location arrangements and behind-the-meter generation (i.e. projects not connected to the grid), among other issues. At the state level, where energy generation and retail sales are generally regulated, the clean transition tariff is a notable example. This model generally enables a large-load customer to work with a utility and regulator to procure additional clean energy through a specific tariff. Google and NV Energy, a regulated utility, demonstrated the clean transition tariff model in Nevada in 2025 by procuring new geothermal capacity. The model has the potential to be replicated in other jurisdictions. For example, Google’s agreement with Xcel Energy, mentioned above, is based on the clean transition tariff structure.

Data centre energy procurement policies can be designed to support system-level objectives. Ireland’s requirement that new data centre generation participate in the wholesale electricity market can help ensure that the additional generation supports wider system quality and reliability. The clean transition tariff model can provide system-level benefits such as improved resource adequacy and improved planning certainty by including the new resources in the state-utility regulatory regime. The examples also illustrate how a single policy or regulatory tool can cover both clean energy and “bring your own” procurement.

IEA. CC BY 4.0.

6. Integrating data centres securely into the grid



An important integration challenge for data centres and other large loads is how they behave as major electricity users once they become operational. If these consumers can operate flexibly by reducing their power draw from the grid during periods of scarcity, they can ease their integration into the system, support additional demand without compromising reliability and, in some cases, provide grid support. At the same time, because individual data centres can be as large as major industrial facilities, their behaviour during both normal operation and periods of system stress needs to be predictable and manageable. This is necessary to ensure that they do not create or worsen disturbances, affect the quality of electricity supply for other users or create new electricity security risks. While flexibility and stability address different needs of the power system, they rely on a common solution architecture: advanced telemetry, workload classification and real-time control systems that enable operators to monitor, forecast and manage data centre electricity demand. These features allow for a flexible response to demand under conditions of normal system operation while also facilitating compliance with emerging technical requirements during system disruptions.

Flexibility

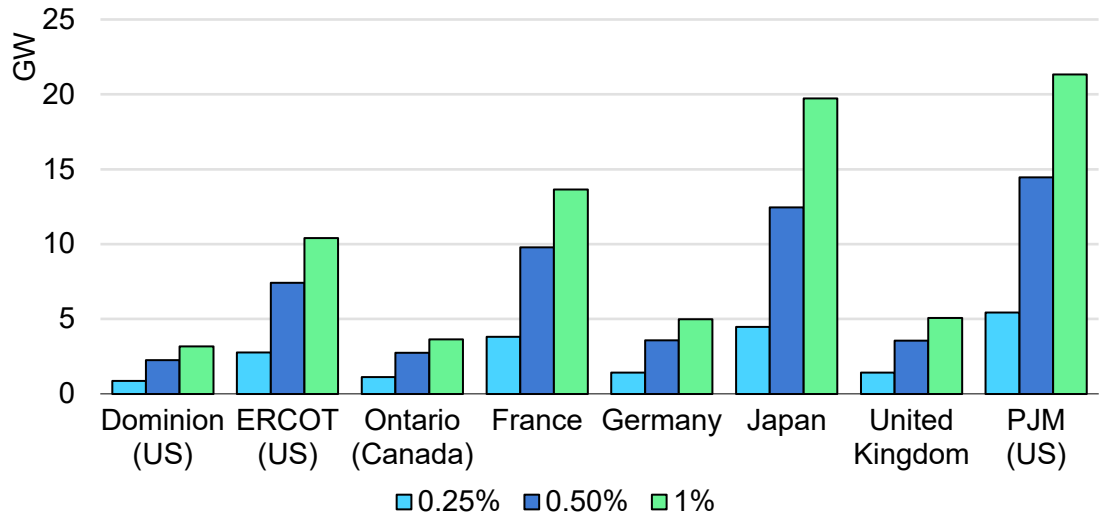
Even limited flexibility from data centres can have significant value for electricity systems. Grid and generation infrastructure are often planned around peak demand levels, which usually occur for only a small number of hours each year. If large loads can reduce or shift part of their demand during these periods, or temporarily meet it using their own power assets, they can ease local congestion, defer reinforcements and support faster connections for new demand without compromising reliability.

IEA. CC BY 4.0.

Data centre flexibility can take several forms. On-site assets such as batteries, uninterruptible power supply systems and backup generation can reduce reliance on the grid during periods of system stress. Non-critical auxiliary loads, especially parts of cooling systems, may be adjusted for short periods. Some IT workloads can also be shifted in time or geographically, reducing consumption in constrained locations during critical hours.

The figure below puts this potential into perspective by comparing how much additional load could be accommodated in selected systems today if data centres were able to reduce their grid demand during a small share of the highest-load hours. These estimates should be interpreted as indicative, as the actual headroom will vary according to local grid constraints, weather conditions, operational requirements and the extent to which recorded peaks reflect system limits.

Headroom in selected power systems for different flexibility cases



IEA. CC BY 4.0.

Notes: CAISO = California Independent System Operator; ERCOT = Electric Reliability Council of Texas. Headroom is calculated as the largest constant additional load that could be added without exceeding the historical peak demand, while keeping the amount of energy that would need to be curtailed, shifted or supplied on-site below the selected flexibility threshold. The calculation uses hourly demand data from 2015 to 2025, adjusted to current demand levels to remove structural growth while retaining weather-driven variability. Historical peak demand is defined as the average of the top 100 hourly observations of adjusted real load over a ten-year period.

Source: IEA analysis based on data from CAISO, PJM, ERCOT, the Alberta Electric System Operator, European Network of Transmission System Operators for Electricity, regional transmission system operators in Japan and Elexon.

