/thoughtworks

## Why green cloud optimization is profitable for you and the planet

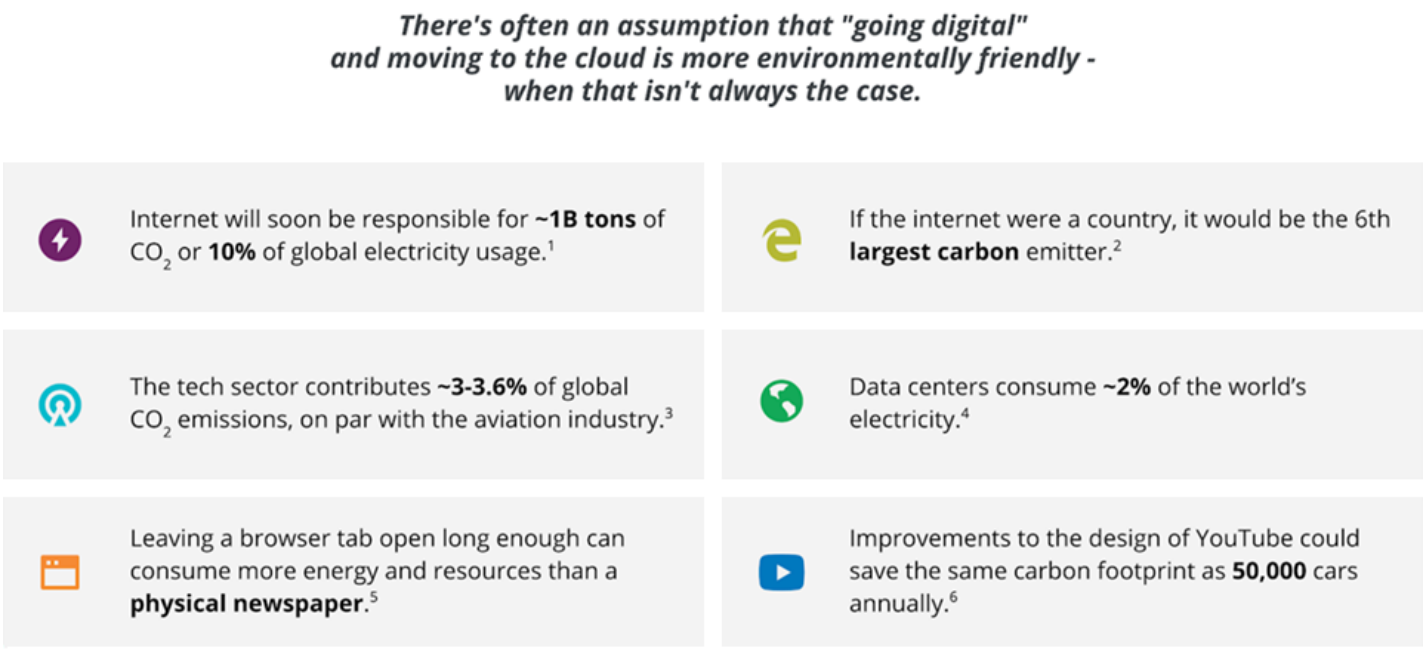
September 29, 2020

**The green cloud challenge**

Enterprise spending on public cloud shows no signs of slowing with some types of cloud spending growing by 92% per year. From a sustainability perspective, that might seem like a good thing — after all, cloud services tend to be more energy efficient than traditional data centers. But as enterprises embrace cloud to support user growth and revenue generation, few stop to consider the sustainability implications beyond choosing a cloud provider. Thankfully there are some simple 'green cloud optimizations' your IT organization can do, which are both good for the planet and your business.

To appreciate the need for greener **cloud computing**, it helps to consider some statistics. The amount of energy needed to power data centers roughly doubles every four years, and even in a best case scenario, the ICT industry is likely to account for 8% of total electricity demand by 2030 — a 15-fold increase from 2010. And when it comes to global greenhouse gas emissions (GHG or CO2e, hereafter referred to as carbon emissions), the ICT industry has contributed 2–6% consistently since 2007 — on par with the aviation industry.

The good news here is that despite the exponential growth of ICT and cloud adoption, the relative energy use and carbon emissions have remained roughly the same; due to significant efficiency gains from hyperscale data centers used by major public cloud providers. But that isn't the full story. Cloud and data centers are increasingly becoming a bigger piece of the ICT industry's carbon emissions pie. To meet the goals of the Paris Climate Agreement, the industry will need to reduce carbon emissions by 45% in the next 10 years. So even with considerable efficiency gains in the cloud, there is still much more to be done.




