451 Research | BLACK & WHITE PAPER

The Carbon Reduction Opportunity of Moving to Amazon Web Services

COMMISSIONED BY

aws

OCTOBER 2019

About this paper

A Black & White paper is a study based on primary research survey data that assesses the market dynamics of a key enterprise technology segment through the lens of the "on the ground" experience and opinions of real practitioners — what they are doing, and why they are doing it.

ABOUT THE AUTHOR

DANIEL BIZO

PRINCIPAL ANALYST, DATACENTER SERVICES & INFRASTRUCTURE

Daniel Bizo is a Senior Analyst for Datacenters and Critical Infrastructure Channel at 451 Research. His research focuses on advanced datacenter design, build and operations, such as prefabricated modular datacenters, highly efficient cooling and integrated facilities and IT management to achieve superior economics.Daniel is also a regular contributor to 451 Research's silicon and systems technology research in the Systems and Software Infrastructure Channel. Dan is also a member of 451 Research's Center of Excellence for Quantum Technologies.

Introduction

Enterprises use public cloud infrastructure for various reasons, but faster time-to-market and the ability to flexibly scale resources to track workload demand are the top two drivers, according to 451 Research's Voice of the Enterprise survey on cloud, hosting and managed services[1]. This is particularly true for organizations further down the line in their digital transformation, which shows enterprises increasingly value the agility and scalability of cloud as they grow more acquainted with it.

Moving workloads to public cloud infrastructure also presents enterprises with the opportunity to dramatically reduce the environmental footprint of their IT operations. Climate change and its effects direct more attention at resource efficiency as a key part of sustainability responsibility, which is of growing importance to businesses. In 2018, 86% of the companies in the S&P 500 index published a sustainability report, up from only 20% in 2011, notes the Governance and Accountability Institute[2].

Enterprises want to be seen as responsible corporate citizens, and many have made sustainability commitments and achieved progress in multiple areas of their operations. Yet even with an emphasis on sustainability, running data centers and IT is not a core competency of most enterprises, many of which lack the expertise and resources to make major investments in infrastructure efficiency. Similarly, most enterprises are not prepared for the effects of climate change (such as extreme weather conditions, drought or floods) on their data center operations.

To estimate the environmental benefits for enterprises moving to its public cloud infrastructure, Amazon Web Services commissioned 451 Research, a technology market research and advisory firm, to conduct a study on the energy and carbon efficiency of enterprise data centers and server infrastructure.

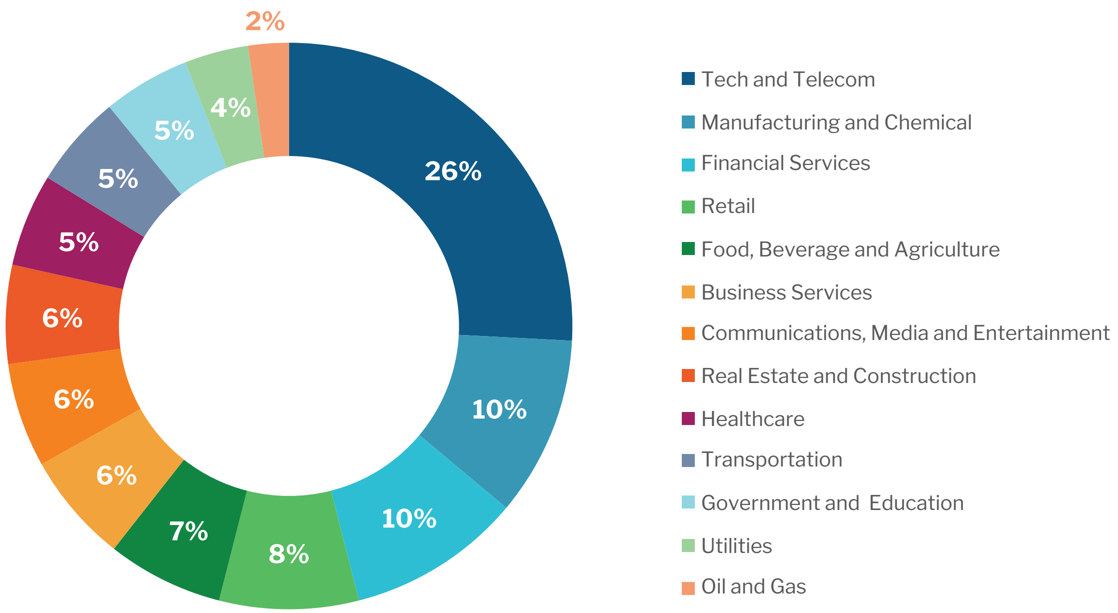
For this study, we focused on US enterprises with revenue between $10m and $1bn, businesses that face IT challenges similar to those of larger enterprises, often with smaller budgets and less IT expertise. Even when enterprises invest in newer infrastructure, they still have to provision for peak demand, and many energy-efficiency measures are only viable at the scale of thousands of servers and when applying advanced design techniques (e.g., efficiency-optimized custom servers, wide temperature bands and indirect evaporative cooling). This means that moving workloads to the cloud can help enterprises steeply reduce energy consumption and their carbon footprint compared to their internal operations.

