Technical white paper

# Increase competitiveness with a comprehensive IT efficiency strategy

## Contents



## 1. Introduction

The exponential growth in the demand for digital technologies and services has no signs of slowing down; rather, the COVID-19 pandemic has hastened the digitalization of our global economies. To outpace the competition, enterprises need to increase efficiencies and optimize their digital infrastructures to achieve more with less. This can not only reduce costs but can also improve agility and minimize the environmental impacts of the IT estate, which is an increasing expectation of customers, investors, and other stakeholders.

**"Challenged by overwhelming demands for real-time data, enhanced agility, and virtually unlimited scalability, CIOs are facing unprecedented complexity in the data center. As leaders prepare for the ubiquity of cloud, IoT, and Big Data, they must still keep efficiency under control."[1]**

Defining and implementing an edge-to-cloud sustainable IT strategy—from procurement to operations to end-of-use management—will enable organizations to get the most out of their IT investments. An IDC study found that 70% of companies prefer to partner with or buy from companies with a demonstrated commitment to sustainability. Meanwhile, 89% of surveyed organizations want to lower overall IT energy consumption and monitor it, and 90% say they want the ability to refurbish and recycle retired assets.[2]

Part one of this paper, accelerate digital transformation with sustainability, IT efficiency, outlined how digital transformation leaders can go about setting a sustainable IT strategy for their organization. It briefly touched upon the importance of identifying key drivers of sustainable transformation, setting clear objectives and targets focused on the most significant areas of impact and opportunity, identifying the right partners, and choosing the right measurement and reporting tools.

This second part will examine the key considerations and technologies that IT leaders should consider when implementing this sustainable IT strategy. The paper first explores opportunities to drive efficiency in the data center infrastructure to reduce operational environmental impacts without sacrificing productivity or performance, before turning to mitigate the upstream and downstream impacts of the IT organization.

## 2. Driving efficiency as a priority

As the demand for digital services grows, so too does the complexity and resource demands of IT infrastructure. IT organizations seeking to scale and meet ever-growing demands from the business are often constrained by resource limitations such as space and power and cooling. Moreover, these resources directly translate into costs to the business, whether from energy consumption or the associated demand for cooling and ancillary equipment. Fortunately, innovations in the IT sector have enabled today's infrastructures to be far more efficient than their predecessors, yet significant opportunities remain to drive more sustainable and energy-efficient operations. This not only enables the IT organization to drive innovation and deliver on business objectives but also to contribute toward sustainability targets and objectives.

While offsetting one's environmental footprint by protecting and regenerating natural capital is a key part of the solution, reducing one's absolute environmental footprint through responsible consumption should remain the number one priority. First, because the more efficient digital transformation leaders are, the less their organization will have to compensate, which typically requires an additional budget. Second, it helps cut cost and may create additional revenue streams.

Because the IT hardware layer is not only the central piece of data centers, where the value resides but also where most of the power consumption occurs, it is key to drive efficiencies at that level first. Once the hardware layer is as efficient as it can be, the second step should be to adapt the mechanical and electrical (M&E) infrastructure around it. If IT leaders start by working on efficiencies at the data center level, the M&E infrastructure will be optimized for a hardware infrastructure that will likely change once they start driving efficiencies at the hardware level, requiring the M&E infrastructure to be optimized again.

At Hewlett Packard Enterprise, we describe this concept as "sustainable IT" or "IT efficiency." Sustainable IT optimizes IT infrastructures, resulting in a variety of benefits such as more efficient asset utilization, reduced total cost of ownership, smaller carbon footprint, and conformance with tightening market requirements. The IT efficiency approach includes:

- **Energy efficiency:** Delivering an optimum level of compute, storage, and connectivity in exchange for the lowest input of energy.
- **Equipment efficiency:** Maximizing IT processing power and storage capabilities with the fewest number of IT assets, helping to solve stranded capacities in compute, storage, and network.