**The business opportunity: A case in point**

Why would a company embrace green cloud optimization? Etsy is a good example. For the first 11 years of its history, Etsy maintained its own servers in a co-located data center. As it grew, purchasing, installing and provisioning servers became time-consuming and hard to plan. Etsy found an astronomical difference between cloud and data center provisioning speed: cloud was an astonishing two million percent faster. Why? Etsy estimated it could take months to provision 150 servers in its data center, but only four minutes to do the same in the cloud.

With the cloud, organizations can scale capacity up and down in response to fluctuating demand. For example, you can run workloads to rotate around the globe, so they fit with local time zones, rather than running 24/7. Shifting towards serverless computing, such as GCP Cloud Functions or AWS Lambda, lets development teams run code without even provisioning or managing servers. In that way, you only pay for the compute time you consume.

Moving to a flexible cloud-based infrastructure enabled Etsy to reduce major idle time and associated energy consumption. As Etsy was transitioning from co-located data centers to the cloud, its combined energy consumption decreased by an estimated 13% (from 7330 MWh in 2018 to 6376 MWh in 2019) at the same time as their business grew.

There are algorithmic efficiency opportunities too. Machine learning, such as that used by Etsy, are targets for algorithmic efficiency, since these algorithms are particularly computationally intensive. While these opportunities exist within on-premise data centers too, cloud has given rise to easily accessible AI and ML services and proliferated their mainstream adoption. Researchers and practitioners are aware of the problems and costs associated with inefficient algorithms. They recommend comparing different models to perform a cost-benefit-accuracy analysis. Changes to these energy-demanding algorithms, can then produce measurably more efficient models and influence hardware choices.

One study of natural language processing models illustrates the extent of the problem. Researchers showed that training one NLP model produced carbon emissions equivalent to 125 roundtrip flights from New York to Beijing.

With the amount of computing power used for deep learning doubling roughly every quarter, when it comes to your AI/ML applications, efficiency needs to be as important a consideration as accuracy. We're beginning to see some tech teams consider notifying developers about carbon emissions associated with compute intensive ML workloads in order to reduce costs and switch those workloads to run in geographic regions with more renewable energy on the grid.

When confronted with this energy challenge, organizations should firstly aim to shift to hyperscale data centers and optimize their cloud usage to reduce overall energy use. And secondly, they can green their remaining energy usage by switching to regions with more direct local renewables and running workloads at a time of day when renewables are available.

**Why now?**

With traditional on-premise data centers, organizations know how much energy they are using from their energy bill. In the cloud, that information isn't readily available. Although hyperscale data centers are more efficient than on premise, their on-demand model for rapidly scaling up and their readily available compute intensive services can mean that it's easy for organizations to unwittingly consume more resources.

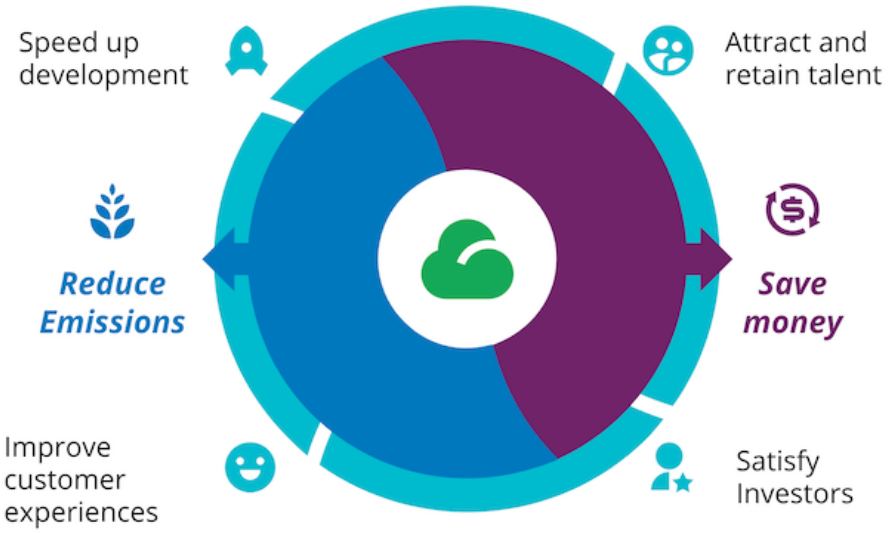
On premise infrastructure also means organizations need excess capacity to handle peak loads. In the cloud, any resources that you don't need can easily be shed. But not every organization is seizing that opportunity. A lift and shift strategy could result in inefficient software and rack up higher costs. And given the ease with which new servers can be added, many organizations have lost control of their cloud estate and are seeing costs balloon, especially if they don't pay attention or precisely forecast their computing needs. With the storage and transfer of data increasing exponentially (which we have seen for

the last three decades) — and the next five years expected to see a doubling in cloud demand and increase in computationally and energy intensive technologies like AI, blockchain — it's up to all of us to ensure our use of technology becomes more sustainable.