1. *Voice of the Enterprise: Cloud, Hosting and Managed Services, Workloads and Key Projects - Quarterly Advisory Report, 451 Research, June 28, 2019*
2. *Flash Report, Governance and Accountability Institute, a corporate risk & sustainability consultancy, May 16, 2019*

We surveyed senior stakeholders at over 300 companies – typically with the titles of CIO, CTO, IT director and data center manager – seeking information about their infrastructure operations at their largest US data center location (on-premises or colocation). Participating companies came from a broad cross section of the US economy representing over 20 sectors, such as IT and telecommunications, media and entertainment, financial services, manufacturing and healthcare. We also conducted 10 in-depth interviews to get more insight into the thinking and challenges of enterprise infrastructure operations.

Figure 1: Demographics of surveyed US enterprises by major industry groupings (n=302)

Source: 451 Research



451 Research devised a carbon efficiency model that offers a grid-to-chip view of efficiency. Considering the power intensity of their operations, servers account for the majority of the enterprise infrastructure carbon footprint and are indicative of relative efficiencies. Also, servers run a growing variety of data storage and network services, with the rise of trends like hyperconvergence and software-defined infrastructure. The model includes data center facility overhead as captured by the widely used power-usage effectiveness (PUE) ratio and server energy efficiency. We used survey results, inputs from Amazon Web Services on its operations in the US, and third-party industry data (including from data center design and operations authority Uptime Institute, a 451 Group company) to populate the model.

We did not include the carbon footprint of water or direct emissions of carbon and other greenhouse gases from sources such as emergency power generators in our carbon model. Water does contribute to the environmental footprint of data centers and is a scarce resource in some regions, but its effect on carbon emissions is much lower than electricity usage. Similarly, direct (Scope 1) emissions in data centers have marginal carbon impact compared to indirect (Scope 2) emissions attributed to energy. Future studies may consider these views as well as embodied emissions (Scope 3) in buildings and hardware for a more complete picture.

Executive Summary

The success of public cloud services has resulted in the creation of cloud data center campuses much larger than enterprise sites. While hyperscale campuses have attracted scrutiny for their energy usage, they are much more efficient and offer a workload carbon footprint that is a fraction of what enterprises typically produce in their on-premises or colocation data centers. There are several technical components to this carbon advantage, but what really makes public cloud inherently more efficient is structural. Cloud operators make the entire technical organization work in unison to attain high infrastructure efficiency by design, while the cloud business model of a shared and monetized infrastructure drives server utilization well above what is possible for enterprises.

