Data Centers for Energy Efficiency

Presentation by Nia Starr

May 4, 2022

The goal of this presentation is to examine best practices for data center energy operations in hopes of establishing industry standards for better climate futures.

Agenda

Data Center Overview

Data Usage
Definitions
Industry Commitments

Energy Efficiency

Hardware Innovations
Cooling Systems
Renewable Energy Sources

Data Center Futures

Establishing a Standard
Green Infrastructure Funding
Policymaking

Data Usage

Current Global Electricity Consumption from Data Centers

1%

Projected Global Electricity Consumption from Data Centers by 2030

3-8%

2%

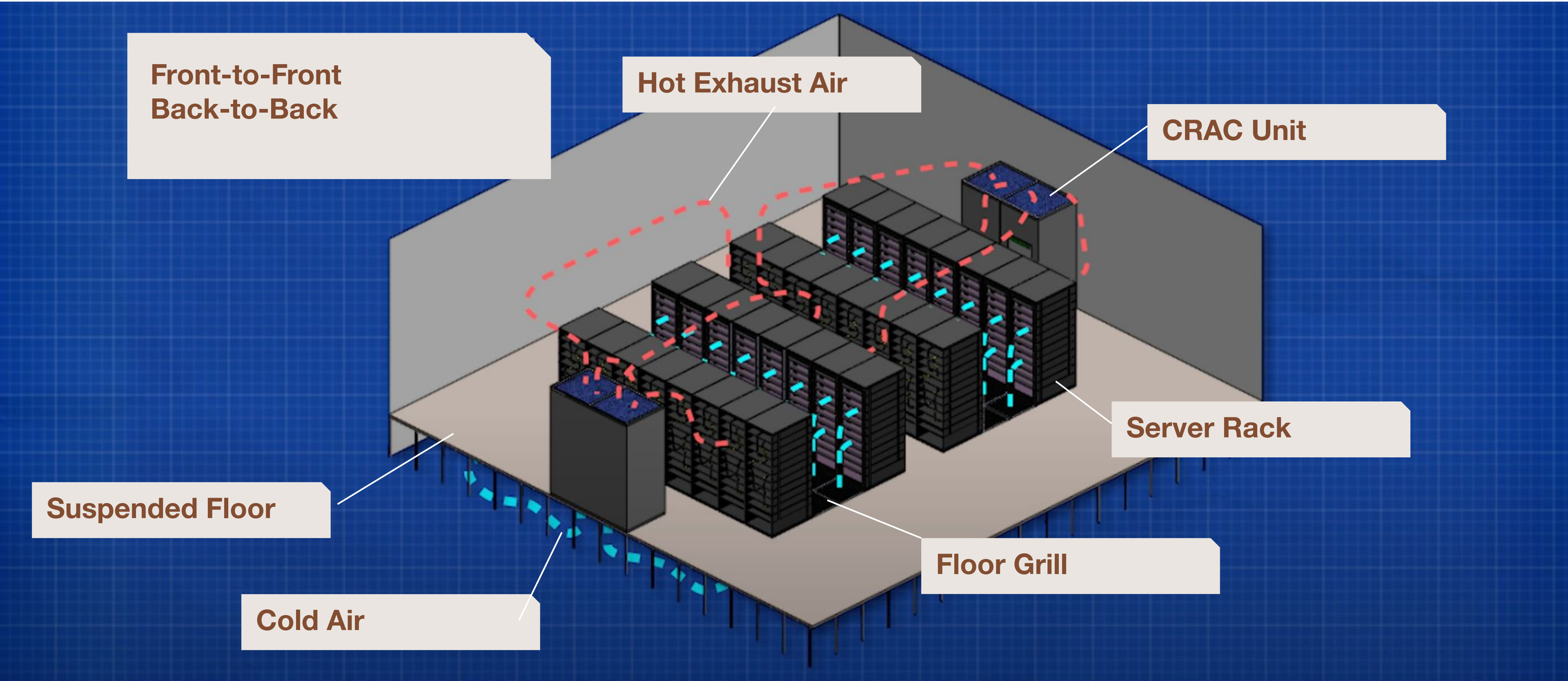
Current Global GHG Emissions from Data Centers