Unfortunately, many organizations think that the efficiencies of moving to the cloud are "enough"; optimizing and reducing their cloud costs and footprint even further is a blind spot. Or, they're so focused on shipping products and on the customer-facing parts of their business, that they don't have enough resources to dedicate towards the numerous internal efficiency and optimization opportunities.

But today's stakeholders, from customers to employees to investors, expect more when it comes to organization's sustainability commitments -- as we will see below -- and the good news is that there is a strong business case for reducing your cloud carbon footprint.

**The benefits of green cloud optimization for business**




The good news is that reducing your cloud carbon footprint is win-win for all stakeholders.

**1. Save money by reducing costs.** The more efficient your cloud usage, the less you'll spend. In 2018, Gartner estimated that organizations that didn't optimize their public cloud usage would spend 40% more by 2020. We've seen this first hand. At an independent news company, Thoughtworks optimized the infrastructure to reduce the number of IBM cloud virtual machines needed to support web applications and allow for better caching. We also increased performance by implementing responsive and dynamic interfaces that refreshed and loaded UI components only as needed, rather than the whole page. Our work reduced the compute resources needed by 50%, which resulted in 25–30% overall infrastructure cost savings.

**2. Meet sustainability goals by reducing your carbon footprint.** Reduce your number of servers and you'll save electricity and cut carbon emissions; switch to cloud providers and services that use renewable energy and you'll cut your carbon footprint even further. Companies like Spotify and SiteGround have switched to the carbon neutral Google Cloud Platform (GCP) in recent years to reduce their carbon footprint. Global retailer Forever New includes its use of containers and virtualization as part of its cloud strategy to reduce overall power consumption and carbon footprint. And MapBox configured its use of AWS regions to those that are covered by renewable energy, in order to achieve its goal of being carbon neutral.

**3. Improve customer experiences with faster, more responsive applications.** Optimizing application performance — with caching, image size, data transferred and so on — speeds up page load times and improves the overall customer experience of your digital channels. Similarly for internal applications, this can improve your employees' efficiency by reducing the time it takes to complete tasks. By reducing the amount of data transferred and distance it travels, this reduces your cloud spend, energy usage and carbon footprint. The paradigm of edge computing which moves computing from central servers closer to the end user further reduces energy and emissions, as well as delivering faster services. Developers can improve algorithm and query performance using services like AWS CodeGuru that let developers know which lines of code need the most resources — and therefore cost the most. The developers can then explore ways of improving that code.

Thoughtworks modernized e-commerce systems at a major retailer, improving both site design and functionality, as well as moving them to the cloud. This involved migrating legacy features to a newer software platform and migrating its infrastructure to the cloud. The resulting site is four and a half times faster, delivers 20% higher revenue and increased conversation rates over the old site by 50%.

And of course communicating your efforts to customers has additional value: 81% of global consumers feel strongly that companies should help improve the environment. Akamai, one of the world's largest cloud content delivery networks reported a 30–40% increase in requests for information about sustainability between 2018-2019. And a 2019 study found that 50% of growth in consumer packaged goods in the last five years came from sustainable products, whose sales grew almost six times faster than non-sustainable products. So whether companies are consumer-facing or B2B, customers are increasingly considering environmental performance as part of their decision making criteria.

**4. Speed up development times by improving developer efficiency.** Optimizing build pipelines and code so that it is faster and more efficient — such as by reducing unnecessary dependencies that consume extra storage or compute resources — helps to reduce costs, reduce energy usage and make your developers' lives easier so that they can focus on delivering value.

At one online travel booking company we worked with, development teams had routinely left code from A/B experiments in their codebase after running tests and simply switched off the test or feature. After thousands of experiments over years, this added up to a significant cost and performance lag. The team identified that removing this excess code would result in notable speed and cost savings, not to mention reduce their overall energy use and carbon footprint.

