On-site Power Generation for Data Centers

Integrating Data Centers with Low Carbon Energy Assets

Reza Ahmadi Kordkheili, Vattenfall

Trevor Hinkle, SDIA

Mohan Gandhi, SDIA

Niek Van der Pas, Legrand

Pooya Davari, Aalborg University

Executive Summary

The Information and communications technology (ICT) sector is one of the fastest-growing sectors in the world. It is apparent that the large amount of information and digital services we consume via the Internet also demands physical resources in the form of large, concentrated information technology (IT) infrastructure. We see this especially in the growth of data centers and their energy and resource consumption. Data centers consumed approximately 270 terawatt hours (TWh) of electricity globally in 2019, or about 1% of total global power consumption, according to the International Energy Agency (IEA). This is expected to rise as 5G, artificial intelligence (AI), and other enabling technologies unlock a host of new applications. Facing this development in a time where renewable energy is still scarce requires data centers to explore and utilize power supply concepts that support the energy system of the future.

The SDIA and its members advocate supporting the integration of data centers with energy assets, be they decentralized renewable energy assets or centralized Combined Heat and Compute (CHC) assets. Such integration would enable distributed, resilient, and environmentally friendly digital infrastructure, and facilitate the decarbonization of our energy system. This paper outlines, at a high level, the concepts that could be employed to deliver the twin transition of decarbonization of our energy systems and the digitalization of our economy. This paper aims to highlight where our focus should turn next in creating a truly sustainable digital economy by 2030.

We call on policy-makers and industry stakeholders to acknowledge energy asset integration and similar concepts as a feasible solution to energy-, climate-, digital sovereignty-, and resiliency-related challenges, and support their development and rollout with respective adjustments in policy frameworks and funding from the public and private sector.

Table of contents

1
Context – The Twin Transitions

1.1
The energy system of the future is transitioning

Europe is in the midst of an energy transition, with the aim to cover energy demand with renewable power while adjusting to the increasing fluctuations of a wind- and solar-driven electrical grid. As intermittent renewables penetrate further into the electricity generation mix, flexibility becomes increasingly valuable. The phase out of flexible supply and inadequate storage capacities add to the potential strain on the electrical grid. Hence, the generation, consumption, storage, and migration of energy will become crucial to balancing the energy grid of tomorrow.

1.2
The Digital Infrastructure Industry is maturing

Data traffic is expected to grow exponentially over the next decade as digital use cases become increasingly embedded in our lives. The smart home, the office in the cloud, the autonomous vehicle, and the robot operator are all becoming increasingly numerous as Internet speed, bandwidth, and accessibility improve. Internet connectivity will connect more consumers and more devices, all of which produce increasingly data- and compute-intensive workloads. The large and growing size of data centers make them particularly applicable as providers of energy flexibility, grid stability, and ancillary services.

The rapid growth of cloud computing, data centers with skyrocketing energy consumption, and the accelerating penetration of renewable energy sources is creating both severe challenges and tremendous opportunities. Data centers as flexible energy operators could open up a unique opportunity to smooth out the significant fluctuation and uncertainty of renewable generation.

1.3
Integration is the key

Grid constraints, expensive storage solutions, and the phasing out of flexible fossil fuels all increase the value of energy flexibility. Co-locating large energy consumers like data centers with renewable assets aligns with the decentralized nature of the grid of the future rather than trying to mimic the characteristics of past “demand-first” energy systems. Large constant base loads could be a burden in the future, especially as prices fluctuate. It is imperative that energy be consumed with proximity to supply and in a flexible manner. Data centers can be proximal, flexible energy suppliers. The SDIA’s Utility of the Future report, where digital and electrical infrastructure combine, elaborates further on how these two critical industries could integrate.