14%

Global GHG Emissions from Data Centers by 2040

0.3g CO_2e for a spam email
4g (0.14oz) CO_2e for a regular email
50g (1.7oz) CO_2e for an email with a photo/attachment
Lancaster University

Data Center Design



Data Center Types

Co-located

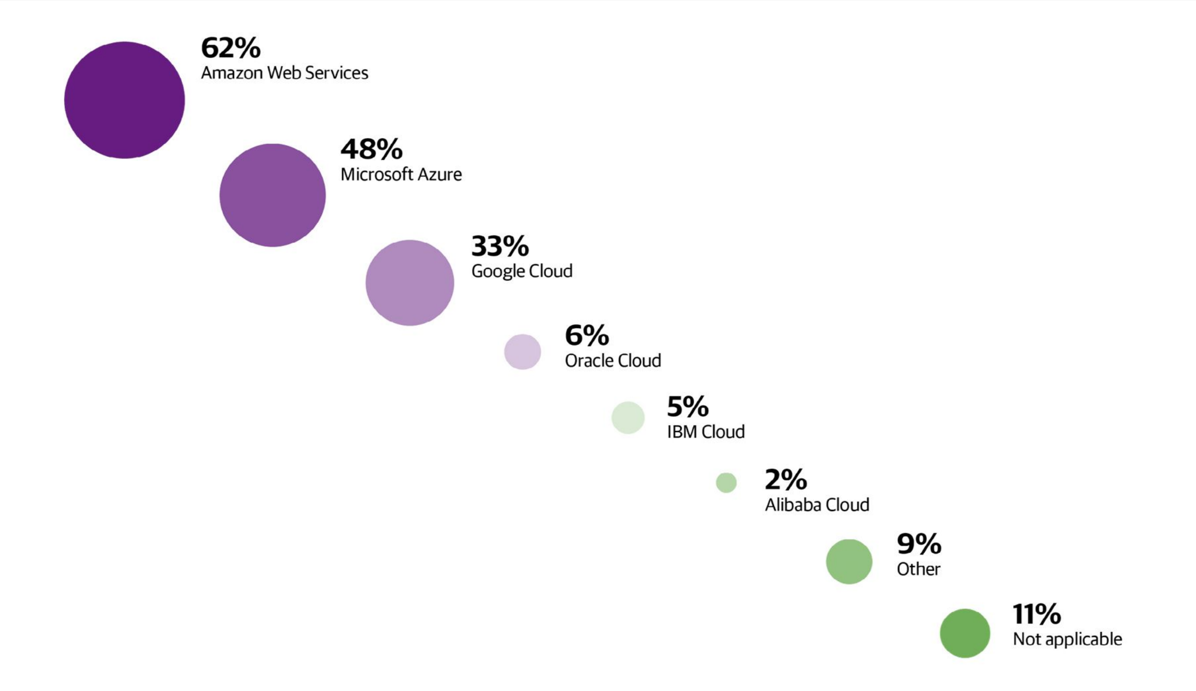
Servers, storage and network elements are owned, physically accessible

Cloud

Virtual servers utilized, high security, better carbon footprint

*On-Premise and Hyperscale are other types, but we will focus on these two main industry standards

Cloud Infrastructure



“The Cloud in 2021: Adoption Continues”

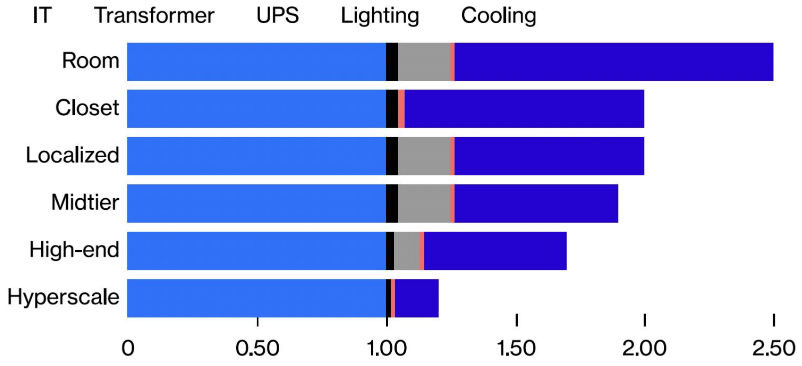
O’Reilly Media, Inc.