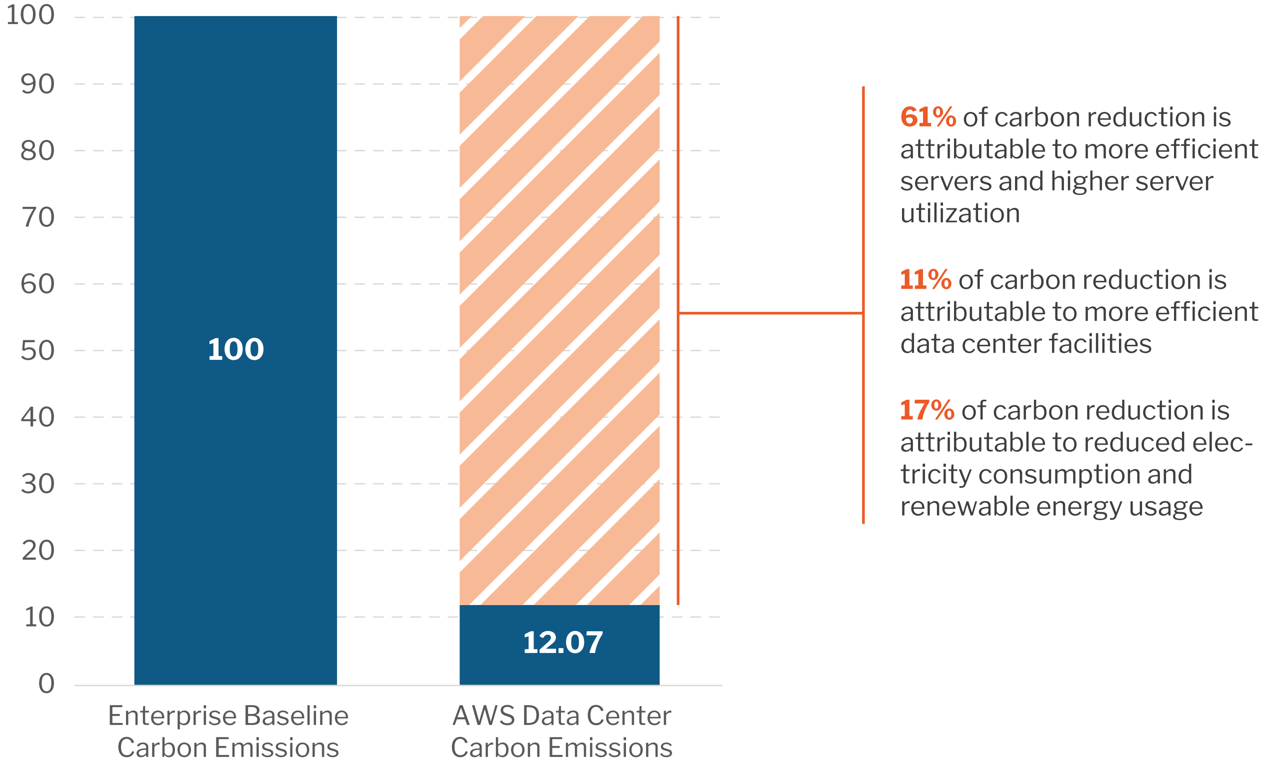
This is contrasted with enterprise infrastructures, which remain fragmented in their management. While efficiency does factor into technology and product choices, efforts are mostly limited to the existing infrastructure and organizational framework, as opposed to fundamental changes to assumptions in engineering and operations. A prime example of this approach is temperature controls, which remain tight at enterprise data centers and are energy-intensive and expensive to maintain year-round. Enterprises also tend to run underutilized or idle servers in significant numbers - the Uptime Institute estimated in a study that roughly 20% of racked enterprise servers are completely unused and abandoned by application administrators, owing to insufficient monitoring of the infrastructure and the lack of a rigorous decommissioning process.

Our results show that AWS's infrastructure is 3.6 times more energy efficient than the median of the surveyed US enterprise data centers. More than two-thirds of this advantage is attributable to the combination of a more energy efficient server population and much higher server utilization. AWS data centers are also more energy efficient than enterprise sites due to comprehensive efficiency programs that touch every facet of the facility.

When we factor in the carbon intensity of consumed electricity and renewable energy purchases, which reduce associated carbon emissions, AWS performs the same task with an 88% lower carbon footprint. (Figure 2).

Figure 2: Carbon efficiency of AWS infrastructure compared to surveyed US enterprises

Source: 451 Research



Even when compared to the top 10% most efficient organizations surveyed, moving to AWS would deliver a 72% reduction in carbon footprint on average. The results suggest that moving to the cloud would reduce workload carbon footprint for virtually any US enterprise in this cohort of companies with revenues between $10m and $1bn.

451 Research expects this carbon benefit to grow in the coming years. Our modeling suggests that AWS should be able to improve its efficiency faster than enterprises using on-premises infrastructure based on currently known server technology roadmaps and expected growth rates in public cloud infrastructure.

Key Survey Findings

Our survey of US enterprises with annual revenue between $10m and $1bn, conducted in the second quarter of 2019, sheds light on the conflicted reality of these organizations when it comes to sustainability and their carbon footprint. On one hand, surveyed businesses monitor at least some efficiency metrics to generate related internal reports, and energy efficiency factors into procurement decisions. Some organizations monitor comprehensively or even track business-level efficiency such as revenue per kilowatt-hour and metric tons of carbon. Virtually all interviewees reported having efficiency initiatives in place, ranging from the servers they purchase to more extensive use of virtualization to more comprehensive data center energy usage assessments.

On the other hand, respondents said that energy efficiency and carbon reduction was not a top priority for their business leaders. Without such a mandate, operations teams tend to lack clear goals, incentives, and the resources to tackle energy efficiency strategically – reliability, availability and performance take precedent on the day-to-day job. Efforts to improve efficiency tend to be limited in scope to work within the existing framework of data center and IT infrastructure – optimizing operations and potentially using more renewable energy rather than making any major and potentially difficult changes to the infrastructure. As one IT manager at a financial services organization put it: “Ideally, we would like to deploy a dedicated team of engineers which can entirely focus on energy efficiency in a data center infrastructure; however, budget and operating expenses are the biggest constraints.”

Enterprise Data Center Facilities – Efficiency Improvements, Not Breakthroughs

For enterprise data center facilities, cooling remains a major source of inefficiency due to tight temperature controls. Most enterprises still aim to keep server inlet temperatures under 72°F, with anything above 75°F setting off alarms, in line with 451 Research and the Uptime Institute’s experience working with enterprise and colocation facilities. This practice is highly energy-intensive because it relies on mechanical refrigeration to keep air in the data hall cool throughout the year. This is not the case with AWS data centers, which rely primarily on the evaporative cooling effect rather than compressors, made possible by designing data centers to operate with wide temperature bands.