2
The Procurement and Consumption of Green Energy

How can data centers consume 100% renewable electricity? The answer to this question has progressed over time. The first step was to offset; the next was to match usage with like-for-like contracted energy purchases, which is typically done by major corporations through large power purchase agreements (PPAs). The final stage is direct consumption of locally generated renewables, either in real time or through recently generated stored energy

2.1
Purchasing offsets

Carbon offsetting is a means of paying others to reduce or absorb emissions to compensate for your own generated emissions. It is fast becoming a quick and easy way of purchasing carbon forgiveness. In many ways, it enables companies to continue to do what they do, and shift responsibility (and guilt) to the end customer.

The problem with this approach is that it does not actually reduce the amount of carbon dioxide (CO2) entering the atmosphere. Tree planting and carbon capture initiatives are overall a good thing, but carbon from fossil fuels is never truly offset. Carbon stored in a tree is not the same as the carbon that would have been left underground had we not consumed the fossil fuel in the first place. When a tree dies and decomposes, it releases that stored CO2 into the atmosphere. If we were to assume, in a roundabout way, that the offset schemes' captured carbon was equal to the carbon released from burning fossil fuels, then by the offsetters own logic, we are only offsetting until that tree decomposes. In other words, it is temporary storage and will only last as long as the tree is alive. It might be better than nothing, but it is not a permanent solution and does nothing to improve the incentives of large emitters to become more efficient, or to consume greener energy.

There is also the issue of trustworthiness. It is difficult to verify claims made by carbon offsetters because these schemes often take place in places where reporting and transparency standards are particularly poor. There have been attempts to create certification and build trust in this industry, but little progress has truly been made. Offsetting, therefore, is largely a public relations exercise by large companies.

2.2
Contracting energy

Overall, green power purchase agreement contracting has resulted in the tech industry being the largest corporate purchaser of renewable energy. Yet, only large corporations with sufficient financial backing are likely to be accepted by a generator as the counterparty to their PPA.

Corporate consumers and energy companies will claim that large green PPAs help to stimulate investment in renewable generation, by giving the generator a guaranteed source of income for a set period of time. The problem with contracting energy is it is simplifying and abstracting the problem, to the extent that the depth of the problem is ignored. While renewable energy is being generated somewhere, that may not be where your data center is consuming electricity.

It is important to distinguish between physical consumption (the physical electrons flowing to the data center) and procurement (the contractual purchases of power by the data center), the latter of which do not necessarily correspond to physical flow of power. At its core, these are accounting methods, and there is not necessarily one “correct” method to describe the power consumption of a data center.

Green PPA’s are not a bad thing, and have certainly helped to stimulate investment in renewable energy projects. However, they are exclusive only to large corporations, and effectively privilege large corporations with green energy (from an accounting perspective) while potentially undermining the ability of the average grid consumer from consuming greener energy themselves. The average grid consumer pays for the downside of the PPA (e.g., increased variability, higher grid fees, etc.) but will not likely see the upside. Hence, there is only a certain amount of PPA that can be made available in a grid zone before it starts to negatively affect other consumers in that zone.

2.3 Direct consumption

The goal for data centers now should become the direct consumption of green energy. The funding, generation and consumption of local renewable energy would reduce the overall carbon intensity of the local grid, and therefore reduce the carbon footprint of energy physically consumed by anyone within that grid zone, including the data center. Low carbon on-site generation would go a step further and ensure data centers were consuming green electrons in a manner that does not compromise the ability of others to themselves consume green electrons. The next chapters detail the challenges and opportunities to on-site integration of data centers with low-carbon energy-generating assets.

The goal for data centers (and all energy consumers) in the future should be the timing of consumption to coincide with when the energy supply is greenest. This acknowledges that timing and volume of the physical consumption of electricity are important factors in determining whether an energy consumer is truly sustainable. An idealized example would be if a significant amount of the computing load was able to shift to a time where the grid would otherwise have wasted “excess” renewable energy (e.g., when wind or solar energy otherwise would have been “curtailed” due to no demand). Google termed this “carbon-aware computing.” Real-time, carbon-aware energy consumption is beyond the scope of this paper; however, it certainly builds on the challenges and opportunities laid out in the following chapters.

