We are *not on track* to meet the Paris Agreement's objectives. What should we do?

World Energy Markets Observatory 2024 | 26th Edition

Capgemini

IN COLLABORATION WITH

Hogan Lovells | vaasaETT | Enerdata

08

How to meet Data center AI/Gen AI consumption increase?

Caroline Vateau
Sustainable IT Director
Capgemini Invent

Caroline Vateau is one of the French national experts on sustainable IT, she has been working for more fifteen years to measure and reduce the environmental impacts of digital technology from semi-conductors, to datacenters, cloud, applications and information systems. She also manage the Digital Tech for Green Program aiming to identify and assess the digital technologies that could be leverage to accelerate the path to ecological transition.

IT needs to accelerate data center decarbonisation

1. Global IT energy use is already larger than that of many countries, yet its flexible nature makes it a strong candidate for rapid decarbonisation.
2. The levers available include shifting compute power around the world, improving optimisation of the use of data centers, improved design, and use of waste heat in heat networks.
3. Recent alliances between technology companies and power companies are evidence of the strategic nature of power to IT.

Introduction

IT is already a major energy user, and with accelerating energy demand due to advances such as AI, 5G and IOT, IT must be a major contributor to decarbonisation. This will be achieved through continuing the efficiency push, increased compute flexibility, careful locating of data centers, and reuse of waste heat. There is a risk that data centers could destabilise grids by rapidly increasing or decreasing demands, so greater collaboration at the data center to grid interface is needed.

Technology advances are also an important enabler to the Energy Transition, and we will need sophisticated energy management systems, advanced designs for turbines, and better predictive models, all of which will be underpinned by AI and generative AI.

The digital economy accounts for approximately 6-12% of global electricity consumption, translating to 1-2% of global energy use.[1] This results in an estimated 0.69-1.6 gigatons of CO_2 emissions, or about 1.5-3.2% of global greenhouse gas emissions in 2020.[2]

While data centers represent just one aspect of the IT sector, they are the most concentrated in terms of energy consumption, accounting for 460 TWh, approximately 2% of global electricity demand in 2022, according to the International Energy Agency (IEA).[3] The electricity consumption magnitude of a country like France!

This figure is expected to increase significantly given the rising demand for new and advancing technologies, like AI and generative AI, as well as blockchain and cryptocurrency operations. By 2025, global data volume is expected to reach 175 zettabytes (ZB), a marked increase from the 2018 level of 33 ZB.[4]

While data center operators have been actively working on energy efficiency improvements for decades and achieved notable results, the rapid and substantial growth in data and compute power requirements from intelligent technologies, AI, IOT-enabled devices, and 5G, is outpacing these efficiency gains. This growth is, in turn, putting immense pressure on some local power systems and grids, which are struggling to provide the necessary electricity at the required speed, especially during peak periods.

To meet the growing demand for compute power while striving to reduce their environmental impact, data centers need to innovate and integrate more sustainable energy production methods into their operations and even become energy producers themselves. In this article, we explore how the dual challenge of increasing energy efficiency and managing escalating demand underscores the critical role of strategic planning and technological advancements in the future of the digital sector.

The impact of data centers on the grid

The energy profile of data centers has a unique impact on the electric systems due to its relatively flat and stable electricity consumption. Unlike other forms of energy usage that experience significant variations throughout the day or year, data centers maintain a consistent level of demand. This, predictability, is advantageous for systems and grid operators, as it allows for more accurate forecasting and planning.

In one way, this makes data centers a prime opportunity to be served through decarbonized energy sources. However, to enable this shift, organizations must address the challenges of:

- Energy storage since renewable energy sources are intermittent while data center needs are constant.
- Local production to avoid grid developments.

At the same time, there are potential drawbacks: If data centers switch from the grid to backup generators, it often means that they will shift from clean energy to fossil fuels. This can negate or undermine the environmental benefits of using green energy sources.

This issue highlights the need for sustainable backup and storage solutions, as well as local green electricity production to ensure that the environmental impact of data centers is optimized even during power disruptions.

1 UN Trade & Development: Digital Economy Report 2024. https://unctad.org/publication/digital-economy-report-2024
2 UN Trade & Development: Digital Economy Report 2024. https://unctad.org/publication/digital-economy-report-2024
3 IEA: Electricity 2024. https://iea.blob.core.windows.net/assets/18f3ed24-4b26-4c83-a3d2-8a1be51c8cc8/Electricity2024-Analysisandforecastto2026.pdf
4 IDC: Data Age 2025. https://www.seagate.com/files/www-content/our-story/trends/files/Seagate-WP-DataAge2025-March-2017.pdf

Decentralizing compute power

The decentralization of compute power is poised to significantly transform the data center model and its associated energy dynamics.

One of the most prominent forms of decentralization is the development of edge data centers. These smaller facilities, located close to the network edge and, by extension, end-users offers several important advantages, including the ability to be powered by local renewable energy sources. This makes them ideal for supporting advanced digital use cases, such as 5G communications, which require high compute and low latency.