The duration and scale of the required response also matter. For example, even small reductions in grid demand could present operational challenges if sustained over long periods. Nonetheless, the estimates show that the system value of flexibility is concentrated in a relatively small number of hours. For example, in France, the analysis suggests that around 6.5 GW of additional data centre load could be accommodated in the current system if the load were flexible for an

IEA. CC BY 4.0.

amount equivalent to 0.25% of its annual electricity use during critical hours. This is equivalent to about 7.5% of the country's current peak load. Extending flexibility beyond the very highest-load hours could increase the potential headroom.

A main barrier to greater flexibility provision today is that the value of high server utilisation, service reliability and speed to market exceeds the value of reducing electricity costs during a small number of hours. Service-level agreements can impose penalties for interruptions, while co-location providers may not control or have visibility over the workloads running inside their facilities. In many cases, the contractual and commercial terms governing flexibility are also not yet clear. As electricity costs typically represent a smaller share of total data centre costs, wholesale electricity price signals alone may not provide a sufficient incentive to adjust operations. Given the importance of speed to market, a faster or more certain grid connection could provide a stronger incentive for credible flexibility commitments than electricity price signals alone.

Emerging models show that flexibility can come both from data centres themselves and from resources around them. Google has signed agreements with utility partners covering 1 GW of data centre demand response, including arrangements to shift or reduce selected workloads during periods of grid stress. Separately, its partnership with Voltus in PJM would create a 100 MW virtual power plant by aggregating distributed resources such as batteries, smart thermostats and other flexible loads from homes and businesses to help support data centre growth. The Japanese Watt-Bit project provides another operational model, using the geographical shifting of computing workloads in response to grid conditions. Octopus and Masdar are deploying flexible connection agreements to accelerate data centre grid connections by integrating them into the portfolio of resources managed by the electricity supplier, thereby enabling the data centre to maintain uninterrupted operation while also helping to manage network congestion.

Regulatory frameworks are also beginning to focus on how flexibility commitments can be made credible once large loads are operational. This requires clear definitions of when flexibility can be called upon, how much load can be reduced, how quickly it can respond, how performance is measured and what happens if the agreed response is not delivered. In Great Britain, Ofgem is examining how large-demand users could contribute to consumer-led flexibility, highlighting the need for arrangements that allow large loads to provide measurable system support.

Policy makers therefore need to develop frameworks that make flexibility usable in real operations, with clear rules on eligibility, dispatch triggers, notice periods, duration limits, measurement, compensation and verification of compliance. The European Union Agency for the Cooperation of Energy Regulators is leading the

IEA. CC BY 4.0.

update of the Connection Network Codes to ensure that new and growing loads, such as data centres, are connected to the grid in a way that supports stable system operation. Once approved, these rules will apply across the European Union, promoting more harmonised approaches to connection requirements, flexibility, queue management and market participation while allowing national regulatory authorities to adopt stricter measures.

Stability

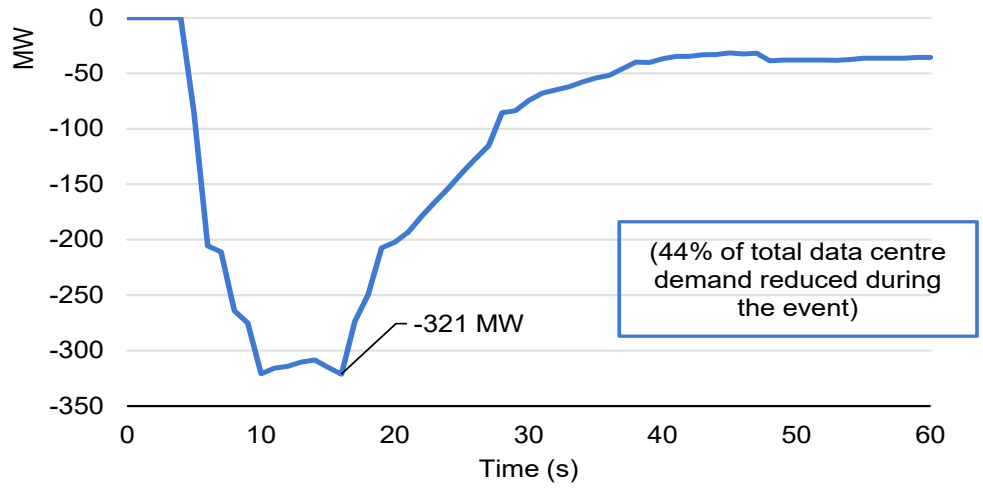
The behaviour of large loads can significantly influence system stability, making their secure integration an important electricity security concern. Their scale, concentration and rapid growth mean that data centres and other large loads can no longer always be treated as conventional passive loads. Several countries are therefore starting to develop stability-related measures traditionally used for electricity supply assets to support the secure integration of large loads into power systems. This section focuses on stability – the ability of power systems to absorb disturbances securely within milliseconds or seconds. It illustrates the key risks associated with large loads and highlights emerging practices to safeguard system stability.

Large electronic loads can create stability risks during grid disturbances if their behaviour is not aligned with system needs. During such events, data centres may disconnect from the grid or sharply reduce grid demand by transferring to backup supply within milliseconds in response to changes in voltage or frequency, for example following a transmission line fault. This occurs because internal systems are typically configured to protect the facility rather than support the grid, and widespread standards for more grid-friendly behaviour have not yet been implemented. The figure below illustrates data centres' collective response to a grid disturbance in Ireland. If several large facilities respond in the same way at the same time, the resulting sudden reduction in demand can worsen the disturbance, while poorly co-ordinated reconnection can create a second shock.