3
The Challenges to Integration

3.1
On-site Integration with solar and wind assets

It is becoming increasingly popular for data centers to consider renewable energy via on-site solar and wind generation to directly offset electricity consumption and contribute to environmental sustainability. The introduction of intermittent and non-dispatchable renewable energy for powering data centers that generally host mission critical workloads presents a significant challenge, however. The carbon footprint can be greatly influenced by the energy sources used, and attention is shifting to metrics such as Renewable Energy Factor (REF). Recently, there have been efforts to exploit and reuse or combine green energy sources in data centers to lower CO2 emissions.

3.2
On-site integration with Energy Storage

There are a number of forms of energy storage, but they all seek to solve a certain need: redundant supply of energy should the grid fail. Most data centers use diesel generators in conjunction with batteries as a form of storage, and these diesel generators can run indefinitely (as long as the supply of diesel can be maintained). They are cheap, well-known to operators, and very reliable. So, any sustainable alternative would need to overcome performance, cost, and trust thresholds before it was even considered by data center operators.

Batteries are a promising form of energy storage. The U.S. Energy Information Administration (EIA) issued a report claiming grid-scale battery-project costs in the United States dropped 70 percent in just five years, from US$2,152 per kilowatt-hour of storage to US$625. In order to deliver tier-3 levels of redundant supply, however, data centers need 36-72 hours of emergency power. Batteries are very large and require large amounts of space, and the price for battery storage will always come on top of the price of energy generation. At the end of the 72-hour blackout period, should the primary supply not be back online, the secondary will fail too. Many argue that a 72-hour blackout is unrealistic, but it does not change the fact that being online and available is the primary purpose of a data center. Larger batteries that might help smooth the variability from on-site renewable supply, but batteries only have a short number of charge and discharge cycles, which significantly affects the financial feasibility of regularly using batteries as well as their sustainability in terms of material consumption and electronic waste (e-waste) generated.

Energy storage can be found in energy wells, however, such forms of energy storage are still largely dependent on weather seasonality. So far, most energy sinks are used to heat and cool buildings, and rarely extend to the charging and discharging of electricity. Typically, energy sinks only offer a limited amount of dispatchable, reliable energy. With data centers becoming more energy dense, and with interest being raised in energy recovery systems, though, heat sinks could play a more prominent role in the integrated energy storage realm.

Overall, the concept of reliably storing energy on-site faces a number of difficulties – not least around cost, performance, and technical development.

3.3
On-site integration with rampable sources

Intermediary storage options are also being increasingly investigated. This is where an energy source is converted into a different form, before being converted back to energy. An example might be the Australian startup Lavo Hydrogen Technology Ltd., which uses power from rooftop solar panels to produce hydrogen from water by electrolysis. The gas is stored on site and converted back to electricity on demand through a hydrogen fuel cell.

Rampable on-site generation sources are not unheard of, but they are very niche. They often fail to capture the economies of scale that large energy generators can provide, and they may well require additional competences to be employed on site. They are also notoriously inefficient because the many transformations reduce the overall round-trip efficiency.

3.4
Grid connections and grid services

Hydrogen fuel cells are particularly rampable. Such an approach can ramp not just to the needs of the data center, but could also be used to deliver energy back to the grid at times of peak demand or short supply. This is a form of demand response. Monetizing demand response is still a challenge. Most consumers cannot trade directly into the electricity markets because they are too small; hence, they use either their energy supplier or an aggregator. Aggregators typically combine the demand response of small customers to produce a pool of flexibility that can be sold into the electricity markets. Every aggregator needs an agreement with each electricity supplier, and every customer needs a clause enabling aggregators to engage in demand response on their behalf. In other words, the contracts are complicated, creating significant risk.