Where possible, tasks should be automated because it allows them to be continuously improved and optimized. One simple example we've seen when working with US organizations to introduce standardized AWS CloudFormation templates that automate infrastructure setup: teams can ensure their default AWS region is set to US-West (Oregon), rather than US-East (Virginia). This would mean all new infrastructure by default uses a greater proportion of direct renewable energy and reduces an organization's cloud carbon footprint, without anyone having to think twice about it.

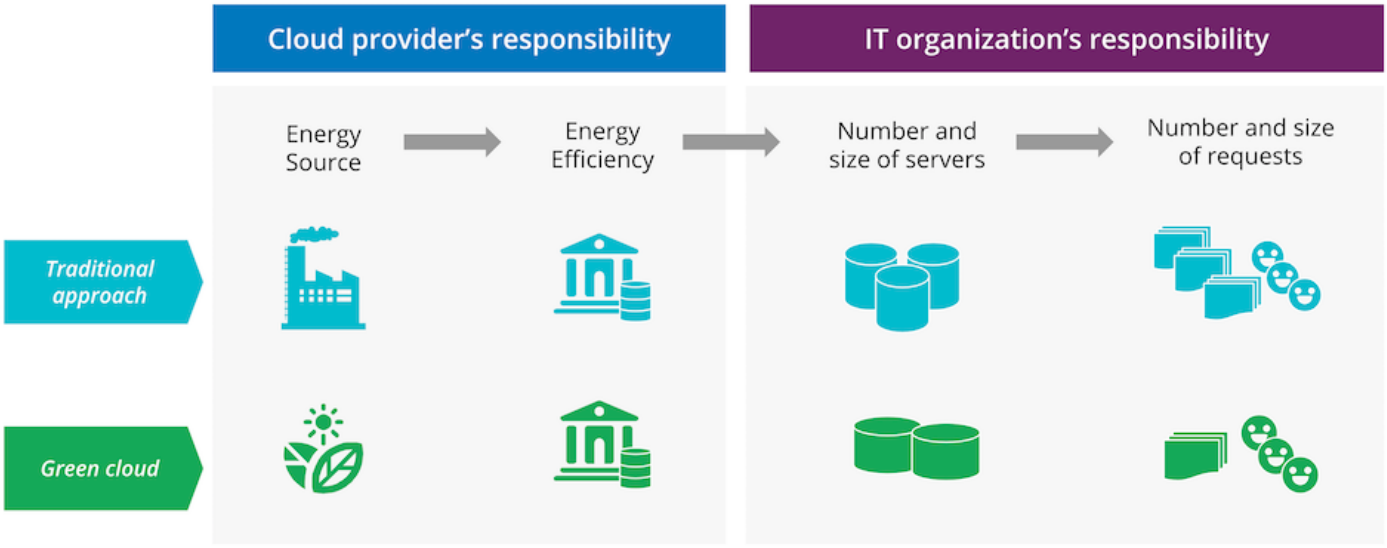
**5. Attract and retain talent with environmental initiatives**. Around 70% of employees said they were more likely to work at a company with a strong environmental agenda, and more likely to stay there long term. In recent years, especially within the tech sector, many employees have been actively demanding greater action from their companies on climate change. Thousands of companies showed their support for the 2019 Global Climate Strikes and almost one fifth of Fortune 500 companies had committed to the Paris Climate Agreement emissions reductions targets in 2018.

**6. Satisfy investors by disclosing carbon footprint information.** Over the past 10 years, sustainable investment funds of companies with good environmental, social and corporate governance (ESG) practices have been shown to have superior long term stock performance to conventional funds. In 2020 investment firm Blackrock, the world's largest asset manager, announced in its annual letter to CEOs that climate risk is financial investment risk and they will look at sustainability or ESG criteria for every investment. Thus more sustainable companies have greater access to capital when needed. Public companies are also facing increasing regulatory and compliance requirements.

**The four pillars of green cloud optimization**

So how can IT organizations ensure a greener cloud approach? One common misconception is that it is solely a matter of choosing the best cloud provider. While this is partially the case, it's not the whole story. There is a lot that tech or IT organizations are responsible for during the development lifecycle too. Your digital actions have an impact: code + data = energy = carbon emissions.

Green cloud computing at its best is using a datacenter fed by local renewable energy that runs software on its infrastructure that is designed and optimized to minimize energy consumption (and costs).




A green cloud has many critical characteristics. There are design, operational and energy consumption considerations, for both physical infrastructure and the virtual software systems that run on it. We delineate and differentiate these requirements into our four pillars of green cloud optimizations.