Despite encouraging advice from climatic industry body ASHRAE to adopt wider temperature bands (recommending a range from 64.4°F to 80.6°F since 2008)[3], most enterprises would find this difficult to implement. First, there are concerns about risks, such as potential application downtime as a result of any increase in IT hardware failures. New hotspots may develop where cold air delivery is insufficient, and recommendations from facilities teams are often met with resistance from IT. Second, data centers that are designed to operate with a tight temperature band are typically not well-suited to run a more flexible climatic regime even when the organization wants to implement this. For example, without changing server settings, server

3. *Thermal Guidelines for Data Processing Environments, Fourth Edition, The American Society of Heating, Refrigerating and Air-Conditioning Engineers, 2015*

fans would start spinning faster, and air handlers would react with more airflow, increasing their energy use and offsetting some (if not all) of the savings from the lower use of compressors in chillers and computer-room air-conditioning units.

Cooling is not the only factor that contributes to facility-related energy overhead. Many respondents reported low electrical distribution efficiency at their largest data center location, pointing to outdated and underutilized equipment that does not operate efficiently at low load. In fact, more than 75% of respondents said their electrical efficiency was below 80%. The biggest hurdle for enterprises to overcome is a combination of the capital cost of electrical upgrades and the risks (perceived or real) to live operations.

Adding it all up, the power usage effectiveness (PUE, the ratio of total datacenter power and the IT power) ranged from 1.63 to 1.70 in the survey sample and was consistent across company sizes. This is in line with broader industry data collected by the Uptime Institute but well below the efficiency of AWS operations. 451 Research does not expect enterprises to improve significantly in the coming years. According to Uptime's research, data center PUE improved from about 2 (facility power losses and cooling energy equaling the IT load) at the start of the decade to about 1.65 by 2013, but gains have been incremental and inconsistent since then – surveyed PUE hit 1.58 in 2018, but edged up to 1.67 in the 2019 reading.[4]

In fairness, there are many potential explanations why PUEs may be on the rise at a given organization (workload consolidation, newer hardware), but by and large, enterprise data center efficiency has stopped improving. 451 Research does not expect the current PUE gap between enterprise and AWS data centers to shrink.

4. Is PUE actually going UP?, The Uptime Institute, May 2019

Enterprise Server Infrastructure – Progress, But Still a Long Way To Go

There are multiple factors that define the efficiency of server infrastructure. In this study, our focus was on server hardware and how close its utilization is to its efficiency optimum in order to capture the fundamentals of energy efficiency in server infrastructure via a view of performance versus power. Other factors such as the resource balance of servers and the broader infrastructure (network and storage) also affect performance and energy usage, but these factors are more specific to workloads as opposed to a more generic view of infrastructure efficiency.

To that end, 451 Research profiled the server infrastructure of surveyed enterprises by asking questions about the composition of the server population by technology generation (a product of hardware age, speed of adoption and change in server numbers) and virtualization and workload consolidation practices, to estimate utilization levels more robustly than singular numbers can capture.

The study shows that, on average, the target group of US enterprises tends to keep their servers for a little over four years before upgrading, although the average masks a wider spread: some keep a server for as long as seven or eight years while others say they refresh after less than three years as a target – a quite aggressive approach. This policy is not strongly correlated with company size or vertical.

Adopting new, energy-efficient server platforms faster also has a material impact on energy efficiency, and larger companies in the survey reported faster speed of adoption. This can significantly contribute to differences in efficiency among enterprises as well as the gap between enterprises and AWS. AWS's early access to the latest server technology – as much as a year ahead of general availability to enterprises – combined with the lag in enterprise adoption, adds up to a significant effect over time by our calculations.

Virtualization has come a long way and is now ubiquitous. About 93% of the enterprises in the sample reported that they use virtualization (the rest of the respondents did not know). The rate of virtualized server adoption is also substantial with the average adoption rate at 48%. To understand how well servers are utilized, we also asked about workload consolidation practices, rating them from aggressively consolidating to using virtualization but not primarily for utilization purposes. With some assumptions, these responses give us guidance on what level of utilization to expect on average.

Figure 3: Virtualized/containerized consolidation practice

Source: 451 Research

Nearly 40% considered themselves pragmatic in workload consolidation, while nearly a third of respondents said they employ aggressive consolidation practices to drive utilization. This is more common of the largest companies in the survey, with nearly 45% of those saying they are on the aggressive side. This factor also seems to correlate with server expansion plans: those respondents who expect their server numbers to grow also tend to favor utilization, with nearly 40% saying they consolidate aggressively.