In addition, smaller, edge-based data sites can be more easily integrated into local energy management schemes, allowing for the reuse of waste heat, as well as better alignment with renewable energy source availability and avoid grid developments. This localized approach not only enhances energy efficiency but also contributes to a more sustainable and resilient energy ecosystem.

Data centers as an energy producer

The decentralization of compute power could lead to a parallel decentralization of energy production. With respect to data centers, this means developing their own electrical production and/or storage capabilities.

With the potential to establish micro-electric power sites or connect directly to renewable energy sources, data centers could become more self-sufficient and reduce their reliance on traditional power grids.

For example, in Ireland, new policies are being introduced that would allow private companies, most notably, data center operators, to build their own energy infrastructure. This could provide a workaround for a de facto moratorium on data center developments in the greater Dublin area that is in effect through 2028.

Taking a holistic approach to energy management

While the growing use of advanced digital technologies are increasing energy consumption, these very same applications can also have the potential to significantly enhance sustainability efforts for energy companies.

For example, AI and generative AI can support the reduction of electricity consumption through optimized energy management and predictive maintenance. AI can also be used to analyze vast amounts of data to identify patterns and inefficiencies within energy grids, buildings, factories, and transportation systems. This analysis can lead to smarter energy distribution and consumption strategies, minimizing waste and lowering overall electricity usage.

The “push and pull” nature of these technologies highlight the need for organizations to take a holistic approach to sustainability measurement, assessing both the direct and indirect impacts of AI implementation. It also requires companies to have a more complete understanding of the total impact of new technologies.

For example, optimizing the efficiency of EVs through AI could lead to more extensive use of these vehicles, potentially offsetting some of the initial gains in efficiency—a phenomenon known as the rebound effect.

To truly enhance sustainability, energy companies must consider the broader implications of AI and generative AI applications, ensuring that the overall environmental impact remains positive. Integrating AI thoughtfully into operations can help energy companies achieve a more sustainable future while navigating the complexities of energy consumption and ecological impact.

Actions: Making IT more sustainable

IT functions are increasingly challenged to make better and more sustainable decisions to reduce their environmental impact. Here we explore several key actions to effect change:

1. Optimize existing resources. Within IT departments there is a notable potential for optimizing existing resource utilization. For example, analysis conducted by the Capgemini Research Institute revealed that 30% of cloud resources were either unused or misused. By addressing such inefficiencies, IT departments can reduce energy consumption and minimize the environmental footprint of their digital operations. This optimization extends beyond mere energy savings, encompassing the broader potential of digital tools to accelerate the ecological transition.
2. Integrate eco-design into project management and digital service conception. Eco-design is a global approach that allows environmental issues to be integrated at the earliest stages of the project lifecycle. This is crucial because an estimated 80% of emission impacts are determined during the design phase, making them difficult to reduce later. To address this challenge, organizations must develop frameworks to apply to projects at the earliest stages to enable teams to make informed choices about functional specifications, architecture, hosting, and other points.
3. Repurpose waste heat and embrace direct liquid cooling (DLC) to improve data center operations. Data center operators have an opportunity to repurpose waste heat generated in data centers. While the waste heat from data centers typically has a lower temperature than what is required for district heating systems, the use of heat pumps or similar converters can be used to increase the temperature and make it suitable for various applications, such as heating swimming pools, supporting agricultural activities in greenhouses, or warming nearby neighborhoods. Integration of these solutions into the broader energy management schemes of new industrial zones can enhance their overall efficiency and sustainability.

Further, data center operators can also implement innovative cooling technologies to further reduce the environmental impact of data centers. DLC, which uses a refrigerant or water directly at the processor and server levels to optimize cooling efficiency, offers a more effective cooling system than traditional air-based systems. As this technique gains traction among data center operators, IT functions have the possibility to substantially lower their energy consumption, contributing to a more sustainable digital infrastructure.

5 Capgemini Research Institute: World Cloud Report. https://www.capgemini.com/insights/research-library/world-cloud-report/
6 ScienceDirect: Sustainable Production and Consumption https://www.sciencedirect.com/science/article/pii/S2352550920314433#bib0040

About Capgemini

Capgemini is a global business and technology transformation partner, helping organizations to accelerate their dual transition to a digital and sustainable world, while creating tangible impact for enterprises and society. It is a responsible and diverse group of 340,000 team members in more than 50 countries. With its strong over 55-year heritage, Capgemini is trusted by its clients to unlock the value of technology to address the entire breadth of their business needs. It delivers end-to-end services and solutions leveraging strengths from strategy and design to engineering, all fueled by its market leading capabilities in AI, cloud and data, combined with its deep industry expertise and partner ecosystem. The Group reported 2023 global revenues of €22.5 billion.

Visit us at

www.capgemini.com/wemo

Scan to download
this year's report