Cloud Infrastructure

Amazon Web Services: 84 Availability Zones within 26 geographic regions around the world, with announced plans for 24 more Availability Zones and 8 more AWS Regions in Australia, Canada, India, Israel, New Zealand, Spain, Switzerland, and United Arab Emirates (UAE).

Power Usage Effectiveness

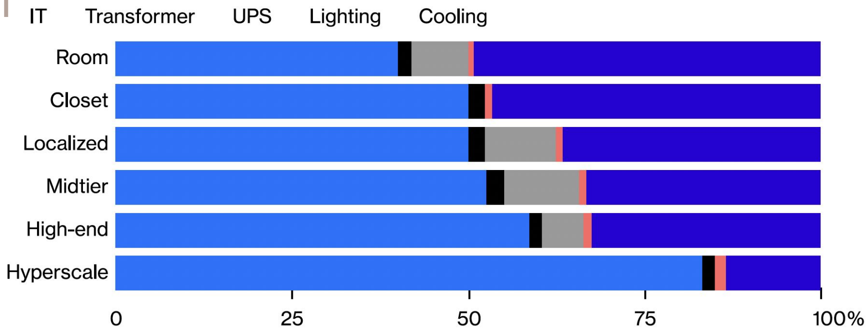
PUE Score



Source: Lawrence Berkeley National Laboratory

By Component

Server power consumption by component, % of total



Source: Lawrence Berkeley National Laboratory

Thank you.

Nia Starr
Professor Richard Leigh
Urban Energy Management
May 4, 2022

Updating Data Storage Infrastructure as an Energy Efficiency Measure

As emerging digital technologies scale up worldwide, understanding the effects that data storage centers transmit to the environment has become increasingly more important. Government and industry standards still have yet to be established around data processing systems to limit future emissions growth and energy demand from this sector. In assessing the different types of data centers, this paper aims to highlight how the cloud model of data center design can lead as an example for how traditional data center models can be transformed in terms of energy efficiency for computing processes.

Current federal legislation concerning data centers in the United States of America was updated in 2020 with the Better Energy Storage Technology (BEST) Act, an amendment to the U.S. Energy Storage Competitiveness Act of 2007.[1] This amendment calls for a "research, development, and demonstration program for grid-scale energy storage systems"[1] and is the first domestic policy piece to address data center practices as a key component in energy efficient futures. Current information on data centers in the U.S. is gathered from extrapolating key findings from market indicators, but the new 2020 policy positions the federal government to update its systems and develop proprietary research for energy efficient models. The policy speaks to a larger trend in the industry and codifies the need for energy efficiency as it relates to data centers for the first time. The policy only speaks to a research and development phase with an implementation plan still yet to be drafted. There is concern that the rapidly growing demand for data will outpace the rate of policymaking, and action is needed presently to set standards for energy use and climate implications.

A data center is a physical facility that houses digital infrastructure meant for processing computing workloads. The typical components of a data center include server racks (computer hardware that provide functions for computer programs) installed atop a suspended floor. Processing data consumes vast amounts of energy,

[1] "H.R. 2986 — 116th Congress: BEST Act." 2019. Retrieved May 2, 2022.
https://www.congress.gov/bill/116th-congress/house-bill/2986/text

and therefore, heat. To account for heat, Computer Room Air Conditioner (CRAC) Units are placed around the perimeter of the room in order to regulate air flow. CRAC Units take in hot exhaust air from server racks, cool it, and then push cold air into the void beneath the suspended floor. The void underneath the floor stores that cool air and circulates it back into the server room through floor grills strategically located in front of the server racks. Fans installed on server racks move air across the circuit boards to cool them. For best practices, vertical server racks are typically arranged face-to-face and back-to-back to ensure the hot exhaust is most efficiently pooled into CRAC units, and not pushed back over the hot hardware.