This is a strong showing from enterprises in 451's view. However, using average utilization assumptions for each consolidation levels and physical servers, the average server utilization rate works out to be approximately 18% in our sample – well below the utilization of AWS servers due to the much more effective workload consolidation of public cloud that dynamically matches workloads to available server resources to minimize slack capacity.

Figure 4: Server metrics of surveyed US enterprises, $10m-$1bn

Source: 451 Research

ADOPTION OF NEW SERVERS	VIRTUALIZATION RATE	COMPOSITE SERVER UTILIZATION (CALCULATED)
4.7 months	48%	18%

The final element to estimating the composition of enterprise server infrastructures is the change in the number of servers over time. Contrary to the perception that enterprise server purchases are on a downward trajectory, most respondents in our survey were bullish about on-premises capacity. Growth means a distribution of server generations skewed toward more recent platforms but still leads to a carbon footprint that is higher than cloud infrastructure. This is because the server distribution at AWS leans heavily toward more recent server generations – owing to the rapid growth in demand for cloud services – that are typically better utilized and run in more energy efficient data centers.

Looking across all these factors, very few companies consistently outperform their peers, indicating that even with initiatives to achieve best-in-class operational efficiency within an organization, they are not generally effective enough to raise all aspects of enterprise operations in line with best practices. For example, in our dataset, only 37 organizations managed to rate in the top half of the overall group across the three metrics of PUE, server virtualization rate and workload consolidation – and only eight ranked in the top quartile across these metrics.

This highlights the difficulty for enterprises to maximize efficiency across their infrastructure in a comprehensive fashion comparable to cloud infrastructure, where the entire stack is designed, built and operated with efficiency in mind. This shows the structural advantage of AWS stemming from its organizational design, which aligns data center facility and IT teams, engineering expertise, and custom hardware with the cloud business model that helps drive server utilization much higher than is possible for enterprises.

451 Research Energy and Carbon Efficiency Model

Using the survey results, 451 Research calculated the relative operational efficiency at the largest data centers of the enterprises surveyed. To establish a baseline, we looked to the power efficiency benchmark database of industry body Standard Performance Evaluation Council (SPEC), SPECpower_ssj2008[5].

This benchmark is a simulation of a Java-based business logic, and it measures performance (operations per second) against server hardware power consumption. 451 Research believes the efficiency improvements between Intel server generations in this database is representative of the real-world experience.

We took 2010 Intel Xeon-based systems at 10% load as a reference point. This means that in our model, the average energy efficiency of a 2-processor 2010 Intel Xeon (codename Westmere) system at 10% processor load (typical of non-virtualized servers) is '1.' Having established the reference point, we modeled the server infrastructure makeup of enterprises by taking into account server lifespan, technology adoption speed and server infrastructure growth dynamic.

With this distribution of server generations (in relative terms), the next step is estimating how well utilized these systems are. To estimate server utilization at the surveyed enterprises, we used responses on server virtualization levels and workload consolidation practices. This is key because server hardware is not evenly efficient across its load curve. Although this has improved in recent years with better power management techniques and improved server designs, the latest-generation Intel servers still show a factor of 2-3x difference between their peak efficiency point (roughly 70% load) and their light load (10-20%) range. See Figure 5 below.

5. *http://spec.org/power_ssj2008/, data as of July 2019; the steering body for the benchmark, the SPECpower committee, includes representatives from AMD, Dell, HPE, IBM, Intel and the University of Würzburg.*

Figure 5: Efficiency curves of Intel server generations per SPECpower_ssj2008 database

Source: SPEC.org, compiled by 451 Research

Load	Westmere	SandyBridge	IvyBridge	Haswell	Broadwell	Skylake	CascadeLake
100%	4.61	6.96	12.11	15.28	17.31	16.40	18.10
90%	4.54	7.20	12.43	15.33	17.55	16.99	19.53
80%	4.35	7.44	12.75	15.25	17.73	17.78	20.11
70%	4.15	7.43	12.15	15.24	17.72	18.08	20.25
60%	3.85	7.31	11.35	14.73	16.83	17.83	19.83
50%	3.49	6.95	10.37	13.90	15.32	16.94	18.43
40%	3.07	6.20	9.09	12.08	13.30	15.37	16.03
30%	2.50	5.15	7.49	10.31	11.43	13.13	13.64
20%	1.81	3.82	5.58	7.98	8.76	10.19	10.58
10%	1.00	2.15	3.19	4.72	5.11	5.98	6.32

With server hardware and utilization profiles complete, we can then calculate the core operational IT energy efficiency. This metric shows how the energy efficiency of the largest data center compares with the reference point of a 2010 Intel server platform at 10% utilization.

