Do You Have Power?

Where does the AI-driven power surge in data centres leave sustainability goals? On the fence.

By Raju Chellam

Here's a geeky futuristic story for the new year: By 2028, AI has become pervasive, and robots have gotten cheaper and smarter. John, a struggling comedian, invests in a robot that can do everything—cooking, washing, cleaning, and taking care of him. John often takes the robot along to his gigs. The robot pulls jokes from the cloud and rehashes old jokes by using AI to tweak them to John's style.

One night when John comes home, he finds his robot missing. After a tiring search, he finds the robot performing stand-up comedy at a nightclub. The "idiot" robot is hamming away with lousy puns and awkward pauses, but the audience is in stitches. The delighted club owner offers John a fat sum of money to let the robot perform every night. John grabs the opportunity. The result? John is now doing everything—cooking, washing, cleaning, and taking care of the "idiot" robot.

If that anecdote made you wink, these estimates should make you blink: Electricity consumption in data centres (DCs) is projected to reach 857 terawatt hours (TWh) by 2028, more than double the 2023 levels, according to IDC Corp. The increasing demand for AI workloads is set to significantly boost DC capacity, energy consumption, and carbon emissions.

> MCKINSEY'S RESEARCH SUGGESTS THAT GENAI COULD GENERATE BETWEEN US$2.6 TRILLION AND US$4.4 TRILLION IN ECONOMIC VALUE GLOBALLY. HOWEVER, ACHIEVING EVEN 25% OF THIS POTENTIAL BY THE END OF THE DECADE WOULD REQUIRE AN ADDITIONAL 50 TO 60 GW OF DC INFRASTRUCTURE IN THE US ALONE.

THE AI IMPETUS

A massive impetus comes from AI, especially GenAI workloads which now account for an increasingly larger share of total DC power usage. This trend is set to fuel energy consumption in DCs at a CAGR (compound annual growth rate) of 44.7% from now to 2027, surpassing 146 TWh by then.

The problem? Electricity prices are rising in tandem due to several factors, including supply and demand dynamics, environmental regulations, geopolitical events, and climate change-induced extreme weather. The rising demand for power and escalating energy costs will make DCs significantly more expensive to operate, though the exact impact remains uncertain.

IDC conducted a scenario for a DC with a 1MW (megawatt) IT load operating at 50% capacity and a PUE (power usage effectiveness) of 1.5. The study examined three energy price growth scenarios using data from Germany, Japan, and the US. In all scenarios, the growth in power expenditure exceeded a CAGR of 15%, with most scenarios shooting expenses at 20%. The study also highlighted that a 10% improvement in energy efficiency could result in substantial savings for DC operators.

Research house Gartner anticipates that by 2027, 40% of existing DCs running AI workloads will face operational problems due to power constraints. By 2027, the power needed for DCs to operate AI-optimized servers will reach 500 TWh annually, about 2.6 times the amount used in 2023.

"Larger DCs are being planned to manage the vast amounts of data required to train and implement the rapidly growing LLMs (large language models) that drive GenAI apps," says Bob Johnson, a Gartner vice president. "Short-term power shortages are likely to persist for years, as new power transmission, distribution and generation capacities will take time to come online. This could disrupt energy availability, leading to shortages that will hinder the growth of new DCs for GenAI and other applications starting in 2026."

What about the US, the largest and fastest-growing market for DCs? The demand for DCs is expected to grow rapidly in the US—from 25 GW in 2024 to over 80 GW by 2030. McKinsey's research suggests that GenAI could generate between US$2.6 trillion and US$4.4 trillion in economic value globally. However, achieving even 25% of this potential by the end of the decade would require an additional 50 to 60 GW of DC infrastructure in the US alone.

Meeting this demand will require significantly more power than is currently produced in the US. This increase in electricity needs is unprecedented, as overall power demand in the US has barely grown since 2007. DCs could account for 30-40% of all new demand added until 2030, with growth also coming from domestic manufacturing, electric vehicles, etc. Between 2024 and 2030, electricity demand for DCs in the US is expected to rise by about 400 TWh, with a CAGR of 23%.