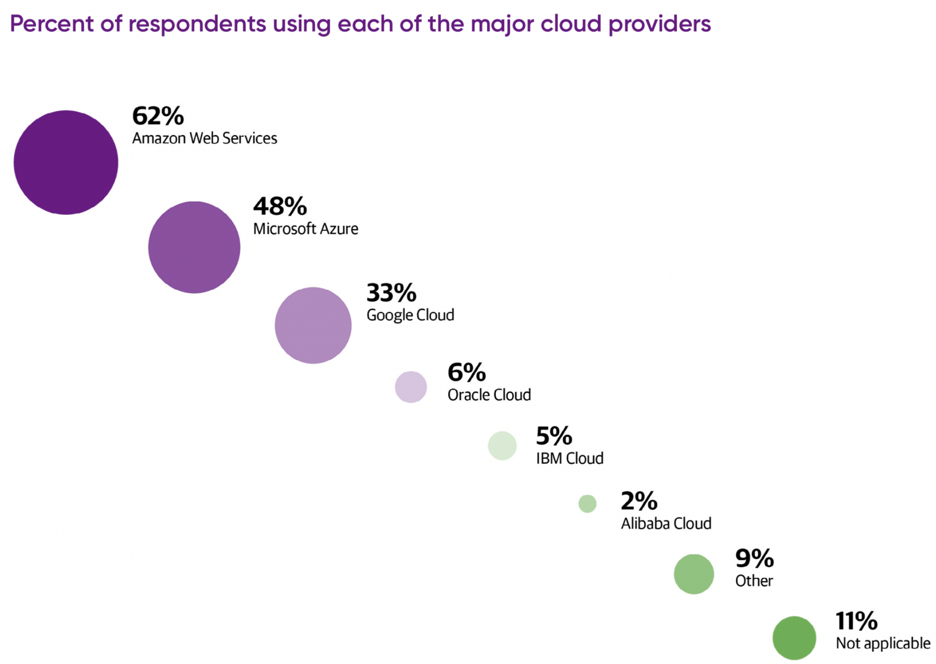
This description of data center design applies to the main types of data centers in use today: co-located, cloud, and hyperscale (a form of cloud data centers, but at an even greater scale). Because the purpose of this paper is to identify better design opportunities of these facilities for the sake of energy efficiency, it is important to outline that the base inventory of data center components is essentially the same. Where better models diverge and pull ahead design-wise is in relation to three criteria: computing workload utility, cooling system management, and energy source. This is where cloud data centers excel in comparison to co-located models.

What is the cloud? The cloud can be defined as the use of virtual servers rather than the use of physical servers to process data for computing processes. Traditionally, computing workloads are assigned to one server to complete, but with innovation in conductive hardware materials, a computing command by an end user can be spread across multiple servers virtually and the computing workload burden eased.[4] This results in faster computation with higher efficiency because of distributed work capacity across multiple servers. In other words, multiple computers within a cloud server stack (or in some cases, locations that are spread across the world) can process more data than traditional server stacks, and at a faster rate. The cloud facilities are owned and operated by just a few large companies who dominate the market.[2] The most efficient top three companies (Amazon Web Services, Microsoft, and Google) are often referred to as “hyperscale” environments because of the grand scale at which they operate and their constant innovation in improving cloud systems (Figure 1). These “Top 3” companies have configured their servers for maximum productivity with high utilization rates. Their hyperscale facilities are highly efficient, employing fewer servers than what is required to provide the same services in traditional, co-located data centers.[3]

---

[2] Loukides, Mike. (2021). “The Cloud in 2021: Adoption Continues” O’Reilly Media, Inc. Accessed May 1, 2022. (Link)
[3] Shehabi, Arman. Smith, S. Sartor, D. Brown, R. Herrlin, M. Koomey, J. Masanet, E. Horner, N. Azevedo, I. Lintner, W. (June 2016). "United States Data Center Energy Usage Report". Lawrence Berkeley National Laboratory. Retrieved May 1, 2022. (Link)

Figure 1



Unlike co-located servers which are housed in standalone facilities, cloud servers can be deployed almost immediately and at a lower initial cost. Companies no longer need to own and maintain their own hardware with virtual servers–they can opt to share resources for computing through cloud servers. The incentives in converting to cloud operations includes lower overhead management, higher security, and built-in customer service. The increased demand for cloud services has led to a steep rise in cloud data center activity with companies migrating their computing workloads to cloud infrastructure at unprecedented rates. As of 2021, cloud computing has been adopted by 90% of all enterprises currently in existence.[5] That overall number is poised to increase in the coming years with the intensity of companies' cloud usage also set to increase.