Our results suggest that there is a factor-of-3 difference between the best and worst enterprises in terms of server infrastructure efficiency. On average, surveyed enterprises scored 8.1 in server efficiency, meaning that their server infrastructure is about 8 times more efficient than non-virtualized servers at the start of the decade. The best surveyed enterprise achieved over 12x, while the worse stood at a notch below 4.5.

These results demonstrate that even laggard organizations have made great efficiency improvements thanks largely to new technology that is inherently more efficient. Also, it must be said that this is not the absolute state of efficiency improvement, since most enterprises in 2010 would have scored below 1 in efficiency with their long tail of legacy systems. Even the worst performing enterprises should be more than 5 times more efficient than at the dawn of the decade.

Data shows that virtualization and level of workload consolidation most strongly explain the difference in efficiency among enterprises because both markedly increase utilization levels. Perhaps more surprising is that the third-strongest factor that correlates with how well enterprises fare against their peers is speed of technology adoption, which is statistically slightly more influential than server lifecycle policy.

Figure 6: Server and combined facility and server efficiency of US enterprises, $10m-$1bn

Source: 451 Research

SERVER EFFICIENCY	COMBINED FACILITY & SERVER EFFICIENCY
8.1	9.9

When combined with data center facility efficiency (PUE), our average energy efficiency metric is 9.9. We calculated this by using a PUE of 1.98 as the baseline average, per data from the Uptime Institute from 2011 survey results[6] to establish a reference point similar to server efficiency calculations. This indicates that on average, midmarket US enterprises have boosted their energy efficiency by 10x in almost a decade. But how does this compare to AWS infrastructure?

6. *Uptime Institute Data Center Industry Survey 2019, the Uptime Institute, May 2019*

Comparing Enterprises to AWS Infrastructure

By the same calculations detailed above, 451 Research estimates that by the server efficiency metric, AWS is over 2.5 times more energy efficient than the median of surveyed US enterprises, with a rating above 20. This difference is the result of much higher utilization of servers and an infrastructure that is heavily weighted toward more recent server technology generations that are inherently more energy efficient. In addition, AWS also designs its own servers for maximum efficiency, while enterprises might give more consideration to other features such as hardware redundancy and expandability.

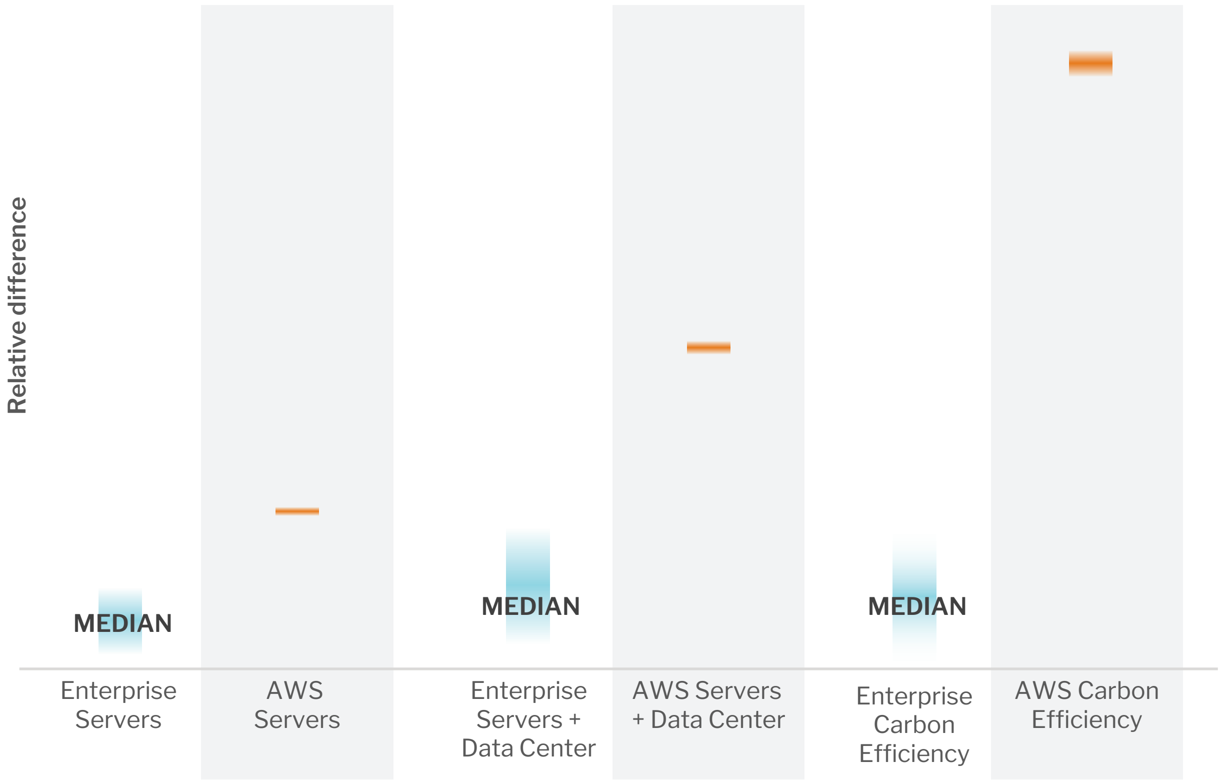
When factoring in data center facility efficiency, AWS is over 3.6 times more energy efficient (in the efficiency-rating range of 34-37) in its data center operations than the median of surveyed enterprises. This additional advantage is due to its more efficient data centers with much stronger PUEs than typical on-premises enterprise data centers, a result of free cooling methods and a leaner electrical infrastructure that introduces lower losses to power distribution.