Market design is therefore the largest barrier to adoption of demand response, and there is little incentive for either the data center or the energy supplier. Participating in demand response or energy flexibility programs may reduce availability or lead to a higher risk of downtime. This risk is exacerbated by the potential surrendering of control to aggregators and energy suppliers. This is control that the data center may not have in the first place – retail colocation data centers simply provide space, security, power, and fiber to tenants who operate their own IT workloads independently. Changes to the energy service level agreements would need to occur between data centers, energy suppliers, aggregators, and transmissions and distribution system operators (TSO/DSOs) to reduce complexity and risk before this option would be considered.

3.5 Heat grid integration

The liquid cooling output temperatures within data centers are steadily approaching the input temperatures required in district heating grids. Output temperatures from a liquid cooling system should be in the region of 75-120 degrees Celsius (167-248 degrees Fahrenheit) to be eligible for conventional district heating systems. Currently waste heat emitted from liquid-cooled data centers is in the region of approximately 50 degrees Celsius, which puts it at the low-quality side. Yet, we are seeing both an increase in output temperatures from data centers as well as more low-temperature district heating grids requiring lower input temperatures. Heat pumps can also be used to convert low-quality heat into high-quality heat before being injected into the heat grid.

Data centers are also becoming increasingly large generators of waste heat. This waste heat is being generated in cities, where district heating demand is growing fastest, and where waste heat can be recycled most efficiently. This also avoids the use of or at least reduces the dependency on natural gas. Additionally, the regulatory environment is shifting the incentive structure more favorably toward recycling recovered heat into district heating grids. A second major benefit to data center operators is the benefit heat recovery might have on the Energy Reuse Factor metric, a metric that is becoming more prevalent by the day.

Despite its promise and increasing adoption rates, challenges persist that hinder the wider-scale adoption of waste heat recycling. The large initial capital costs are a political barrier to adoption, even though lower operating expenditures generally support the economic case. Utilities are also wary of leaving large capital expenditures idling should the regulatory, and therefore the business environment, change. Hence a stable and structured approach to regulatory change is required to ensure utilities and data centers are confident enough to make the high initial investments.

Valorizing heat is perhaps the biggest hurdle for data centers. Putting a price on the heat that can be injected into district heating grids is so far very difficult. Heat production profiles from data centers would also need to match the heat demand profiles of district heat grids (or independent heat consumers), otherwise the risks involved to the data center may well be high. Additionally, one should not underestimate the physical disruption caused by building or upgrading district heating systems.

4
Opportunities of integrated digital and energy infrastructure

4.1
Supporting the twin transitions

The integration of the twin transitions will enable our energy system to switch to a more renewable, lower carbon system. Distributed, sovereign renewable energy infrastructure is the perfect starting point for realizing a distributed and resilient digital infrastructure. The increase of own-consumption on-site reduces requirements for increased grid capacity. Physical connection and proximity results in greater incentives for synchronizing consumption with renewable supply, thereby reducing imbalances in the grid as well as helping to integrate more renewables. In many cases, limited local supply of renewables and market monopolies limit access to renewable energy sources for data centers. On-site generation allows smaller and medium-sized data centers to access renewable energy where previously only large operators could procure or contract such amounts of renewables.

4.2
Resource & energy efficiency

On-site generation allows for the redesign of specific components, which has potential to reduce electrical losses through improved energy density and reduced transmission and conversion losses. Power often gets converted from alternating current (AC) to direct current (DC) and vice versa multiple times as it travels from the grid to IT equipment, which can lead to significant power losses. As a result, you end up paying for energy they do not actually use (and contributing more CO_2 emissions if powered by fossil fuels).

Given the increasingly variable cost of grid generated energy, the opportunity for energy arbitrage exists where the data center optimizes their energy costs when the cost of generating power onsite is less than the cost of buying the power. They can save even more by using the heat associated with the generation of power that is ordinarily wasted. Energy cost savings alone can often justify the installation of distributed generation, and more than 100 cogeneration facilities currently operating in our system have rewarded their owners with substantial energy cost savings.