GenAI is a game changer for more sustainable IT. How it changes the game, though, is up to every organization that uses it. The crux? "Up to 99% of executives are unable to quantify their IT ops' exact carbon footprint," says an IBM IBV (Institute for Business Value) study. "The good news: GenAI can help solve the very problem it may contribute to over time. It is creating new opportunities to transform IT and make it more sustainable."

About 25% of businesses are already leveraging hybrid cloud solutions to significantly boost the sustainability and energy efficiency of their IT ops. "Nearly half (46%) report a substantial positive impact on their overall IT sustainability," the study reports.

WHITHER SUSTAINABILITY?

Where does the AI-driven power surge in DCs leave sustainability goals? On the fence.

Zero-carbon sustainability goals will also be adversely affected by short-term measures to increase power supply, as the rising demand forces suppliers to boost production by any means necessary. This sometimes involves keeping fossil

> DC PROVIDERS, INCLUDING CLOUD AND COLOCATION SERVICES, SHOULD CONTINUE TO PRIORITIZE INVESTMENT IN RENEWABLE ENERGY SOURCES AND HELP TO INCREASE OVERALL SUPPLY WHILE HELPING THEIR CUSTOMERS MEET THEIR SUSTAINABILITY GOALS.

fuel plants operational beyond their planned retirement dates.

"Increased DC usage will lead to higher CO_2 emissions to generate the necessary power in the short term," Johnson notes. "This will make it more challenging for DC operators and their customers to meet ambitious CO_2 emission reduction targets."

DCs require continuous power availability, which renewable sources like wind or solar cannot provide without alternative supplies during non-generating periods. Reliable and continuous power can only be generated by hydroelectric, fossil fuel, or nuclear power plants at this point. In the long term, innovative technologies such as improved battery storage (sodium-ion batteries) or clean power (small modular nuclear reactors) might become available to help achieve sustainability goals.

In the US, the industry faces the daunting challenge of decarbonizing its footprint to achieve carbon-free energy usage by 2030. "While the carbon emission intensity for power grids is set to drop in the next ten years, generation from natural gas is expected to increase," McKinsey says. "At the same time, most grid decarbonization timelines (if they exist) exceed the targets set by major hyperscalers."

Moreover, increasing rack power densities (from 50 to 100 kilowatts per rack) are leading to higher power ratings for equipment such as transformers, switchgear, and PDUs (power distribution units). This requires changes and innovations in product lines and could create opportunities for entrants. Specializing in modularization and prefabrication technologies, such as mechanical, engineering, and plumbing packages, could help hyperscalers build new DCs more quickly.

Would companies have the means to continue to pursue sustainability goals? "Organizations should re-evaluate sustainability goals relating to CO_2 emissions in light of future DC requirements and power sources for the next few years," Gartner recommends. "When developing GenAI applications, they could focus on using a minimum amount of computing power and look at the viability of other options such as edge computing and smaller language models (instead of LLMs)."

What about other options to increase DC efficiency? Technological solutions like improved chip efficiency and liquid cooling could help. So would redesigning DCs and power distribution. "Providing energy-efficient solutions is only part of the equation for meeting customer needs," says Sean Graham, IDC research director for DC trends. "DC providers, including cloud and colocation services, should continue to prioritize investment in renewable energy sources and help to increase overall supply while helping their customers meet their sustainability goals."

The bottom line: Solar and wind power offer significant environmental advantages while also providing the lowest LCOE (levelized cost of electricity), which reflects the average net present cost of power generation over a generator's lifetime. IDC recommends collocating facilities near the source of renewable energy generation. Providers can reduce construction costs and energy losses associated with distribution and enhance overall efficiency and sustainability.

Since we started with a geeky futuristic quip, let's end with a cheeky one: AI systems have become more pervasive and intuitive by 2028. A swarm of AI-powered robots have formed their own comedy troupe. They hit the stage to perform for a packed house, with the crowd being a mix of humans and robots. The show starts with a robot walking onto the stage and asking the audience: "Why did the robot cross the road?" The audience leans forward in anticipation. The robot pauses for a moment before deadpanning: "To get to the other circuit board." DQ

Raju Chellam is a former Editor of Dataquest and is currently based in Singapore, where he is the Editor-in-Chief of the AI Ethics & Governance Body of Knowledge and Chair of Cloud & Data Standards.

maildqindia@cybermedia.co.in