When we factor in grid carbon intensity and renewable energy to calculate relative carbon efficiency, AWS's advantage extends even more, to over 8.5. This means that the carbon footprint for the same server performance on the AWS cloud is 88% lower than the median of surveyed enterprises.

While enterprises vary considerably in their efficiency with larger ones typically performing better due to having a higher mix of more energy-efficient servers and better utilization, AWS's sustainability advantage is robust. The middle 90% of enterprises (between 5th and 95th percentiles) are likely to see a reduction of 80% to 93%. Figure 7 depicts the spread of results for the surveyed enterprises and AWS. Against the top 10% of the sample, AWS still delivers an average reduction of 72%.

Figure 7: Enterprise data center efficiency compared to AWS

Source: 451 Research



This additional benefit of using AWS infrastructure is due to the fact that the US electrical grid is generally powered by fossil energy sources such as goal and gas. Carbon intensity in most major US regions is between 300 and 500 grams of CO_2 per kilowatt-hour, according to data collected by the US Energy Information Administration[7], with some exceptions in the Northeast and California. AWS extensively reduces the carbon intensity of its electrical purchase well below these levels via renewable power-purchase agreements (PPAs) and associated renewable energy credits (RECs).

For a typical 1-megawatt enterprise data center at 30% electrical utilization, moving to AWS infrastructure would reduce workload carbon emissions by hundreds of metric tons – 400-1,000 metric tons per year, according to the model calculations, depending on the location of the enterprise site. This is the equivalent of up to 215 average passenger vehicles driving a total of nearly 2.5 million miles a year, according to data from the United States Environmental Protection Agency[8].

7. *Assessing the evolution of power sector carbon intensity in the United States Greg Schivley, Ines Azevedo and Constantine Samaras, Carnegie Mellon University, June 2018*
8. *Greenhouse Gas Equivalencies Calculator, http://epa.gov, December 2018*

451 Research expects the implied energy efficiency gulf between enterprise operations and AWS to widen in the next 24 months. Our modeling suggests that the faster growth of cloud combined with Intel's and AMD's currently known server technology roadmap will push the efficiency of AWS infrastructure ahead faster.

This does not even account for processing accelerators such as GPUs and FPGAs and new memory technologies that make large persistent memory capacities possible. These are readily available to AWS customers and should increase workload energy and carbon efficiency even further for select applications. We also expect hyperscale operators to continue to innovate in cooling and electrical design, which will push the efficiency of both the data center facility and servers beyond what's currently possible.

While we did not quantify it in this study, 451 Research surmises that enterprises would also enjoy carbon benefits in the form of lower embodied emissions when moving workloads to the cloud as opposed to building their own infrastructure capacity. Hyperscale data centers are leaner and better utilized, and hyperscale server supply chains are highly optimized for lower use of materials, which in effect lowers the energy required to produce systems. We also note that cloud providers tend to use higher-performance processors than enterprises because they can monetize them more effectively, so they require fewer servers to deliver the same overall performance. These combined factors result in lower carbon overhead for a unit of performance.

Appendix: Methodology

The focus of this model is to capture the carbon impact of key design and operational features of enterprise data centers compared to AWS, and to understand how key components impact the overall efficiency picture. The core of the model centers on Scope 2 emissions (electricity), and this analysis does not include Scope 1 emissions (direct emissions from site operations such as vehicle emissions, cooling system refrigerants, diesel engine power generator emissions, etc.). Our model focuses on key factors that enterprises have direct control over and that substantially affect energy efficiency and carbon footprint. Scope 1 emissions do not reflect the core operational efficiency of a datacenter. For example, virtually all operators need generators that run tests or when the grid fails – there is little room for differentiation. Future versions of the model may include some Scope 3 emissions with embodied carbon calculations for facility and IT systems, but we don't expect that to meaningfully alter the conclusion of the analysis.