During normal operation, rapid changes in computing activity can cause large, oscillatory swings in power demand. Power conversion equipment linking data centres to the grid can create harmonics, potentially causing equipment to overheat or interfering with nearby grid controls. Overall, the behaviour of data centres may become increasingly important as they expand, cluster in the same locations and support more dynamic AI workloads. If not properly managed, their operation can affect the wider grid through both additional costs and security risks.

IEA. CC BY 4.0.

Data centre demand reduction in response to a transmission system voltage disturbance in Dublin, 26 January 2025



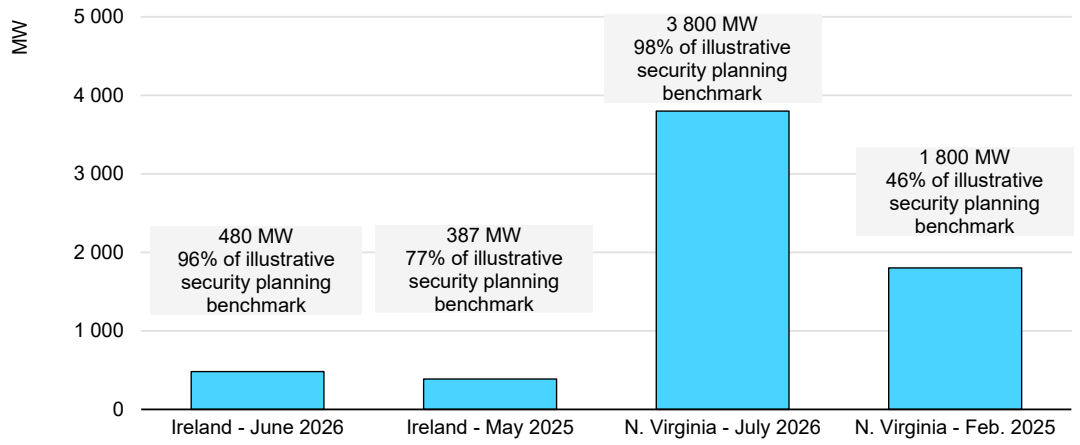
IEA. CC BY 4.0.

Source: IEA (2025), Powering Ireland's Energy Future.

Recent events show that stability risks are emerging. In the all-island Irish system, five transmission faults between 2022 and 2026 led to reductions of up to 52% in total data centre demand under conditions that generators, batteries and interconnectors would normally be required to ride through. The largest event, in June 2026, involved a 480 MW reduction in data centre demand. This was nearly equivalent to the loss of the system's largest cross-border interconnector, whose disconnection at 500 MW has traditionally been used as a reference in all-island contingency planning. In North America, the North American Electric Reliability Corporation (NERC) has identified more than 30 unplanned demand-loss events involving data centres or cryptocurrency mining facilities since 2023. In the largest such event in the history of PJM, a transmission line fault in Northern Virginia in July 2026 caused 3.8 GW of data centre demand to transfer to on-site backup supply, requiring immediate corrective action by the grid operator. For illustrative comparison, this was roughly equivalent in magnitude to the 3.875 GW Resource Loss Protection Criteria used for the Eastern Interconnection. The event led PJM to evaluate changes to reliability requirements for large loads, including ride-through standards.

IEA. CC BY 4.0.

Unplanned data centre-initiated demand loss during recent routine grid faults, 2025-2026



IEA. CC BY 4.0.

Notes: For Ireland, the all-island value of 500 MW (disconnection of the largest interconnector) was used as an illustrative benchmark. For Northern Virginia, NERC's Resource Loss Protection Criteria for the Eastern Interconnection was used.

Sources: IEA analysis based on EirGrid and SONI (2025), Large Demand Facility Fault Ride Through Issue and Proposed Solutions; EirGrid and SONI (2026), Fault Ride Through / Active Power Recovery / Rate of Change of Frequency; NERC (2025), 2025 Frequency Response Annual Analysis; NERC (2025), Comments of the North American Electric Reliability Corporation.

Although outages were avoided, these events show that routine grid faults can trigger load losses at data centres comparable in scale to the loss of major power plants or interconnectors. This creates security risks because losses of data centre demand could occur alongside the loss of other assets, bringing the grid close to or beyond secure operating limits. In 2026, NERC issued an alert requiring action on the modelling, monitoring and operation of computational loads.

Two approaches being used by grid operators to mitigate these stability risks are operational practices and the procurement of system services. Operational measures include limiting vulnerable demand, managing interconnector flows and co-ordinating reconnection. System services can entail procuring support from batteries or other resources capable of absorbing sudden demand losses. However, these measures come with additional costs and, as indicated in analysis by the system operators of the all-island system covering Ireland and Northern Ireland, may not be enough to ensure stability without technical performance requirements for large loads. This concern is increasingly relevant in Ireland, where peak data centre demand reached around 1 GW in June 2026. The grid operator EirGrid estimates that credible faults could cause data centre demand reductions of up to 900 MW, equal to the maximum instantaneous demand loss for which the all-island system is currently secured.

In this context, several jurisdictions are developing technical performance requirements for large loads, although these are at different stages of maturity. ERCOT has approved dedicated rules that took effect in August 2026, while

IEA. CC BY 4.0.

Denmark and France already apply ride-through and recovery requirements to transmission-connected demand. Ireland and Northern Ireland have developed a broader package covering technical performance, modelling and data provision, with the proposed requirements remaining subject to regulatory approval. Australia and Great Britain are also developing dedicated frameworks, while Japan is examining potential requirements through an international review of grid code approaches commissioned by the Organization for Cross-regional Coordination of Transmission Operators. The European Union Agency for the Cooperation of Energy Regulators has established an informal workstream on data centre requirements as part of the ongoing revision of the Demand Connection Code. The revised code could provide a common framework for harmonising technical connection requirements across EU member states. The table below summarises the key emerging requirements and their status.