The first two pillars are primarily under the cloud provider's control of their physical infrastructure, and IT organizations can choose to use one or more cloud providers and switch locations within a provider.

**1. Energy source.** This means using cloud infrastructure powered by renewable energy rather than fossil fuels — or use Renewable Energy Credits (RECs) to "neutralize" fossil fuel emissions by matching out each "dirty" megawatt of electricity a datacenter or customer uses with a "clean" megawatt represented by an REC. However, there is work being done to make data centers run directly off of wind and solar power, leveraging large battery storage, or even towards Pausable Data Centers, that shift workloads to different times of the day to take advantage of weather patterns.

**2. Energy efficiency.** Cloud infrastructure that has high energy efficiency when measured through PUE (power usage effectiveness). Typically, that's achieved through innovative approaches to cooling. For example, Microsoft has built some of its data centers underwater and Google has data centers cooled by water pipes which are less energy intensive than traditional air conditioning. Vigilant provides smarter AI-driven cooling that adjusts depending on the server workloads and corresponding heat generated in the room.

The second two pillars are primarily under the cloud customer IT organization's control in how they configure and use the cloud, although of course cloud providers can also optimize their services to perform better too.

**3. Number and size of servers.** Servers are what physically consume energy; by reducing the number of servers you use, you reduce your carbon footprint. Organizations can optimize the performance of their applications and reduce unnecessary storage, but it will still have the same cost and energy usage unless they also reduce the number or size of servers on which their tech estate is running.

**4. Number and size of requests.** Cloud services generate significant network traffic, thanks to the requests and responses made. And while the likes of Netflix or YouTube aren't completely responsible for how many hours of video users watch, the shift towards a sustainable economy means that end users' energy use of digital products increasingly need to be considered as part of "Scope 3" emissions in an organization's carbon reduction goals. To minimize traffic, you can firstly use or optimize caching so there is less distance to transfer the data from a local edge server to the user, and secondly reduce the amount of data being transferred by dynamically reloading only the components that are needed and prioritizing mobile-first user experiences since smaller devices and screen sizes require less data.

**Getting started**

For any tech or IT organization, there are simple steps you can take to reduce your carbon footprint and optimize your cloud usage and costs.

**1. Choose a greener cloud provider or region.** Choose cloud providers, cloud services and SaaS tools that use renewable energy rather than ones running on fossil fuels. The Green Web Foundation provides a Green Host Directory where you can look up renewable energy powered providers, as well as two different open source tools you can incorporate into build pipelines to reduce fossil fuel consumption: Greenhouse and the Sustainable web plugin. One non-profit we worked with chose to host their application with Data Center Light, due to it being 100% renewable energy powered from hydropower and solar power in the Swiss Alps, and it also helped that it follows EU privacy laws, which are stricter than the US.

Microsoft, including Azure, matches 100% of its global annual energy use with renewable energy credits (RECs) and will shift to use direct local renewable energy within the next five years. Google, including GCP, also matches 100% of its annual energy use with clean power, and is working towards matching it on a 24/7 basis with local clean energy so it has carbon-free operations in the next 10 years. Amazon plans to match 100% of annual energy use with renewable energy within the next five years and be net zero carbon within the next 20 years. AWS's annual energy consumption was matched with over 50% renewable energy two years ago, and matches 100% of its energy use with clean energy in five specific regions, in North America and Europe, out of its 22 global regions.

This illustrates that not just choosing a cloud provider, but selecting a region using direct local renewable energy has an impact too. One client identified their environmental impact could be reduced by changing their GCP region from Belgium to Switzerland (with a higher proportion of renewable energy in their grid due to hydropower), and by shifting workloads to a time of day when renewable energy supply is abundant.

While development teams may not be able to choose 100% direct, local, renewable energy for all cloud and SaaS services you use, it can be factored into their decision making process as another cross-functional requirement along with price, security, performance and so on. That way development teams become aware of the impact and tradeoffs and can make more informed choices. Over time the percentage of direct renewable energy in your digital supply chain should increase. Organizations can also request their cloud and SaaS providers to increase the speed and scope of their transition to renewable energy usage and to provide greater transparency into the location- and market-based carbon footprint of their individual usage.

**2. Cloud optimization.** The majority of cloud optimizations which reduce cost also directly reduce server usage and thus energy use and carbon emissions. Be aware though, this isn't always the case: for instance with contract agreements or bulk/upfront purchase discounts.