4.3 Enhanced availability & reliability

On-site generation can provide an essential redundancy to power provided by the local electric utility. One can use on-site power generation to fully support data centers and replace the increasingly unpredictable electricity grid altogether. On-site generation in concert with small power storage systems can offer improved power quality. Harmonics, voltage sag, frequency variations and momentary outages (enough to trip expensive computers) are all increasingly common on the electricity grid. The power electronics inside data centers are more sensitive to such grid imperfections than ordinary consumer items. Data center operators can therefore improve power quality and hence equipment reliability through on-site generation. This approach leads to increased reliability and stability, enhanced energy efficiency and, depending on the source of that power, greater environmental sustainability - all of which can result in significant long term - operational cost savings.

Reducing the dependence on the electricity grid can also help to increase the uptime of a data center. Local distribution grids are increasingly overcrowded as demand has outpaced capacity. This is especially true in dense urban settings. On-site power generation allows data center operators to act independently of the existing electrical infrastructure, while operating independently of any changes in demand to that same infrastructure.

5
Conclusion: We call for more support for integrating data centers with low carbon energy assets

In its “Shaping Europe’s Digital Future” strategy document, the European Commissions mandates the following: “Data centres and telecommunications will need to become more energy efficient, reuse waste energy, and use more renewable energy sources. They can and should become climate neutral by 2030.” Without proper engagement from industry, researchers, and policy-makers alike, defining such ambitious targets will only be a theoretical milestone, leading to yet another disappointing outcome.

The data center industry as a whole, including all its value chain, needs to proactively develop and implement new solutions in order to mitigate the gap between the current status quo and the EU commission’s 2030 targets.

Power consumption is undeniably one of the cornerstones of any efforts toward a more sustainable future. We believe that the direct integration of data center infrastructure and renewable energy assets can open a great window of opportunity for all the industry players, while significantly boosting the industry’s sustainability metrics and revamping society’s perspective on the industry as a whole. Direct integration via on-site power production from renewable assets (e.g., wind, solar, hydro, etc.) paired with data centers, not only guarantees a physical delivery of green electrons to the data center, but also helps minimize the pressure on the grid and reduces the losses in the energy chain. Such an alternative has certain challenges such as intermittency, backup power, and cost, that need to be addressed. Our group is inviting all interested parties to join forces in order to find the right solutions to these challenges without sacrificing the core performance criteria of a data center to be the highly available home of the digital economy.

We ask all parties involved in the decision-making process, including but certainly not limited to regulators, policy-makers, industry players, universities, and research centers, to join us in shaping the right solutions and addressing the main challenges. We believe that we all need to work together in order to frame the challenge correctly, hence developing a holistic, system-based and sustainable solution that addresses our society’s concerns and creates a positive future for the next generations.

About SDIA

Established in 2019 and co-based in Germany and the Netherlands, the Sustainable Digital Infrastructure Alliance e.V. (SDIA) is a nonprofit network of more than 65 members and partners working to catalyze the transition to a sustainable digital economy. The SDIA brings together stakeholders from across industries and fields, both public and private, to realize its Roadmap to Sustainable Digital Infrastructure by 2030. It is meant to offer a holistic, systems-thinking approach to solving the challenges facing ICT sustainability, ranging from energy supply and data centers to fiber-optic networks and software.

If you' d like to join the Roadmap and participate in the Steering Group work, please apply here. In case you have a questions on the above paper, reach out to Mohan Gandhi, Head of Research & Analysis author at:
mohan.gandhi@sdialliance.org

Explore the Roadmap >>

Learn more about the Alliance >>

The publication of the above paper was possible due to the across industry boundaries collaboration on progressing the Roadmap to Sustainable Digital Infrastructure by 2030.