McKinsey & Company

What is a data center?

A data center is a facility that houses and runs IT infrastructure that's critical to the digital economy, particularly gen AI.

July 2025

When you type a question into a gen AI platform, you receive an answer so fast that it may feel like magic. Of course, it's not: The models that many of us rely on, both personally and professionally, are the result of decades of research and trillions of dollars of investments—not to mention vast and ever-increasing amounts of energy.

Data centers are specialized facilities that manage IT infrastructure, including servers, storage devices, and network equipment. They play a critical role in processing, storing, and distributing large amounts of data, making them essential to gen AI and the rest of the digital economy.

McKinsey analysis indicates that by 2030, data centers will need $6.7 trillion of worldwide investment to keep pace with the demand for compute power—around 70 percent of which will come from AI workloads (Exhibit 1). "Over the next decade," says McKinsey Senior Partner Pankaj Sachdeva, "the industry will go through an S-curve of demand growth to support the infrastructure that will power the digital revolution and continue to power the cloud revolution." Not only will existing data centers need to become more powerful, but new data centers will also need to be built apace. How can the world meet this quickly growing demand? Find out this and more in this *McKinsey Explainer*.

Learn more about McKinsey's Technology, Media & Telecommunications Practice.

What is compute power?

Compute power is emerging as one of this decade's most critical resources. The rise of AI has led to skyrocketing demand for compute power, or the hardware, processors, memory, storage, and energy needed to operate data centers.

What are the core components of a data center?

There are four core components to most data centers:

— *IT equipment.* Data centers host servers, storage devices, and network devices that handle data processing, storage, and transmissions needs.

— *Infrastructure and utilities.* Data centers are equipped with air-conditioning, redundant electricity systems, and electricity conditioning to ensure uninterrupted operations.

— *Connectivity.* Data centers are typically located near high-bandwidth fiber networks that enable low-cost, high-speed data exchange.

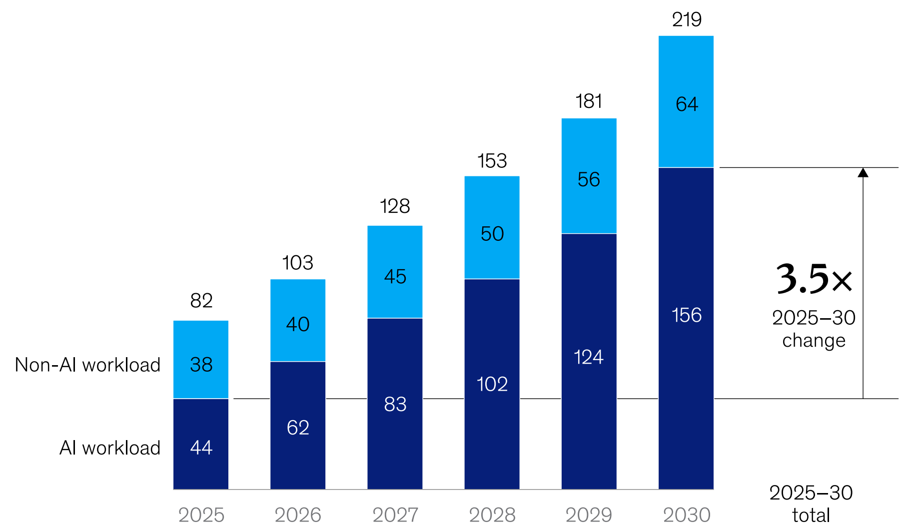
— *Physical security.* Robust physical security measures, including fire suppression systems and restricted access, are typically implemented to protect the center's equipment and data.

Learn more about McKinsey's Technology, Media & Telecommunications Practice.

Exhibit 1

Both AI and non-AI workloads will be key drivers of global data center capacity demand growth through 2030.