Reducing your cloud or SaaS usage requires three things: first, having visibility into your usage, followed by steps to optimize your current setup, and lastly ongoing governance processes to ensure lower costs and energy usage are maintained. We've seen many organizations in which development teams aren't aware of their cloud usage and costs, because it hasn't been properly tracked and made available to them. For example, AWS tags allow organizations to track and allocate costs to different teams, enabling them to understand their individual spend, however we know of many organizations that don't set up tags in the first place and lack access to this data. AWS is now making this feature even more accessible by allowing organizations to set Tag Policies that spans multiple AWS accounts and Organizational Units. Conversely, we've seen organizations where every development team has a cloud spend champion who is notified when their costs are increasing.

Lastly, cloud optimizations will only be successful in reducing cost, server usage and emissions if it becomes a part of the continuous software delivery process. This is thus a management challenge to create a process of continued review and optimization — both for cost and usage patterns.

**3. Carbon measurement and reporting.** Multiple cloud optimization SaaS tools have now appeared in the market that allow companies to track and reduce their cloud spend. We're also starting to see tools emerging that connect cloud spend with energy and carbon emissions metrics. This is helpful for sustainability reporting, such as Azure Sustainability Calculator, Etsy's Cloud Jewels and Green Web Foundation's Green Cost Explorer. While these early approaches are paving the way, the market is yet to mature in this space and so most organizations have to determine their cloud carbon footprint on an individual basis since most cloud and SaaS providers don't share this information.

We predict this may soon change, given the rise of carbon footprint calculations in code such as through standard emissions index APIs, automated emissions data gathering through IoT sensors and renewable energy credits on demand via APIs. We're also working on a cloud carbon footprint estimator building on these early concepts to help advance the wider IT industry in green cloud optimization practices.

Organizations should be connecting their corporate sustainability teams with their IT cloud center of excellence or cost optimization teams to understand their cloud footprint, otherwise they are missing opportunities to reduce their impact.

**Conclusion**

Empowering developers with green cloud optimization tools is best for business: it lowers costs and GHG emissions and improves customer experience and development speeds. So why isn't everyone doing it?

We see two common reasons. The first is that organizations are so focused on shipping products or their customer-facing parts of the business, that they don't have enough resources to dedicate to internal optimization. This is an important consideration and where many organizations already have DevOps teams, platform teams or more recently, cloud centers of excellence: monitoring and reducing cloud costs and carbon emissions might fall under their remit. Or, it could be the remit of a new green cloud optimization squad.

Secondly, this is an emerging best practice. **Cloud-native** development is a relatively new organizational capability and so the subset of developers who are well-versed in cost optimization as well as environmental considerations is even smaller. Thus organizations need to build up this capability. There are benefits from so doing — especially as environmental credentials become a mainstream business requirement for employees, customers and investors.

We suggest creating a small, dedicated green cloud optimization squad as part of a new or closely-aligned existing team that can help implement and spread best practices internally. You can take a look at our Green Cloud Optimization Checklist to get started. Follow these steps to choose greener alternatives, optimize your cloud usage, and score a win-win for all stakeholders. To get a better picture of your cloud emissions and to begin implementing these best practices, try out Cloud Carbon Footprint, our open source solution for measuring and analyzing cloud carbon emissions.

References: **1.** "If the Internet was a country, it would be the 6th largest polluter in terms of CO2 emissions" https://www.sustainablewebmanifesto.com/; **2.** "The internet will soon be responsible for nearly 1 billion tonnes of CO2 or 10% of global electricity usage." https://serving.green/#co2-generation; **3.** The tech sector contributes around 3 - 3.6% of global CO2 emissions which is on par with the aviation industry https://medium.com/@AINowInstitute/ai-and-climate-change-how-theyre-connected-and-what-we-can-do-about-it-6aa8d0f5b32c; **4.** Data centers consume around 2% of world's electricity https://www.forbes.com/sites/ibm/2019/12/09/ibm-tech-trends-to-watch-in-2020--and-beyond/#5da9ae634c1c; **5.** Leaving a browser tab open long enough can consume more energy and resources than a physical newspaper, Frick, Tim. Designing for Sustainability: A Guide to Building Greener Digital Products and Services**6.** Improvements to the design of YouTube could save the same CO2 footprint as 50,000 cars annually https://www.fastcompany.com/90346595/the-internets-youtube-habit-has-the-carbon-footprint-of-a-small-city

*This article was amended on October 5, to include a reference to cloud demand doubling in the next five years.*