ENERGY INNOVATION
POLICY & TECHNOLOGY LLC

How Much Energy Do Data Centers Really Use?

March 17, 2020 (https://energyinnovation.org/2021/05/04/)

Energy Innovation partners with the independent nonprofit Aspen Global Change Institute (https://www.agci.org/) (AGCI) to provide climate and energy research updates. The research synopsis below comes from Eric Masanet (https://www.mccormick.northwestern.edu/research-faculty/directory/profiles/masanet-eric.html) and Nuoa Lei (https://www.linkedin.com/in/nuoa-lei-70846790/) of the McCormick School of Engineering and Applied Science (https://www.mccormick.northwestern.edu/) at Northwestern University, and a full list of AGCI's quarterly research updates covering recent climate change research on clean energy pathways is available online at https://www.agci.org/solutions/quarterly-research-reviews (https://www.agci.org/solutions/quarterly-research-reviews)

Data centers can be thought of as the "brains" of the internet. Their role is to process, store, and communicate the data behind the myriad information services we rely upon every day, whether it be streaming video, email, social media, online collaboration, or scientific computing.

Data centers utilize different information technology (IT) devices to provide these services, all of which are powered by electricity. Servers provide computations and logic in response to information requests, while storage drives house the files and data needed to meet those requests. Network devices connect the data center to the internet, enabling incoming and outgoing data flows. The electricity used by these IT devices is ultimately converted into heat, which must be removed from the data center by cooling equipment that also runs on electricity.

On average, servers and cooling systems account for the greatest shares of direct electricity use in data centers, followed by storage drives and network devices (Figure 1). Some of the world's largest data centers can each contain many tens of thousands of IT devices and require more than 100 megawatts (MW) of power capacity—enough to power around 80,000 U.S. households (U.S. DOE 2020).

As the number of global internet users has grown, so too has demand for data center services, giving rise to concerns about growing data center energy use. Between 2010 and 2018, global IP traffic—the quantity of data traversing the internet—increased more than ten-fold, while global data center storage capacity increased by a factor of 25 in parallel (Masanet et al. 2020). Over the same time period, the number of compute instances running on the world's servers—a measure of total applications hosted—increased more than six-fold (see Figure 3) (Masanet et al. 2020).

These strong growth trends are expected to continue as the world consumes more and more data. And new forms of information services such as artificial intelligence (AI), which are particularly computationally-intensive, may accelerate demand growth further. Therefore, the ability to quantify and project data center energy use is a key energy and climate policy priority.

Data center energy use estimates: A tale of two methods

Official statistics are not currently compiled on data center energy use at national or global levels. Therefore, mathematical models must be used to estimate this energy use. So-called “bottom-up” models account for the installed stocks of IT devices in different data centers and their energy use characteristics to arrive at an estimate of total energy use. While bottom-up studies offer many insights into the drivers of energy use, they are also very data and time intensive; therefore, they don’t appear often. For example, the most authoritative bottom-up study in the last decade appeared in 2011 (Koomey 2011), and it estimated that data centers accounted for between 1.1 percent and 1.5 percent of global electricity use in 2010.