Estimated global data center capacity demand, 'continued momentum' scenario, gigawatts



	2025	2026	2027	2028	2029	2030	2025–30 total
Incremental AI capacity added per year, gigawatts	13	18	21	19	22	31	124

Note: Figures may not sum to totals, because of rounding.
Source: McKinsey Data Center Demand Model; Gartner reports; IDC reports; Nvidia capital markets reports

McKinsey & Company

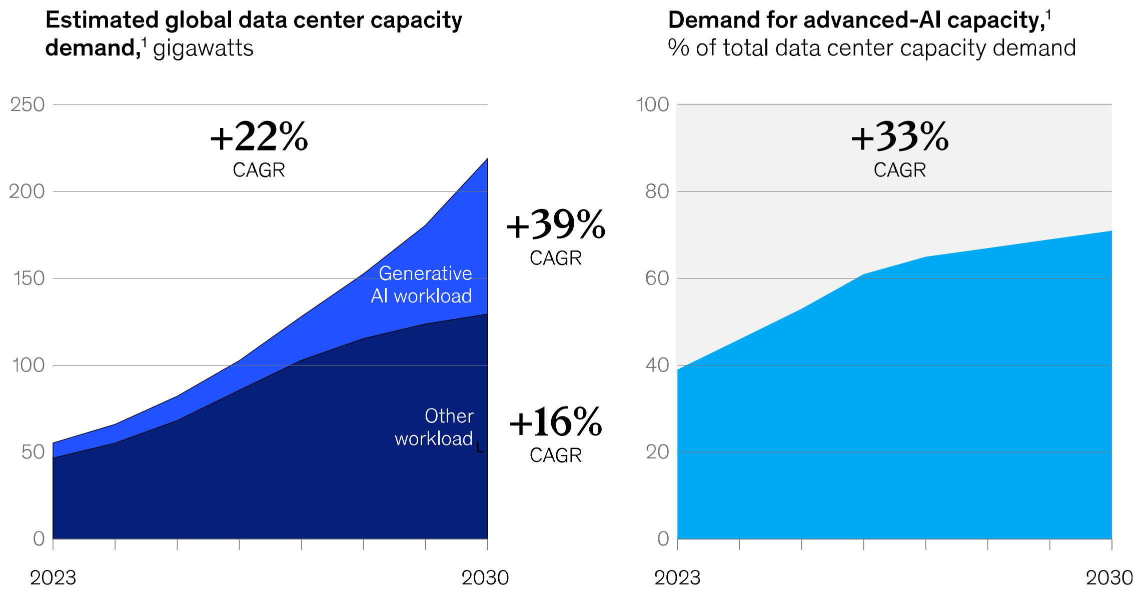
How is AI influencing the growth of data centers?

AI's boom has fueled skyrocketing demand for power. As noted above, 70 percent of the projected demand for data center capacity will come from AI-based workloads by 2030. McKinsey analysis suggests that in a midrange scenario, demand for AI-ready data center capacity will rise at an average rate of 33 percent per year between 2023 and 2030. Gen AI, currently the fastest-growing advanced-AI use case, will account for around 40 percent of the total demand (Exhibit 2).

To keep up with the rapid rise of AI, data centers have become bigger and more powerful. Ten years ago, a center with 30-megawatt capacity was considered large; today, a 200-megawatt campus is considered normal. AI-ready data centers consume an especially large amount of

Exhibit 2

AI is the key driver of growth in demand for data center capacity.



[1]Midrange scenario is based on analysis of AI adoption trends; growth in shipments of different types of chips (application-specific integrated circuits, graphics processing units, etc) and associated power consumption; and the typical compute, storage, and network needs of AI workloads. Demand is measured by power consumption to reflect the number of servers a facility can house.
Source: McKinsey Data Center Demand model

McKinsey & Company

energy because of their high average power densities—that is, the energy consumption of servers in the racks. Average power densities have more than doubled in just two years and are expected to rise nearly four times by 2027.

Hyperscalers including Amazon Web Services, Google Cloud, Microsoft Azure, and Meta are the companies fueling most of today's incremental demand for AI-ready data centers. That's because these hyperscalers require massive capacity to host both the large models they develop in-house, such as Google's Gemini, and those developed by AI companies, such as OpenAI's ChatGPT. Cloud service providers currently own more than half the world's AI-ready data centers. McKinsey estimates that by 2030, up to 65 percent of AI workloads in Europe and the United States will be hosted on hyperscalers' infrastructure.

Most other companies are using off-the-shelf models that are largely hosted on a public cloud. But as AI matures, more organizations are likely to build and train their own models based on internal data, which could increase demand for private hosting.

What do new AI-ready data centers require?