EXTENDED CARBON MODEL W/ EMBEDDED EMISSIONS

				CORE ENERGY CARBON MODEL				
Embodied facility	+	Embodied IT	+	Grid - offset by renewable purchases	x	PUE	x	IT efficiency

The carbon emissions model consists of five major areas: embedded (or embodied) emissions of both data center facilities and IT hardware; carbon intensity of the grid; then facilities and IT operational emissions. The objective of the model is to show the difference between enterprise and cloud operations. The output of the model is a ratio that shows the relative carbon efficiency difference between enterprises and the AWS cloud. The model incorporates survey data, third-party industry sources and data from AWS. Some survey questions are not directly used in the model calculations but are for further background analysis.

Our survey focused on understanding some of the characteristics of US enterprises with $10m-1bn in annual revenue that influence efficiency metrics – including policies and attitudes regarding consolidation levels, speed of server technology adoption and typical server lifespan. We believe that such an approach, while requiring careful tuning of some assumptions, creates a much more robust picture and provides better context than asking exclusively for technical specifications and operational metrics, many of which may not be tracked with the required detail.

Embedded facility and IT emissions (the carbon related to manufacturing/building) – This factor relies on third-party industry data, such as studies from Schneider Electric and IBM, and AWS's input on how its own facilities (both structural and M&E) and IT server build practices compare. This can include differences in building materials, redundancy levels, power density, battery chemistry, etc. While 451 Research expects AWS to command a clear lead due to its leaner design, we note that this factor primarily serves the comprehensiveness of the model, and the overall difference in the effect of design choices is relatively modest compared to other areas of carbon emissions.

Facility and IT utilization and power density – These factors dictate how much the effective rate is for embedded facility carbon. The lower the space utilization and power density of a facility, the more embedded carbon falls onto IT systems to carry. We also factor in IT power efficiency (useful work per energy) from later questions to account for effective embedded carbon per unit of work, using survey data and AWS input. To keep the survey brief, we decided to ask about space utilization only, which we think is easier to track than power for most enterprises.

Grid carbon intensity – Carbon emissions per kilowatt-hour of energy; third-party data.

PUE – Power-usage effectiveness; shows the facility energy overhead as a ratio of the IT load.

Server hardware power efficiency – The inherent design power efficiency of the server that is calculated using *server distribution by age, server utilization and power efficiency* data from the Standard Performance Evaluation Council's database specpower_ssj2008.

- *Server age distribution:* proxy for server technology generation that largely defines the server's efficiency potential. To gain this distribution, we asked for average lifespan, speed of adoption of new tech (to account for additional lag compared to the cloud) and capacity change (skew of distribution). Q10 *How long is the typical lifespan of the servers running in your largest data center?*
- *Server utilization*: Instead of asking for server utilization, which we have found is not practical to obtain, we asked about the maturity of IT operations by gauging virtualization levels, any projects in motion that aim to increase virtualization levels over time, and aggressiveness of consolidation. 451 Research based our assumptions on how these responses translate into utilization numbers on third-party industry data.
- *Power-efficiency data from SPEC*: SPEC maintains a database on server power efficiency per a test suite that simulates a complex business logic and benchmarks performance against power use across the load curve. Using this data, the model can assess the relative power efficiency of servers based on their technology generation (age) and utilization.

While server makers do aggressively fine-tune hardware and software specifically for the specpower_ssj2008 benchmark to attain the best possible result in ways that arguably do not represent a typical deployment case, we relied on averages across multiple submissions and used the data to calculate efficiency improvements with newer server generations and with better utilization. We believe these are representative of real-world behavior of hardware and software in a generic enterprise IT environment.

While SPEC data is in 10% increments, we needed finer granularity of 1% for our calculations as we modeled IT operational efficiency of the surveyed US enterprises. We did that by using linear interpolation between measurements as an approximation to an implied efficiency curve.

Based on virtualization and consolidation levels, we calculated composite average utilization of each server generation for each enterprise, then weighted such efficiency readings by distribution of server generations per enterprise. We tested this against a more detailed hourly workload simulation (e.g., internet traffic profile during a day) where a more complex calculation using hourly utilization and energy-efficiency readings would generate the efficiency reading, but the total difference from using a simple average utilization and associated energy efficiency reading was typically 1% or less.

About 451 Research

451 Research is a leading information technology research and advisory company focusing on technology innovation and market disruption. More than 100 analysts and consultants provide essential insight to more than 1,000 client organizations globally through a combination of syndicated research and data, advisory and go-to-market services, and live events. Founded in 2000 and headquartered in New York, 451 Research is a division of The 451 Group.

NEW YORK
Chrysler Building
405 Lexington Avenue, 9th Floor
New York, NY 10174
+1 212 505 3030

SAN FRANCISCO
505 Montgomery Street,
Suite 1052
San Francisco, CA 94111
+1 212 505 3030

LONDON
Paxton House
30, Artillery Lane
London, E1 7LS, UK
+44 (0) 203 929 5700

BOSTON
75-101 Federal Street
Boston, MA 02110
+1 617 598 7200

451 Research