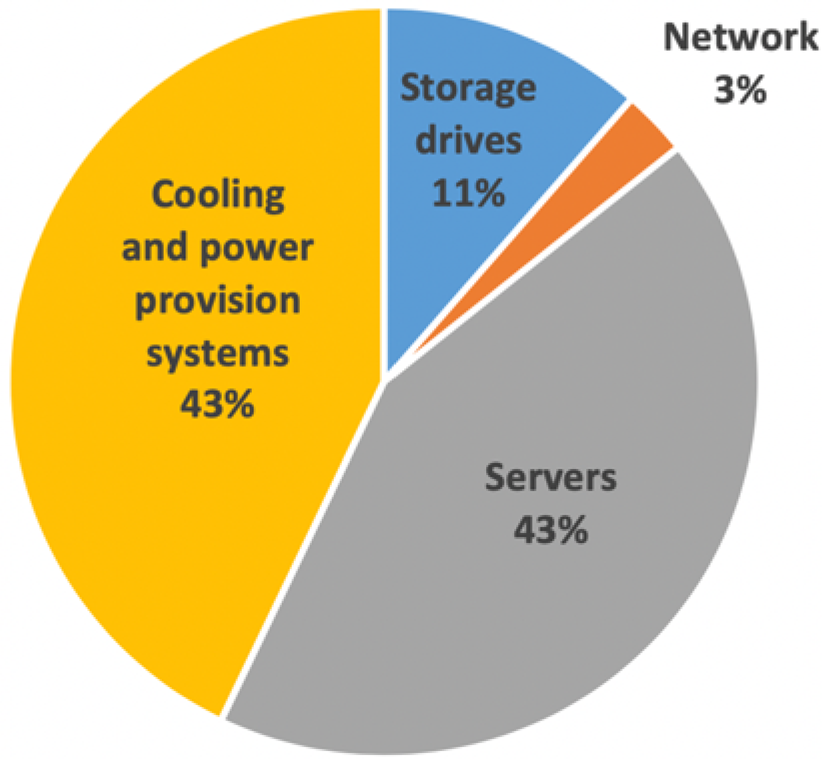

Figure 1. Fraction of U.S. data center electricity use in 2014, by end use. Source: Shehabi 2016.

In contrast, extrapolation-based models estimate total energy use by taking previous bottom-up values and scaling them up on the basis of data center market growth indicators, such as global IP traffic (Andrae and Edler 2015) or data center investments (Belkhir and Elmeligi 2018). Because extrapolation-based approaches are much simpler, they have been used to fill the temporal gaps left behind by sporadic bottom-up studies.

However, such extrapolations tend to estimate large increases in data center energy use, given that the market indicators upon which they rely also grow rapidly. For example, some oft-cited extrapolations have suggested that global data center energy may have doubled since 2010, and by extending this historical logic, that it will continue rising swiftly in the future (Andrae and Edler 2015, Andrae 2017, Belkhir and Elmeligi 2018, Bawdy 2016). These estimates have received significant attention (Jones 2018), reinforcing the common belief that rapidly growing demand for data equals rapidly growing data center energy use.

However, new results from the bottom-up perspective indicate otherwise: Despite rapid growth in demand for information services over the past decade, global data center energy use likely rose by only 6 percent between 2010 and 2018 (Masanet et al. 2020). These new results were based on integration of numerous recent datasets, which better characterize the installed stocks, operating characteristics, and energy use of data center IT devices, as well as structural shifts in the data center industry, compared to past studies.

The finding that global data centers likely consumed around 205 terawatt-hours (TWh) in 2018, or 1