Examples of emerging technical requirements for large loads to safeguard system stability

Area	Risk mitigation approaches	Recent developments
Behaviour during disturbances	Requiring large load facilities to remain connected during short disturbances and to restore demand in a controlled manner	**Texas**: From August 2026, new rules require large computational loads to withstand defined voltage and frequency disturbances. **Ireland and Northern Ireland**: Requirements on ride-through, recovery and frequency disturbances are awaiting regulatory approval. **Denmark**: Transmission-connected facilities of 200 MW or more must meet ride-through and recovery requirements. **France**: Transmission-connected demand must meet ride-through and recovery requirements.
Behaviour during normal operation	Limiting sharp demand changes, voltage disruption, harmonics and repeated fluctuations that could affect both the local and the wider grid	**Alberta**: In addition to disturbance response, technical requirements for transmission-connected data centres address voltage fluctuation and rapid demand changes. **Finland**: Transmission-connected users ≥ 30 MW must meet, among others, power quality and oscillatory behaviour requirements; broader rules for distribution-connected users are under development. **North America**: NERC guidance recommends collecting data on harmonics, ramp rates and oscillations.
Information, modelling and verification (cross-cutting)	Requiring developers to provide information on how facilities will behave, submit usable models, verify performance and share event data	**Ireland and Northern Ireland**: System operators have issued dedicated requirements for the submission and validation of large load models. **Denmark**: Requires submission of several types of models.

IEA. CC BY 4.0.

While requirements are advancing in several jurisdictions, industry-wide standards and harmonised practices remain at an early stage. Policy makers can accelerate progress where needed by clarifying institutional responsibilities and supporting co-ordination and data sharing among system operators, regulators, network companies, technology providers and data centre developers. Policy makers also need to ensure that system operators and regulators have the tools and mandate needed to manage growing data centre demand without compromising system stability. Alongside jurisdiction-specific grid codes, greater harmonisation of technical, modelling and information requirements could reduce fragmentation and support more predictable facility behaviour.

IEA. CC BY 4.0.

Annex

Data centre and large-load policies by jurisdiction

Selected data centre and large-load measures

Jurisdiction	Category	Relevant instruments and measures
Canada	Permitting and queue management	• British Columbia Industrial Electricity Allocation Framework: AI/DC projects ≥10 MW enter competitive allocation before BC Hydro interconnection; 2026 call allocates 300 MW for AI and 100 MW for data centres (DCs).
	Connection arrangements	• Alberta Data Centre Regulation; Alberta Electric System Operator (AESO) Large Load Integration: framework for transmission-connected DCs ≥75 MW; tethered/bring your own generation projects prioritised, with limited bridged grid access.
	Cost allocation and tariffs	• Alberta Corporate Tax Act amendments: DC levy linked to electricity drawn from the public grid; non-grid/tethered supply can receive 0% levy rate.
	Power sourcing strategies	• Alberta Data Centre Regulation: Bring your own generation/tethered DC model incentivises new/self-supply rather than pure grid draw. • British Columbia Industrial Electricity Allocation Framework: Competitive allocation directs clean electricity to selected AI/DC projects using provincial-benefit criteria.
	Operational integration	• Alberta AESO transmission-connected data centres guide: Transmission-connected DCs subject to technical reliability and operating data requirements.
China	Load projections and spatial planning	• Data Centre Green and Low-Carbon Development Special Action Plan (2024): New large/ultra-large DCs prioritised in national hub clusters; no new large/ultra-large projects where local DC utilisation is <50%.
	Data centre design and efficiency	• Data Centre Green and Low-Carbon Development Special Action Plan (2024): New/expanded large and ultra-large DCs target PUE ≤1.25; national hub DC projects ≤1.2; WUE requirements of below 1.4 L/kWh in water-rich areas and below 0.8 L/kWh in water-deficient regions. • New Data Centre Three-Year Action Plan (2021-2023): Real-time PUE/WUE monitoring and DC energy-efficiency inspections.
	Power sourcing strategies	• Multi-user Green Power Direct Connection Notice (2026): Priority support for computing facilities to use green-power direct connection; annual self-use ≥60% of generation and ≥30% of load, rising to ≥35% before 2030.
European Union	Data centre design and efficiency	• Commission Delegated Regulation (EU) 2024/1364 of 14 March 2024 on the first phase of the establishment of a common Union rating scheme for data centres: DCs above 500 kW IT power demand report energy, water, cooling, waste heat and sustainability key performance indicators. • Energy Efficiency Directive; 2023/1791: DCs >1 MW total rated energy input must use/recover waste heat unless not technically or economically feasible. • Data Centre Energy Efficiency Package: EU rating scheme planned in 2026; first labels in 2027; minimum performance standards assessment by 2027.
	Permitting and queue management	• European Grids Package: Recommends first-ready-first-served, maturity criteria, milestones/penalties and regular queue clean-up. • Strategic Roadmap for Digitalisation and AI in Energy: Promotes use-it-or-lose-it for DC grid requests through future tripartite agreements.
	Cost allocation and tariffs	• Future-proof electricity bills proposal *(forthcoming)*: Expected to clarify network-charge principles so tariffs better reflect system use and flexibility.
	Power sourcing strategies	• Strategic Roadmap for Digitalisation and AI in Energy: Tripartite model promotes power purchase agreements and additional clean generation for DC integration.
	Operational integration	• Electricity Directive; 2019/944: Enables demand response, aggregation and flexibility procurement; DC flexibility can participate through markets where technically eligible.

IEA. CC BY 4.0.