The higher energy demand and power density, as well as the complexity of different AI workloads, are leading to rapid change in three main areas of data center construction:

— *Location and power infrastructure.* As data centers proliferate, power supply is becoming an issue in markets that have traditionally attracted clusters of data centers, such as Northern Virginia and Santa Clara, California, in the United States. Many utilities find that they haven't been able to build transmission infrastructure quickly enough, raising concerns that they may be unable to generate sufficient power in the future. Inadequate power generation can slow data center expansion and affect the overall consumer and business use of AI.

— *Mechanical (cooling) system design.* AI servers consume so much energy that they get physically hot, so much so that air-based cooling systems can't keep up. This issue has prompted a shift to approaches that remove heat directly from racks by using liquid, which is more efficient at absorbing and transferring heat than air: for example, rear-door heat exchangers, direct-to-chip technology, and liquid immersion cooling.

— *Electrical system design.* AI workloads call for larger power distribution units that can cope with higher power densities—and in response, many data center operators are installing larger switchgear and floor-mounted distribution units. These changes reduce the complexity as well as the capital and operational costs of installing and maintaining multiple smaller units.

Learn more about McKinsey's Technology, Media & Telecommunications Practice.

Do data centers also support non-AI tasks?

AI workloads may dominate the conversation, but non-AI processing loads and cloud make up a significant portion of data center activity. These workloads include traditional enterprise IT tasks such as web hosting, enterprise resource planning systems, email, and file storage. Non-AI tasks require less compute power and can operate efficiently on central processing units, rather than the specialized graphics processing units or AI accelerators that AI workloads require.

Non-AI loads also tend to have more predictable usage patterns and lower power densities, which allow for less demanding cooling and energy requirements. As a result, data centers that focus on non-AI processing typically have different infrastructure needs, capital intensity, and operational considerations compared with those that are intended primarily for AI workloads.

What regional challenges does the data center sector face?

In Europe, the data center sector faces several challenges, including limited sources of reliable power, sustainability concerns, insufficient power infrastructure, land availability issues, shortages of power equipment, and a lack of skilled electrical tradespeople. In major markets, it can take up to five years or more to supply power to new data centers, and the power grid is increasingly strained. Meeting these demands will require a lot of clean energy, which will in turn require the construction of more new energy systems that can be turned on in times of especially high demand.

In the United States, the data center sector faces more significant challenges, particularly in terms of power connections and labor constraints. There is also a shortage of electrical trade workers, which affects the ability to execute projects on time and can delay the building of data centers and associated power infrastructure. Tariffs have also increased an element of uncertainty and could present additional supply chain complexities.

Learn more about McKinsey's Technology, Media & Telecommunications Practice.

How can energy players get involved in the data center sector?

Investors have plenty of opportunities to participate and can enable solutions for power access and sources. Here are four high-potential areas:

— *Secondary markets with access to reliable, cheap power.* There's a timing gap between data center builds—which can be done in 18 to 24 months—and power infrastructure development, which can take anywhere from three to ten years (sometimes even more) to complete. But there are creative ways to bridge this gap. Many hyperscalers are building out capacity in new, nontraditional locations outside core data center markets because these areas have access to cheaper, available power as well as the potential to build carbon-free infrastructure. In the United States, Iowa, Wyoming, Indiana, and Ohio each house or have received investments from at least two of the top four hyperscalers.

— *Behind-the-meter solutions.* These solutions provide power in areas where utilities providers cannot keep up with the pace of demand or reliability requirements as transmission constraints or the availability of local power supply worsen. For example, there are opportunities for investors to build power that can be fully islanded outside the grid or provide supplemental power (such as nuclear) to complement the existing grid.

— *Sustainability ambitions driven by renewable-energy providers.* Hyperscalers that have made climate commitments will require hundreds of terawatt-hours of clean energy to meet future demand. Solar and onshore wind are expected to generate most of the world's new clean energy, but other clean energy technologies can also supply energy in the medium to long term. These sources include offshore wind, nuclear, geothermal, gas, carbon capture and storage, and clean fuels.

— *Transmission and distribution investments.* Power availability is critically important to meeting data center demand, and utility companies are taking notice. Investors can funnel investments into utility companies to build out transmission and distribution infrastructure in key markets.

Learn more about McKinsey's Technology, Media & Telecommunications Practice.

What role can real estate organizations play in the data center industry?

Data centers present three main opportunities to real estate organizations. At the most basic level, real estate organizations can buy the land for data centers—specifically, parcels of land that they think will get the power, network connections, and customer demand needed for data centers—then sell or lease this land to data center developers.