Currently, one percent of all global energy consumption (200-250 TWh as of 2020[4]) is consumed by data center electricity demand.[3] This one percent accounts for

[3] 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 (2011): 161.
[4] IEA analysis based on Masanet et al. (2020) and Malmodin (2020). (Link)
[5] Cisco, "Cisco Global Cloud Index: Forecast and methodology, 2016–2021 white paper" (Cisco, document 1513879861264127, 2018).
[6] Loukides, Mike. (2021). "The Cloud in 2021: Adoption Continues" O'Reilly Media, Inc. Accessed May 4, 2022. https://get.oreilly.com/ind_the-cloud-in-2021-adoption-continues.html
[7] 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

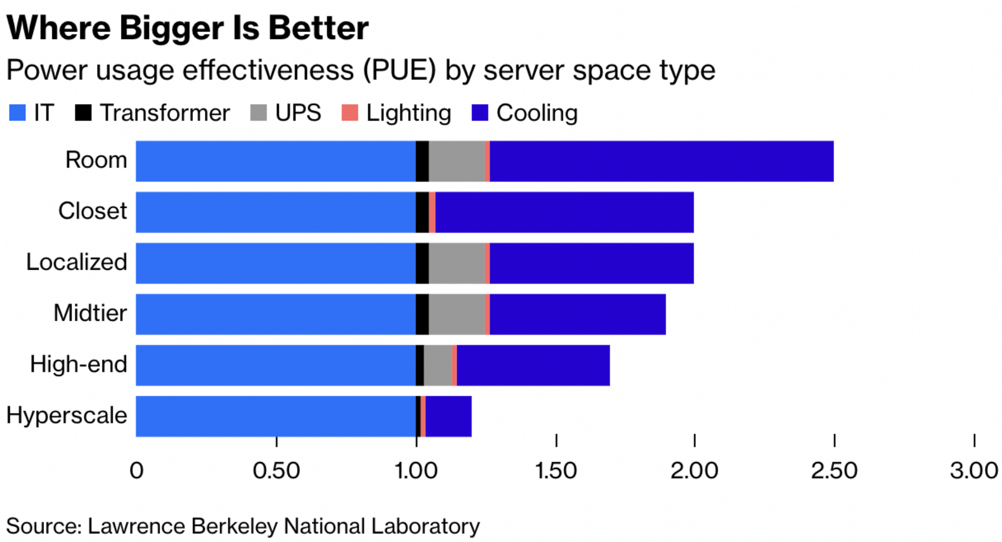
all of the servers that companies are using to power their websites, daily business operation applications, along with consumers rely on everyday for online browsing. Everytime a video is streamed, information uploaded to the internet, or an online application is fired up, web servers are called upon to process this workload at a data center. With new capabilities increasingly being added to the Internet of Things (IoT)—augmented reality (AR) services, automated vehicles, or cryptocurrency transactions to name a few examples—a rehabilitation of core data infrastructure is needed if humanity hopes to curb its impact on climate change. By 2030, that one percent of energy usage is projected to grow to eight percent of all global electricity consumed by data centers.[8] All of these internet activities sum up to two percent of the world's greenhouse gas (GHG) emissions with projections of increase to fourteen percent by 2040.[13]

In the data center industry, there is a standard measure of energy efficiency called the Power Usage Effectiveness (PUE) ratio. The PUE standard attributes a score to data servers by dividing the total energy used by the amount of energy consumed specifically for computational processes. The ideal rating would be a PUE score of 1, where one hundred percent of electricity consumed by data servers is attributed to computational processes.[7] This is not typically the case because a large proportion of electricity is needed for the network of cooling systems that data centers employ. Heat must constantly be removed from server rooms inside data centers, otherwise the circuit boards will overheat and risk destruction. Removing all of this heat consumes additional energy on top of computing, therefore increasing PUE scores to a number between 1 and 2. In observing the data center landscape, services utilizing the cloud model are best equipped to efficiently manage this heat (Figure 2) because of the way their facilities are designed.