Selected data centre and large-load measures (continued)

Jurisdiction	Category	Relevant instruments and measures
Finland	Load projections and spatial planning	• Fingrid Grid Scope map: Indicative current/future main-grid connection capacity for consumption; non-binding and not a reservation tool.
	Permitting and queue management	• Fingrid connection-contract enquiry procedure: New/increased consumption connections in southern Finland require Fingrid approval; initially applies to consumption connections above roughly 2 MW to 10 MW.
	Cost allocation and tariffs	• Electricity tax reform: Data centre electricity moved from lower tax category II to general category I, increasing duty from 0.05 c/kWh to 2.24 c/kWh.
	Operational integration	• Fingrid Technical requirements for large demand facilities/KJV2026: Demand facilities >30 MW, including DCs, must withstand disturbances and be able to limit power intake.
France	Load projections and spatial planning	• RTE Fast-Track site certification: State-identified sites for very high-power projects, including DCs, based on land availability, 400 kV proximity and manageable grid congestion.
	Data centre design and efficiency	• Code de l'énergie L236-1: DCs ≥500 kW installed IT/server-room power must report and publish energy, environmental and operational data annually. • Code de l'énergie L236-2: DCs ≥1 MW must valorise waste heat; deemed met if energy reuse factor ≥0.20, with techno-economic derogations possible.
	Permitting and queue management	• RTE electrification connection framework (announced/by end-2026): First-ready-first-served to replace first-come-first-served for industrial connection queues, including DCs. • RTE Fast-Track procedure: Accelerated connection route for 400 MW to 1 GW consumers on state-identified 400 kV sites; project must show credible delivery prospects. • RTE Fast-Track site certification: 30% down payment at connection-commitment acceptance, 60% at works start; building permit proof within 15 months and land-rights retention required to avoid termination.
	Cost allocation and tariffs	• RTE Fast-Track site certification/Loi de simplification 2026: Fast-track projects pay connection-work contribution plus site-specific congestion/operating-measure contribution; calculation principles require CRE approval.
Germany	Data centre design and efficiency	• Energy Efficiency Act/EnEfG §11: DCs currently must meet PUE limits; existing DCs ≤1.5 from July 2027 and ≤1.3 from July 2030; new DCs from July 2026 ≤1.2 within two years. • EnEfG amendment *(cabinet-approved draft)*: Relaxes existing-DC PUE to ≤1.6 from July 2027 and ≤1.4 from July 2030; new DCs keep ≤1.2 but with four-year ramp-up. • Energy Efficiency Act/EnEfG §11: New DCs must meet waste-heat reuse shares of 10% from July 2026, 15% from July 2027 and 20% from July 2028, subject to exemptions/relief. • Energy Efficiency Act/EnEfG §13; Data Centre Energy Efficiency Register: DC operators must annually report/publish energy efficiency data to the national register, including electricity use, PUE and reused-energy share.
	Cost allocation and tariffs	• StromNEV §19(2): Large end consumers can apply for individual lower network charges if consumption behaviour qualifies; permanent-load privilege applies from >10 GWh/year at one offtake point, but reform is under way. • BNetzA AgNes industrial network-tariff reform *(under development)*: Future special network tariff planned from 2029 to incentivise flexible use by electricity-intensive consumers; DCs noted as growing high-consumption users.
	Power sourcing strategies	• Energy Efficiency Act/EnEfG §11: DC electricity consumption must be covered by renewable electricity on a net basis; 50% from 2024 and currently 100% from 2027. • EnEfG amendment *(cabinet-approved draft)*: Postpones the 100% renewable electricity requirement from 2027 to 2030.

IEA. CC BY 4.0.

Selected data centre and large-load measures (continued)

Jurisdiction	Category	Relevant instruments and measures
Ireland	Load projections and spatial planning	• Large Energy User Action Plan: Plan-led approach for very large energy users, including hyperscale DCs, with future Green Energy Parks co-locating demand with renewable energy supply. • CRU Large Energy Users Connection Policy; EirGrid Data Centre Constrained Area Overview: SOs assess DC location against constrained/unconstrained grid conditions; system operators (SOs) must publish capacity/constraint information.
	Data centre design and efficiency	• Government Statement on Data Centres; Climate Neutral Data Centre Pact (voluntary benchmark): Favours DCs aligned with PUE ≤ 1.3 in cool climates.
	Permitting and queue management	• CRU Large Energy Users Connection Policy: Applies to new DC applications ≥1 MVA; SOs may refuse/terminate applications that do not meet location, dispatchable generation/storage and renewables criteria. • CRU Large Energy Users Connection Policy: SOs required to publish DC engagement and connection application process, including early engagement, project evidence and required generation documentation.
	Connection arrangements	• CRU Large Energy Users Connection Policy: DCs ≥10 MVA cannot become operational or ramp up demand without delivery of on-site/proximate dispatchable generation or storage covering the maximum import capacity (MIC) on a derated basis. • CRU Large Energy Users Connection Policy: If required energy services underperform, SOs may reduce MIC or move the DC connection to non-firm status.
	Power sourcing strategies	• CRU Large Energy Users Connection Policy: DCs ≥1 MVA must meet at least 80% of annual demand with additional renewable electricity generated in Ireland, with a six-year glide path and annual reporting. • CRU Large Energy Users Connection Policy: Renewables can be delivered through corporate PPAs or direct development; unsupported/additional Irish renewable generation required.
	Operational integration	• CRU Large Energy Users Connection Policy: DCs ≥10 MVA must provide separately metered on-site/proximate dispatchable generation or storage capable of Single Electricity Market participation. • CRU Large Energy Users Connection Policy: DCs of 1 MVA to 10 MVA must provide an auto producer unit matching 100% of the MIC on a de-rated basis and participating in the Single Electricity Market.
Italy	Permitting and queue management	• Decree-Law 21/2026, converted by Law 49/2026, Art. 8: Single authorisation procedure for DC construction, expansion and operation, including grid connection networks of any voltage; maximum 10 months, extendable by up to 3 months.
	Connection arrangements	• Decree-Law 21/2026, converted by Law 49/2026, Art. 8: DC applications may indicate a temporary medium-voltage connection solution, with the definitive grid-connection solution integrated later under the competent authority's procedure.
Japan	Load projections and spatial planning	• GX Strategy Area System: Supports DC clusters based on efficient power and communications infrastructure; candidate municipalities are not publicly named to avoid land/power hoarding. • TEPCO Welcome Zone Map *(regional example)*: Indicates special-high-voltage supply potential, including ≥50 MW and ≥400 MW substation areas.
	Data centre design and efficiency	• METI/ANRE DC Guideline: Covered DC businesses ≥1,500 kl crude-oil equivalent report DC energy data; average PUE benchmark ≤1.4 by fiscal year 2030. • METI/ANRE DC Guideline: From fiscal year 2029, new covered DCs should target design PUE ≤1.28 and operational PUE ≤1.3 after the two-year post-start period.
	Power sourcing strategies	• Decarbonised Power Regional Contribution Investment Promotion Project: Supports DC capital expenditure where projects use 100% decarbonised electricity and contribute to the power source host region; support strength linked to PPA/self-generation, power source type and local contribution.