Real estate companies can also do partial data center development, then sell their facilities to developers or final customers. Finally, companies can create a "co-location" model: They buy the land, construct the building, provide access to power and connectivity, and build out the interior. These facilities are often leased by enterprises or hyperscalers.

"Compared to other real estate asset classes," says McKinsey Senior Partner Pankaj Sachdeva, "data centers historically have had higher yield. Due to supply-side constraints, we don't anticipate that real estate players or data center developers will face major yield compressions over the long term. Real estate investors are seeking exposure to data centers to get higher growth and sustained levels of higher yield."

How can telecom operators capitalize on the growing demand for AI infrastructure?

Beyond providing the infrastructure that powers communication and connects people, telecom operators can build the infrastructure that will realize AI's full potential. One way is laying down fast internet cables (or fiber to connect new data centers, which will help people and businesses access powerful cloud services). Telecom operators can also turn unused space into profit by offering AI computing power, known as graphics processing units, or GPUs, as a service for rent. Another option is building and running their own data centers to support the high computing power and internet speed that AI requires. They could also address additional gaps to bridge connectivity platforms.

Because they already have wide coverage and experience managing networks, telecom companies are in a strong position. But they can also partner with companies that manufacture computer chips, build data centers, or offer other tech services.

It's important to have a clear aspiration and ambition for capabilities and investment needed to succeed, especially as operators go further up the stack from the core connectivity infrastructure.

To reduce risk, telecom firms can start small by investing gradually in AI-ready infrastructure and making use of what they already have. Clear communication and a nimble approach will be key to success.

Learn more about McKinsey's Technology, Media & Telecommunications Practice. And check out job opportunities related to data centers if you're interested in working with McKinsey.

Articles referenced:

— "The cost of compute: A $7 trillion race to scale data centers," April 28, 2025, Jesse Noffsinger, Mark Patel, and Pankaj Sachdeva, with Arjita Bhan, Haley Chang, and Maria Goodpaster

— "AI infrastructure: A new growth avenue for telco operators," February 28, 2025, Abhyuadaya Shrivastava, Brendan Gaffey, Gustav Grundin, Sebastian Cubela, and Tomás Lajous, with Akshat Agarwal, Dapo Orimoloye, Lorraine Salazar, Miguel Frade, and Nicholas Shaw

— "How hyperscalers are fueling the race for 24/7 clean power," December 18, 2024, Lorenzo Moavero Milanesi, Tjark Freundt, and Yuito Yamada, with Fridolin Pflugmann and Marc Ludwig

— "AI power: Expanding data center capacity to meet growing demand," October 29, 2024, Bhargs Srivathsan, Marc Sorel, and Pankaj Sachdeva, with Arjita Bhan, Haripreet Batra, Raman Sharma, Rishi Gupta, and Surbhi Choudhary

— "The role of power in unlocking the European AI revolution," October 24, 2024, Anna Granskog, Diego Hernandez Diaz, Jesse Noffsinger, Lorenzo Moavero Milanesi, and Pankaj Sachdeva, with Arjita Bhan and Sofia von Schantz

- "What the real estate industry needs to know about data centers," October 15, 2024, Pankaj Sachdeva

- "How data centers and the energy sector can sate AI's hunger for power," September 17, 2024, Alastair Green, Humayun Tai, Jesse Noffsinger, and Pankaj Sachdeva, with Arjita Bhan and Raman Sharma

- "The hard stuff: Navigating the physical realities of the energy transition," August 14, 2024, Mekala Krishnan, Chris Bradley, Humayun Tai, Tiago Devesa, Sven Smit, and Daniel Pacthod

- "Investing in the rising data center economy," January 17, 2023, Srini Bangalore, Arjita Bhan, Andrea Del Miglio, Pankaj Sachdeva, Vijay Sarma, Raman Sharma, and Bhargs Srivathsan

Get to know and directly engage with senior McKinsey experts on data centers

Alastair Green is a senior partner in McKinsey's Washington, DC, office; **Andrea Del Miglio** is a senior partner in the Milan office; **Humayun Tai** is a senior partner in the New York office; **Jesse Noffsinger** is a partner in the Seattle office; **Mark Patel** is a senior partner in the Bay Area office; **Marc Sorel** is a partner in the Boston office; **Pankaj Sachdeva** is a senior partner in the Philadelphia office.