[1] Data center cooling and efficiency: Thinking outside the box, William Gast, Datacenter Dynamics, 2016
[2] IT's role in driving sustainability progress, IDC, 2020

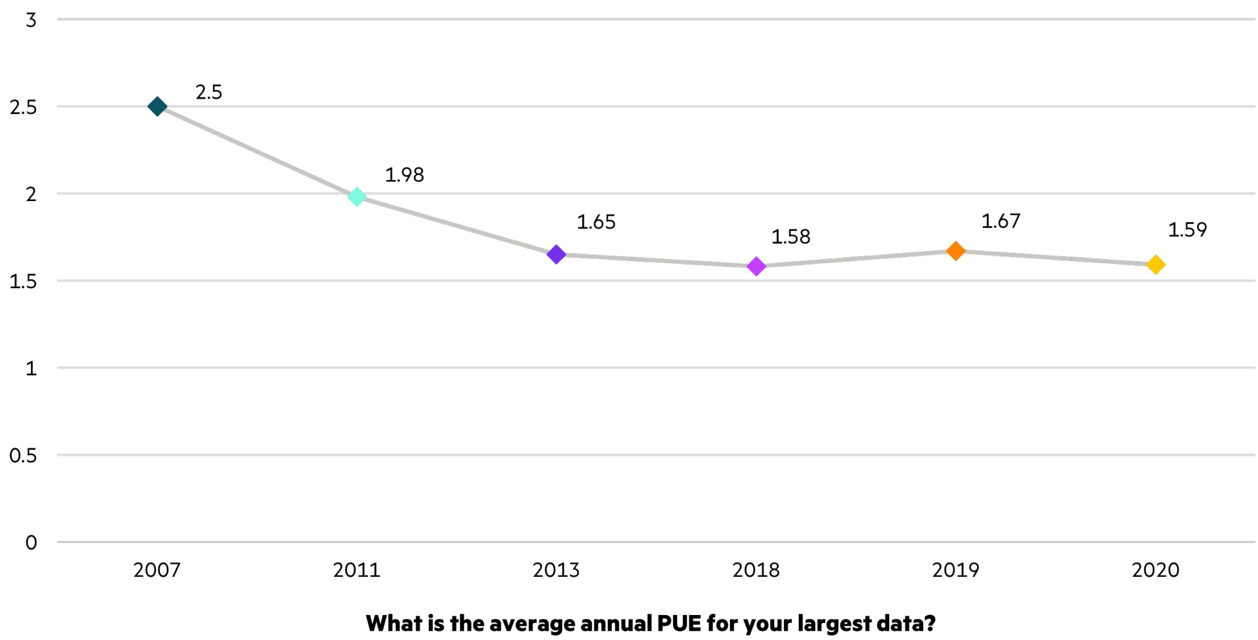
- **Resource efficiency:** Engineering products to work efficiently from edge to cloud while requiring the least amount of support, staff, and equipment for power conversion, cooling, and resiliency.
- **Software efficiency:** Writing efficient code and using intelligent software to automate environments, drive efficiencies, and improve management practices.

This section will explore this concept in-depth, beginning with hardware-level efficiency, where the IT efficiency journey should begin, followed by resource efficiency opportunities at the data center infrastructure level.

## a. Making your IT infrastructure efficient

Various industry surveys find that IT leaders have been successful in improving efficiencies by focusing on their data center's power usage effectiveness (PUE).[3] Significant PUE savings opportunities, mostly residing in M&E changes, have already been achieved and PUE has begun to stall in recent years (Figure 1). It is now time for IT leaders to focus on the IT hardware efficiency dimension, where immense untapped opportunities still exist. This dimension is also described as energy and equipment efficiency.

The PUE is a ratio that describes how efficiently the M&E components of a data center use energy. It is calculated by dividing the total amount of energy consumed in the entire data center by the energy consumed by the IT equipment. The perfect PUE ratio is 1. The 2020 Uptime Institute Survey[4] shows an industry average PUE of 1.67 in 2019, while best practitioners can reach a PUE below 1.1.