IEA. CC BY 4.0.

Selected data centre and large-load measures (continued)

Jurisdiction	Category	Relevant instruments and measures
Korea	Load projections and spatial planning	• MOTIE AI/data centre regional dispersion measures: Discloses 345 kV substations with spare capacity and supports faster power-system impact assessment for non-capital-region AI DCs.
	Permitting and queue management	• Special Act on Activation of Distributed Energy; Enforcement Decree: Users needing ≥10 MW in designated areas must undergo power system impact assessment before project approval; assessment covers electricity flow, quality, supply stability and impacts on power use. • AIDC Special Act *(takes effect February 2027)*: One-stop permitting and timeout system for AI data centres; non-capital-region AI data centres below a future size threshold can be exempted from power system impact assessment.
	Connection arrangements	• KEPCO Electricity Supply Terms: New/increased customers >400 MW may receive 154 kV supply with customer-funded demand limiter, maximum demand ≤500 MW, multi-circuit supply and reserve supply unless outage risk is contractually accepted.
	Operational integration	• Electric Utility Act Enforcement Decree Art. 5-5: Supply may be refused for users ≥5 MW if it would undermine electricity quality or power-system reliability. • Special Act on Activation of Distributed Energy: Power system impact assessment can require mitigation measures; non-compliance can lead to compliance orders or project suspension.
Netherlands	Load projections and spatial planning	• Besluit kwaliteit leefomgeving / Bkl hyperscale data centre rules: New hyperscale DCs >10 ha and ≥70 MW are excluded from most of the Netherlands, with limited exempted locations. • Netbeheer Nederland Capaciteitskaart: National offtake-capacity map showing where transport capacity is available, limited, under study or congested with waiting list.
	Permitting and queue management	• ACM Code Decision on prioritising transport requests 2025: network operators must apply a prioritisation framework in congested areas; priority applies to congestion mitigators, safety/security and basic-needs functions, not ordinary DCs by default. • ACM amendment to the Netcode elektriciteit: Requires grid operators to allocate scarce transmission capacity by a priority framework rather than first-come, first-served.
	Cost allocation and tariffs	• ACM time-bound transport tariffs: Large users can receive lower tariffs when avoiding peak hours; large users using full capacity during peak hours pay more.
	Connection arrangements	• ACM alternative transport-rights package: Large users in congested areas can use non-firm/time-bound transport rights where residual grid capacity is available; high-voltage duration-bound contracts guarantee access for at least 85% of the time.
	Operational integration	• ACM congestion-management framework: Large users can be asked or paid to shift/curtail demand to relieve grid congestion; congestion mitigators receive highest priority in the transport-request framework.
Norway	Load projections and spatial planning	• Kart- og planforskriften: DC added as a specific regulatory plan sub-purpose, enabling municipalities to zone sites explicitly for data centre use.
	Data centre design and efficiency	• Forskrift om datasenter: DC operators must register with Nkom and provide location, ownership/legal status, security, preparedness and risk management information. • Energy Act/surplus heat cost-benefit regulation: New or upgraded DCs above 2 MW must submit a Norwegian Water Resources and Energy Directorate-approved cost-benefit analysis for surplus-heat utilisation before construction or upgrade.
	Permitting and queue management	• Statnett increased capacity process: Capacity requests ≥5 MW or 20 GWh require Statnett clarification; queue position depends on a mature order, maturity assessment, cost commitment and binding progress plan. • Statnett capacity reservation rules: Reservation can be cancelled or reduced if the project is delayed, not realised or needs less capacity than reserved.
	Cost allocation and tariffs	• Anleggsbidrag/construction contribution rules: Customers triggering new or increased grid capacity pay a proportional share of required network investment; mandatory across all grid levels.
	Connection arrangements	• Statnett connections with terms framework: Where ordinary capacity is unavailable, grid connection can include conditions such as disconnection or load limitation to enable faster or cheaper access.

IEA. CC BY 4.0.