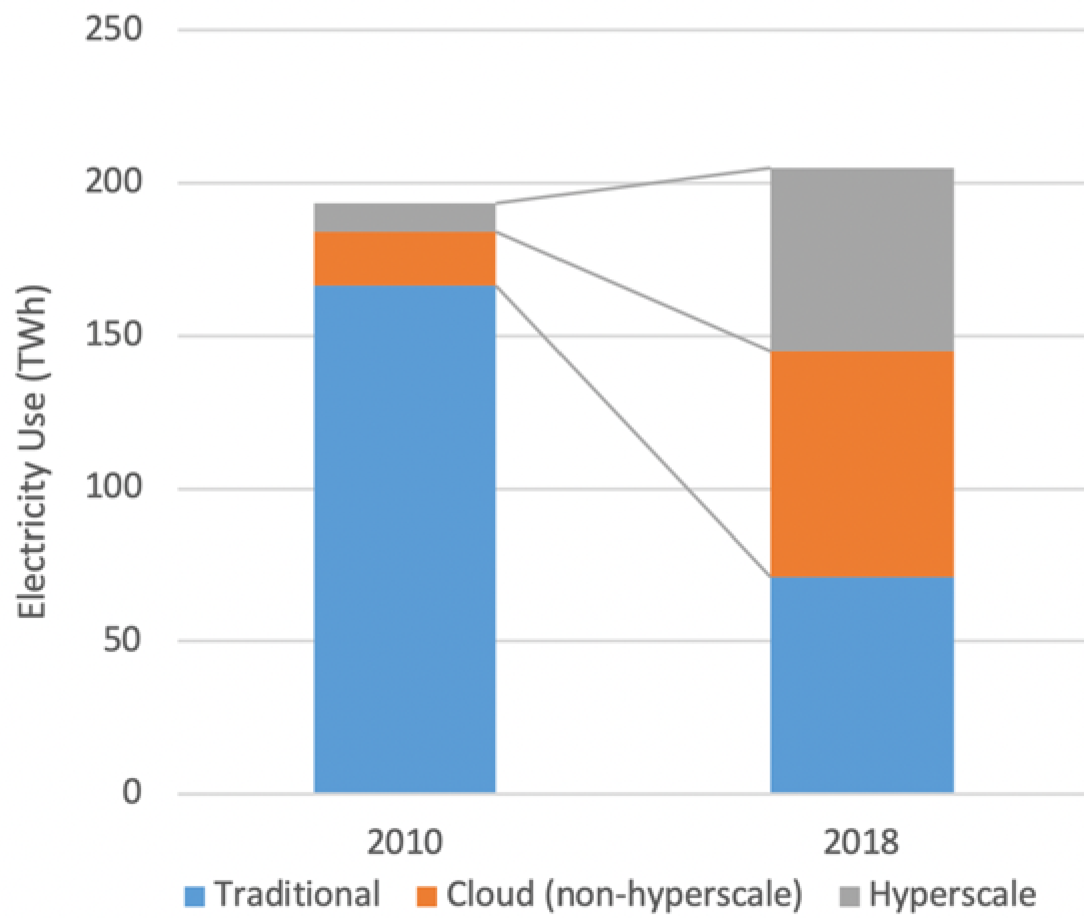

Figure 2. Estimated global data electricity use by data center type, 2010 and 2018. Source: Masanet et al. 2020.

percent of global electricity use, lies in stark contrast to earlier extrapolation-based estimates that showed rapidly-rising data center energy use over the past decade (Figure 2).

Three primary efficiency effects explain this near-plateau in energy use: First, the energy efficiency of IT devices—and servers and storage drives in particular—has improved substantially due to steady technological progress by IT manufacturers. Second, greater use of server virtualization software, which enables multiple applications to run on a single server, has significantly reduced the energy intensity of each hosted application. Third, most compute instances have migrated to large cloud- and hyperscale-class data centers, which utilize ultra-efficient cooling systems (among other important efficiency practices) to minimize energy use (Figure 2).

These efficiency effects are not well captured in extrapolation-based approaches, given their lack of technological and structural detail. In other words, while extrapolation-based approaches generally capture the drivers of data center demand that would push energy use up (top half of Figure 3), they fail to adequately capture strong countervailing efficiency trends (bottom half of Figure 3) that keep energy use in check.

What about CO2 emissions?

The substantial electricity use of data centers also gives rise to concerns over their carbon dioxide (CO_2) emissions. Unfortunately, it is not yet possible to accurately estimate total CO_2 emissions, due to a lack of data on the locations of the vast majority of global data centers and the emissions intensities (measured in grams CO_2 per kilowatt-hour) of their actual electricity sources. Only a handful of companies, including Google, Apple, Switch, and Facebook publicly report such data, indicating a growing trend among some of the world's largest data center operators toward renewable energy procurement.

Knowing the electricity use of global data centers, however, provides a useful benchmark for testing claims about the CO_2 implications of data center services. For example, one oft-repeated claim is that the world's data centers emit as much CO_2 as the global aviation industry (Pearce 2018), which is roughly 900 billion kilograms of CO_2 (Air Transport Action Group 2020). Considering that global data centers recently consumed around 205 billion kWh, for this claim to be true, their average electricity emissions intensity would have to be around 4.4 kg CO_2/kWh.

Yet the average coal-fired power plant, the most carbon-intensive option available, has an emissions intensity that is less than one-fourth of this value—around one kilogram CO_2/kWh (U.S. EIA 2020). And it's clear that all the world's data centers are not running on coal, especially in light of renewable energy use by some large data centers that account for increasing shares of global compute instances.

Another recent claim is that "the emissions generated by watching 30 minutes of Netflix (1.6 kg of CO_2) is the same as driving almost four miles." This claim is backed up by assumptions that data centers providing Netflix streaming services would consume around 370 TWh per year (Kamiya 2020). Yet this value is 1.8 times larger than the 205 TWh estimated for all of the world's data centers combined, which provide society with myriad other information services beyond just streaming Netflix videos. (For a more complete assessment, see Kamiya 2020.)

Therefore, the improved clarity that these recent bottom-up estimates have brought on global data center use can also enable "reality checks" that expose the implausibility of some attention-grabbing and widely-circulated claims about data centers' contribution to climate change.

The path forward

That said, in the coming decade, a significant risk exists that rapidly growing demand for information services—and compute-intensive applications like AI in particular—will begin to outpace the efficiency gains that have historically kept data center energy use in check. Potential still remains for substantial efficiency gains but investments in next-generation computing, storage, and heat removal technologies will be required to avoid potentially steep energy use growth later this decade. And parallel investments in renewable power sourcing will be required to minimize the climate implications of unavoidable data center energy use (Masanet et al. 2020).

Better modeling capabilities are required for decision makers to confidently evaluate future efficiency and mitigation options, so developing more robust and predictive methods that increase the frequency of bottom-up insights and overcome the limitations of extrapolation-based forecasts are a key priority for the energy analysis community.