Source: Uptime Institute global survey of IT and Data Center Managers 2020, n = 445

**Figure 1.** Data center energy efficiency gains have flattened out

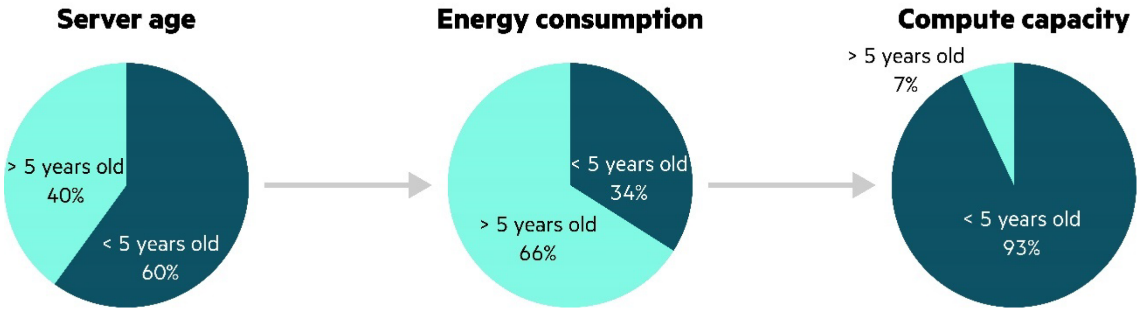
Improving IT hardware efficiency will necessitate both the procurement of energy-efficient solutions, as well as the active management of this equipment to operate at the highest levels of efficiency. But achieving energy efficiency without equipment efficiency is a job half-done.

[3, 4] Beyond PUE: Tackling IT's wasted terawatts, the Uptime Institute, 2020

### Reducing IT sprawl and increasing output

It is common knowledge among IT professionals that data center operations can be inefficient, with low utilization rates, overprovisioning, and obsolete or stranded equipment. To address these inefficiencies, IT leaders must first assess their IT estate and monitor key indicators for energy and equipment efficiency.

On average, more than 65% of the power used by IT in data centers is used to process just 7% of the work (Figure 2).[5] That staggering number illustrates the challenge at hand: How can such a large number of the assets that organizations pay for, both at purchase and during operations, be responsible for such a small portion of the overall outcome they are intended to deliver?




Source: European Union Resource Efficiency Coordination Action

**Figure 2.** Average output and energy consumption per server age

### Causes

There are numerous possible causes for these equipment inefficiencies, and they may vary by time, type of organization, or even geography. Some of the root causes identified include:

- **Aging infrastructure:** On average, 40% of enterprise data center servers are five years old or more.[6] Considering the leap in energy effectiveness experienced by the industry over time, it is likely that such a large number of old assets will contribute unfavorably to energy-efficiency objectives. More specifically, the number of operations per watt that the latest generations can achieve can easily be three times more than that of older generations.[7]
- **Underutilization:** On average, servers in non-virtualized environments are underutilized by 90%.[8] Servers in virtualized environments enable better utilization but they still run, on average, underutilized by 75%.[9] While there are several reasons to not run servers at 100% utilization, such as resilience and peak load events, average industry benchmarks are far from optimization. Often, the memory is the bottleneck that inhibits more utilization.
- **Zombie servers:** On average, 25% of data center servers are zombies, meaning they are plugged-in; consume power and space and cooling; and are the source of countless operating and maintenance costs yet do little to no work at all.[10]

Due to these inefficiencies, the average data center is overprovisioned by 59% for compute and 48% for storage.[11] The potential cost savings from addressing these inefficiencies is significant, including reductions from the cost associated with purchasing, operating, and maintaining hardware by reducing power, cooling, space, and software license expenses. Driving IT efficiency would not only allow IT infrastructures to expand when needed without significant additional investment, but would also help free up funding for innovation or additional resources where needed.

To put this in perspective, it would be irrational to purchase a highly energy-efficient washing machine and then wash just one sock per cycle. The same should stand true for IT assets, particularly given that between 60 and 80%[12] of the environmental footprint of data center IT assets results from the use phase.

[5, 6, 8, 9] Beyond PUE: Tackling IT's wasted terawatts, the Uptime Institute, 2020
[7] Based on SPECpower_ssj2008 benchmark
[10] Zombie & Comatose Servers Redux, Koomey and Taylor, 2017
[11] Can enterprises achieve both scalability & control when it comes to cloud? 451 Research, 2016
[12] Product Carbon Footprint, PAIA, 2020

**Indicators**

Key indicators that should be tracked include utilization (processor or storage utilization) and performance per watt (operations or terabyte per watt).