Selected data centre and large-load measures (continued)

Jurisdiction	Category	Relevant instruments and measures
Singapore	Load projections and spatial planning	• Green Data Centre Roadmap: Plans at least 300 MW of additional DC capacity in the near term, with potentially another 200 MW or more through green energy deployments.
	Data centre design and efficiency	• DC-CFA2; BCA-IMDA Green Mark for Data Centres 2024: Selected DC projects must obtain Green Mark Platinum and achieve PUE ≤1.25 at 100% IT load.
	Permitting and queue management	• Second Data Centre Call for Application/DC-CFA2: Government allocates at least 200 MW of new DC capacity through competitive application, rather than open-ended first-come grid queue.
	Cost allocation and tariffs	• SP Group Transmission Service Rate Schedule: HT/EHT/UHT customers pay contracted-capacity, peak/off-peak, reactive-power and uncontracted-capacity charges; new contracted capacity is binding for five years.
	Power sourcing strategies	• DC-CFA2: At least 50% of proposed new DC capacity must be powered through eligible green energy pathways. • Green Data Centre Roadmap: Recognised green energy pathways include bioenergy, fuel cells with carbon capture, low-carbon hydrogen/ammonia and vertical/building-integrated photovoltaics.
Spain	Load projections and spatial planning	• Real Decreto-ley 7/2026; Red Eléctrica demand access capacity map: Grid operators must report requested/granted demand capacity >1 kV; Red Eléctrica publishes available and occupied demand access capacity by transmission node/zone, updated monthly.
	Data centre design and efficiency	• Draft Royal Decree on energy efficiency and sustainability for data centres *(draft)*: DCs ≥500 kW would report energy, water, cooling, renewable share and sustainability indicators; DCs >1 MW would assess waste-heat reuse; projected DCs >100 MW would need to show top 15% performance against relevant indicators.
	Permitting and queue management	• Real Decreto 1183/2020; CNMC Circular 1/2024: Demand facilities require access and connection permits; demand access capacity can be allocated through standard temporal priority or demand contests in scarce nodes. • Real Decreto-ley 7/2026: Demand permits ≥1 kV must include activity code; activity must be maintained for three years or permits automatically lapse. • Real Decreto-ley 7/2026: High-priority demand projects can suspend competing applications at the same node/capacity zone; liberated capacity >5 MW is reserved for managed reallocation.
	Cost allocation and tariffs	• Real Decreto-ley 7/2026: Consumers with demand access permits ≥1 kV pay a monthly capacity-reservation charge until activity starts; non-payment can trigger permit expiry.
Sweden	Permitting and queue management	• Principles and Guidelines for Connecting to the National Grid: Grid capacity is allocated on a first-come-first-served basis for complete applications, subject to energy analysis, schedule and maturity requirements; reserved allocation can be lost if deadlines or progress requirements are not met.
United Kingdom	Load projections and spatial planning	• AI Growth Zones: Designated AI data centre locations supported through improved access to power and planning support; government proposes reserving/reallocating grid capacity for AI Growth Zones and other strategic demand projects. • National Planning Policy Framework *(England)*: Planning policy now recognises the need for data centres; large DCs can opt into the nationally significant infrastructure planning route.
	Data centre design and efficiency	• Climate Change Agreement for Data Centres *(voluntary government scheme)*: Participating DC operators receive Climate Change Levy discounts in return for energy efficiency targets; 2026-2030 sector target is 14.5% improvement against a 2022 baseline. • Climate Neutral Data Centre Pact *(voluntary industry benchmark)*: Signatories target PUE ≤1.3 in cool climates, WUE ≤0.4 L/kWh in water-stressed areas and assessment of heat-reuse opportunities.
	Permitting and queue management	• TMO4+ Connections Reform: Great Britain grid queue moves to a gated "first ready and needed, first connected" model; demand projects, including DCs, currently need to meet readiness criteria rather than the Clean Power 2030 strategic-alignment filter applied to generation/storage.
	Cost allocation and tariffs	• AI Growth Zones pricing support mechanism *(planned from April 2027)*: Targeted electricity cost reductions for AI Growth Zone data centres in selected regions where additional demand can reduce system costs. • Connection and Use of System Code User Commitment: Transmission-connected demand users face security/final-sums liabilities for connection works; reforms are under way to align treatment across demand users.

IEA. CC BY 4.0.

Selected data centre and large-load measures (continued)