This is how they're able to reduce their Power Usage Effectiveness score to almost 1.

[8] Andrae, A. & Edler, T. "On Global Electricity Usage of Communication Technology: Trends to 2030" 117–157 (2015). (Link)

Figure 2



While cloud computing is efficient, it still requires massive needs for electricity generation and its energy intensity is continuously slated to increase year over year as the IoT absorbs new functions in society.[8] As "Big Tech" gains public scrutiny for carbon emissions related to energy intensive operations, the move to increase power purchase agreements and scaling vertical renewable energy projects have become a key action to reduce their environmental impact and improve the reputations of their brands. To create guardrails against electrical power volatility hyperscale data companies invest heavily in renewable energy tech to power their data centers and reduce their PUE score. In the past couple of years, Google (12 TWh in 2019), Apple (1.7 TWh in FY2020) and Facebook (7 TWh in 2020)[9] matched one hundred percent of their electricity consumption through purchasing or generating renewable energy. Amazon Web Services launched an initiative that would allow the company to monitor its renewable energy assets (with AWS data centers located nearby) across 30 different countries in near-real time.[10] The alternative to using renewable energy has been to

---

4

[9] IEA (2021), Data Centres and Data Transmission Networks, IEA, Paris. (Link)
[10] Li, Xin. Singh, A. Thompson, G. Beer, J. (October 2021). "How Amazon Achieves Near-Real-Time Renewable Energy Plant Monitoring to Optimize Performance Using AWS" AWS for Industries. Accessed May 2, 2022. (Link)
[11] Data Center Consolidation Initiative. September 2011. United States General Services Administration. (Link)
[12] Project Natick. Microsoft. (Link)
[13] Assessing ICT global emissions footprint: Trends to 2040 & recommendations Belkhir L., Elmeligi A. (2018) Journal of Cleaner Production, 177 , pp. 448-463. (Link)

burn fossil fuels to support cooling efforts and hence, the large global footprint of greenhouse gas emissions attributed to the internet business.

The strategic location of data centers has become a method for increasing energy efficiency. Some data centers are purposefully built in cold parts of the world to make use of naturally cold air. Both Google and Microsoft have built hubs in Finland, and Meta has deployed data centers in Denmark and Sweden for both the natural temperature and renewable energy capabilities.[7] Others have even been trialed underwater to reduce the energy consumption of cooling, like Microsoft's undersea Project Natick in Scotland's Orkney Islands.[12]

On top of innovating hardware and switching to low-carbon energy resources, cloud facilities have also begun to reform the way energy is operationalized. The traditional model for regulating temperature inside of data centers sets the thermostat to 72 degrees fahrenheit, a temperature status that optimizes human comfort. Cloud facilities have recalibrated temperature comfortability to server operations at 80 degrees fahrenheit, cutting down on energy intensity for cooling capabilities in alignment with the U.S. General Services Administration recommendations.[10]

All of these undertakings by large internet companies with highly efficient cloud facilities have served as a major catalyst for the rest of the data center industry to adopt low-carbon measures. There is a great opportunity to take these models and apply them to traditional data center models in a hyperscale shift if governments hope to achieve carbon neutrality in the internet sector. Identifying funding for renewable energy products has historically been difficult, but reports of Big Tech companies account for half of global corporate investments in renewable energy (over 15 GW).[9] Governments should take advantage of this vast network of global data centers to fund green infrastructure projects while holding the carbon-intensive internet sector accountable. As cloud services increase in the next couple of decades, so too will the demand for more energy. Big technology companies like the top three cloud computing giants, are seeking ways to increase renewable energy production and store it so that clean energy can be truly attributed through their operations, rather than utilizing the power payment agreement system and using power from a fossil-fuel electrical grid. The U.S. government should establish industry standards as a next step for the BEST Act of 2020, with cloud infrastructure and hyperscale business models serving as a beacon.