- **Utilization:** For servers, the largest IT power consumer,[13] tracking processor utilization is paramount. The main task of a server is to compute information and the processor is the core electronic component of computation. Considering utilization is defined as a requirement divided by a capacity, processor utilization (in percentage) is the primary metric to consider when measuring server efficiency. Processor utilization is a key indicator of the capacity of the server—for instance, is it close to maximum utilization and, therefore, at risk of becoming a bottleneck? Or is it idling most of the time, wasting energy, space, cooling, and software licenses while doing no work? Through measurement of processor utilization, IT organizations will find the latter is most likely to be true. These assets should be handled first considering the cost they incur with little to no value added.

  So, how pervasive is this problem? Processor usage averages 23% for servers, with 80% at less than 30% utilization according to a 2018 study.[14] To illustrate how pervasive this issue is, consider that Alibaba, Amazon, Mozilla, and Google™ were all cited in the past decade (albeit at different times) as operating servers at extremely low utilization rates of 10%, 7%, 6%, and 20–40% respectively.[15, 16]

  Further, the Uptime Institute states that more recent processor utilization rates in virtualized environments could be around 20 to 30%, which is still low.[17] A subcategory of underutilized servers are comatose servers, which can be defined as those that have not delivered information or computing services in six months or more. A 2015 study by Koomey and Taylor suggests that 30% of servers are comatose and sitting idle.[18] Not only are these wasted resources that consume power, cooling, and space, they are also a source of wasted software licenses. The study estimates the cost of an idling server to be $3,000 per year. When extrapolated, the worldwide estimated cost of comatose servers amounts to $30 billion per year.

  Similarly, storage can also suffer from underutilization, leading to unnecessary, costly operations. Traditionally, businesses have asked storage administrators for more capacity than an application would need to ensure that they wouldn't run out. There is little recent research on this issue, but a Gartner report estimates that 40% of storage capacity is unutilized,[19] and HPE estimates that the overall cost of keeping this free space unutilized for 500 TB of storage infrastructure over five years amounts to more than $2 million.[20]

  However, identifying and implementing solutions that deliver better utilization and equipment efficiency can be complex. Recent innovations have come to market to address IT sprawl. Software-defined architectures have the power to run efficient, rightly utilized data centers. As an example, HPE developed a new class of hardware, known as HPE Synergy composable infrastructure that consists of a pool of compute, storage, and networking resources in a common frame.

  Unused capacity can be repurposed for other applications. HPE Synergy lowers IT infrastructure costs by 22% and cost of operations by 35%.[21] This is mainly due to its large internal memory (up to 6 TB), capable of hosting significantly more applications than traditional architectures. More applications in one frame mean higher utilization ratios, and higher utilization ratios mean reduced costs and environmental footprint. In general, larger server memories can help with higher utilization rates.

- **Performance per watt:** The second indicator, performance per watt, is important as well. Being equipped with an architecture that is not only fully utilized but also consumes the least amount of energy per workload is essential in delivering a coherent sustainable IT strategy.

  Server performance per watt, known as energy effectiveness, increases with each new generation of server. This performance improvement is clear when viewing results of the SPECpower_ssj2008 performance benchmark, which was developed by an industry organization known as the Standard Performance Evaluation Corporation. The results demonstrate application performance and power effectiveness gains over time, as shown in Figure 3. Over four years, power effectiveness gains, while slowing down, still provide substantial power savings opportunities.

[13] Recalibrating global datacenter energy-use estimates, Koomey, 2020
[14] Right-sizing server capacity headroom for global online services, Verbowski et al., 2018
[15] Datacenter Servers Suck—But Nobody Knows How Much, Wired Magazine, 2012
[16] The Datacenter as a Computer, Mark D. Hill, 2013
[17] Beyond PUE: Tackling IT's wasted terawatts, the Uptime Institute, 2020
[18] New data supports finding that 30 percent of servers are Comatose, Koomey & Taylor, 2015
[19] IT Key Metrics Data 2019: Key Infrastructure Measures: Storage Analysis, Gartner, 2019
[20] Understanding the Hidden Cost of "Free Space" in Datacenter Storage, HPE, 2020
[21] Quantifying the Business Value of Enhanced IT Agility and Performance with HPE Synergy, IDC, 2021