Jurisdiction	Category	Relevant instruments and measures
United States (federal)	Load projections and spatial planning	• Executive Order 14318/Accelerating Federal Permitting of Data Centre Infrastructure: Directs faster federal permitting and use of federal land/resources for qualifying AI data centres and supporting energy infrastructure. • Department of Energy AI data centre federal site selection: Identifies Idaho National Laboratory, Oak Ridge Reservation, Paducah and Savannah River as potential sites for AI data centre and energy infrastructure development.
	Permitting and queue management	• FERC RM26-4 large-load interconnection proceeding *(not final)*: Considers reforms for interconnection of large loads >20 MW, including data centres, to the interstate transmission system.
	Connection arrangements	• FERC PJM co-location order: Requires PJM to create transparent tariff rules for AI data centres and other large loads co-located with generation, including interim non-firm transmission service while network upgrades are completed. • FERC-approved SPP CHILLS tariff *(regional RTO example)*: Conditional high-impact large-load service gives eligible large loads non-firm transmission service using available capacity for up to seven years while pursuing firm service.
	Cost allocation and tariffs	• White House Ratepayer Protection Pledge *(voluntary federal pledge)*: Participating AI/DC companies commit to pay for required power delivery upgrades and not pass new data centre infrastructure costs to ordinary households.
United States (Pennsylvania)	Data centre design and efficiency	• Governor's Responsible Infrastructure Development (GRID) *(state support eligibility)*: GRID projects must report expected PUE/WUE, water use, co-located generation/storage and non-emitting power share; sustainability plan must use Leadership in Energy and Environmental Design Gold for Data Centres, Energy Star, ISO 50001/14001 or equivalent.
	Permitting and queue management	• Governor's Responsible Infrastructure Development (GRID) *(state support eligibility)*: Qualifying DCs receive GRID certification for PA Permit Fast Track, computer equipment sales tax benefits and preferential tax zone participation; GRID Ready Report and annual documentation required.
	Cost allocation and tariffs	• Governor's Responsible Infrastructure Development (GRID) *(state support eligibility)*: Qualifying DCs must pay all transmission, distribution, network upgrade, energy and ancillary service costs caused by the project, avoiding cost shifts to other ratepayers. • Pennsylvania PUC Large Load Model Tariff Framework: Applies to >50 MW individual or >100 MW aggregate loads; requires cost responsibility, deposits/collateral, study timelines, contract terms, queue transparency and stranded-cost protection.
	Power sourcing strategies	• Governor's Responsible Infrastructure Development (GRID) *(state support eligibility)*: Qualifying DCs must build, bring or buy incremental generation in the same PJM Locational Deliverability Area as the project; clean-firm share rises from 10% in 2027 to 14.5% in 2030 and 32% in 2035. • Governor's Responsible Infrastructure Development (GRID) *(state support eligibility)*: DC facilities >100,000 ft² must be solar-ready for rooftop/building-integrated PV.
United States (Texas)	Permitting and queue management	• ERCOT PGRR145/NPRR1325 Batch Zero process: Transitional batch-study process for 75 MW+ large-load interconnections meeting maturity and commitment criteria; replaces individual studies for Batch Zero.
	Connection arrangements	• ERCOT NPRR1325 Batch Zero process: Allows large loads to be evaluated as Provisional Controllable Load Resources or Withdrawal-Limited Private Use Networks where controllability or associated generation affects grid withdrawals.
	Cost allocation and tariffs	• Texas Senate Bill 6: Requires large-load interconnection standards, site control evidence, financial commitment standards and at least a USD 100 000 initial transmission screening study fee for qualifying large loads.
	Operational integration	• Texas Senate Bill 6/ERCOT reliability service: Requires protocols for new post-2025 transmission voltage loads to be curtailable during firm load shed, and a demand management service using ≥75 MW large-load reductions during anticipated emergency conditions. • Texas Senate Bill 6: Large loads with non-exporting on-site backup generation able to serve ≥50% of on-site demand must disclose it and may be directed during emergencies to deploy backup generation or curtail load.
United States (Virginia)	Cost allocation and tariffs	• Virginia SCC GS-5 large-load rate class: Dominion customers ≥25 MW enter a new rate class; minimum charges require 85% of contracted transmission/distribution demand and 60% of generation demand.
	Operational integration	• Dominion Energy Virginia Facility Interconnection Requirements: Large data centre customers must demonstrate voltage stability and ride-through capability and may need monitoring/mitigation to avoid harmful oscillations or unintended large-load loss.

IEA. CC BY 4.0.

Glossary

ACER	European Union Agency for the Cooperation of Energy Regulators
AEMO	Australian Energy Market Operator
AEP	American Electric Power
AI	Artificial intelligence
AWS	Amazon Web Services
BYOP	Bring your own power
CAISO	California Independent System Operator
CO_2	Carbon dioxide
CRE	Commission de régulation de l'énergie
CRU	Commission for Regulation of Utilities
DBSCAN	Density-based spatial clustering of applications with noise
EDO	Energy Demand Outlook
EMS	Directorate of Energy Markets and Security
EMO	Energy Modelling Office
ENTSO-E	European Network of Transmission System Operators for Electricity
ERF	Energy reuse factor
ERCOT	Electric Reliability Council of Texas
ESM	Electricity Systems and Markets Division
EU	European Union
FERC	Federal Energy Regulatory Commission
G7	Group of Seven
GPU	Graphics processing unit
GW	Gigawatt
IT	Information technology
kV	Kilovolt
kW	Kilowatt
kWh	Kilowatt-hour
MISO	Midcontinent Independent System Operator
MVA	Megavolt-ampere
MW	Megawatt
NERC	North American Electric Reliability Corporation
Ofgem	Office of Gas and Electricity Markets
PPA	Power purchase agreement
PUE	Power usage effectiveness
RTE	Réseau de Transport d'Électricité
SERT	Server Efficiency Rating Tool

IEA. CC BY 4.0.

SME	Small and medium-sized enterprise
SMR	Small modular reactor
SONI	System Operator for Northern Ireland
SPEC	Standard Performance Evaluation Corporation
SPP	Southwest Power Pool
TEPCO	Tokyo Electric Power Company
TPU	Tensor processing unit
TWh	Terawatt-hour
UPS	Uninterruptible power supply
USD	United States dollar
VPP	Virtual power plant
WUE	Water usage effectiveness

See the IEA glossary for a further explanation of many of the terms used in this report.

IEA. CC BY 4.0.

International Energy Agency (IEA)

This work reflects the views of the IEA Secretariat but does not necessarily reflect those of the IEA's individual Member countries or of any particular funder or collaborator. The work does not constitute professional advice on any specific issue or situation. The IEA makes no representation or warranty, express or implied, in respect of the work's contents (including its completeness or accuracy) and shall not be responsible for any use of, or reliance on, the work.



Subject to the IEA's Notice for CC-licenced Content, this work is licenced under a Creative Commons Attribution 4.0 International Licence.

This document, as well as any data and map included herein, are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.
Unless otherwise indicated, all material presented in figures and tables is derived from IEA data and analysis.

IEA Publications International Energy Agency Website: www.iea.org
Contact information: www.iea.org/contact

Typeset in France by IEA - October 2026
Cover design: IEA
Photo credits: © Shutterstock