These models will be needed by policy makers and energy planners for monitoring future data center energy use trends, understanding key energy use drivers, and assessing the effectiveness of various policy interventions for managing possible energy growth. Because data centers are present nearly everywhere, such capabilities will be required at both national and global levels, and particularly for China, where data center capacity is expanding rapidly.

Analysts should consider several key priorities. First, the development and open sharing of reliable data sources on installed stocks, configurations, and the energy use characteristics of IT devices and cooling/power systems will enable more common and accurate technology representation across models. Such technological detail will also ensure proper consideration of important efficiency trends. Second, models should be shared and inter-model comparisons should be conducted so that analysts can develop best practices, increasing the confidence in model outputs. Third, analysts must work together to develop methods for modeling emerging trends, such as AI, the rollout of 5G, and increased edge computing, giving policy makers early insights into their possible energy use implications. Fourth, more reliable data should be developed and shared openly for the Asia Pacific region, and for China in particular, where data center demand is growing rapidly. Lastly, better methods are needed for prospective analysis of next-generation computing, storage, and heat-removal technologies for accelerating investments in technologies that might avert future energy use growth.

References

Air Transport Action Group (2020). Facts And Figures. Https://Www.Atag.Org/Facts-Figures.Html (Https://Www.Atag.Org/Facts-Figures.Html)

Andrae, A.S. And Edler, T. (2015). On Global Electricity Usage Of Communication Technology: Trends To 2030. Challenges, 6(1), Pp.117-157.

Andrae, Anders. (2017). Total Consumer Power Consumption Forecast. Https://Www.Researchgate.Net/Publication/320225452_Total_Consumer_Power_Consumption_Forecast (Https://Www.Researchgate.Net/Publication/320225452_Total_Consumer_Power_Consumption_Forecast)

Bawdy, T. (2016). Global Warming: Data Centres To Consume Three Times As Much Energy In Next Decade, Experts Warn. The Independent. Saturday 23 January 2016.

Belkhir, L. And Elmeligi, A. (2018). Assessing ICT Global Emissions Footprint: Trends To 2040 & Recommendations. Journal Of Cleaner Production, 177, Pp.448-463.

Jones N. (2018). How To Stop Data Centres From Gobbling Up The World's Electricity. Nature. 2018 Sep;561(7722):163-166. Doi: 10.1038/D41586-018-06610-Y.

Kamiya, George (2020). Factcheck: What Is The Carbon Footprint Of Streaming Video On Netflix? Carbon Brief, February 25th. Https://Www.Carbonbrief.Org/Factcheck-What-Is-The-Carbon-Footprint-Of-Streaming-Video-On-Netflix? (Https://Www.Carbonbrief.Org/Factcheck-What-Is-The-Carbon-Footprint-Of-Streaming-Video-On-Netflix?)

Koomey, Jonathan (2011). "Growth In Data Center Electricity Use 2005 To 2010." A Report By Analytical Press, Completed At The Request Of The New York Times 9 (2011): 161.

Masanet, Eric, Arman Shehabi, Nuoa Lei, Sarah Smith, And Jonathan Koomey. "Recalibrating Global Data Center Energy-Use Estimates." Science 367, No. 6481 (2020): 984-986.

Pearce, F. (2018). "Energy Hogs: Can World's Huge Data Centers Be Made More Efficient?" Yale Environment 360, April 3rd, Https://E360.Yale.Edu/Features/Energy-Hogs-Can-Huge-Data-Centers-Be-Made-More-Efficient (Https://E360.Yale.Edu/Features/Energy-Hogs-Can-Huge-Data-Centers-Be-Made-More-Efficient)

Shehabi, Arman, Et Al. (2016). United States Data Center Energy Usage Report. No. LBNL-1005775. Lawrence Berkeley National Lab. (LBNL), Berkeley, CA (United States).

U.S. Department Of Energy (2020). Annual Energy Outlook 2020. Https://Www.Eia.Gov/Outlooks/Aeo/ (Https://Www.Eia.Gov/Outlooks/Aeo/)

U.S. Energy Information Administration (2020). How Much Carbon Dioxide Is Produced Per Kilowatthour Of U.S. Electricity Generation? Https://Www.Eia.Gov/Tools/Faqs/Faq.Php?Id=74&T=11 (Https://Www.Eia.Gov/Tools/Faqs/Faq.Php?Id=74&T=11)

Published on March 17, 2020 (https://energyinnovation.org/2020/03/17/)
Posted in Guest Articles (https://energyinnovation.org/category/guest-articles/) | Tagged Energy efficiency (https://energyinnovation.org/tag/energy-efficiency/), Power Sector Transformation (https://energyinnovation.org/tag/power-sector-transformation/), Emissions (https://energyinnovation.org/tag/emissions/)