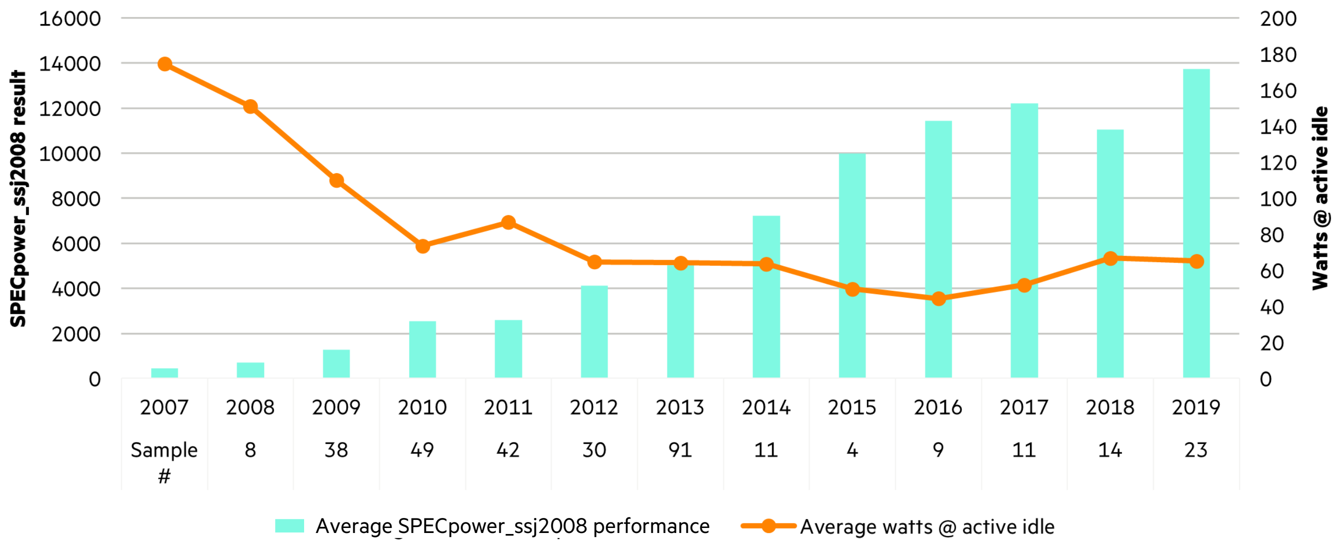



**Figure 3.** Active idle power consumption and energy effectiveness over server generations[22]

For instance, the HPE ProLiant DL360 Gen10 server delivers five times better compute power performance than the older Gen7. Moreover, some configurations are even more performant than others. With the intent of postponing the end of the Moore law, HPE worked with AMD to make energy effectiveness a key attribute of AMD's EPYC processors. This resulted in HPE ProLiant DL385 and DL325 Gen10 servers with AMD's second-generation EPYC processors beating previous energy effectiveness records by 28%.[23]

A more radical example of a high-density energy-efficient server is the HPE Moonshot architecture. Due to low-power system on chips (SoCs), integrated solid-state drive (SSD) storage, and integrated workstation GPU, HPE Moonshot enclosures house up to 45 servers in a 4.3U space. This takes up significantly less space and lowers power consumption by up to 70% in comparison to rack-mounted servers.[24]

Regarding storage, new technologies such as SSDs flash memory and non-volatile memory express (NVMe) SSDs are significantly more energy efficient than HDDs. An SSD doesn't have to expend electricity spinning up from a standstill. Consequently, none of the energy consumed by the SSD is wasted as friction or noise, rendering them more efficient and thereby resulting in a lower energy bill. For all common workloads such as web, file, and mail, SSDs are one to two orders of magnitude more energy efficient than HDD RAID.[25]

Finally, regarding networking, while technologies driving a better energy efficiency output are not as obvious as with compute and storage, significant progress has been achieved through photonics. As much as 40% of electricity is spent just moving information inside data centers, so this is not an area to underestimate.[26] Computers today send signals via electrons over copper wire and these electrical currents lose energy as they travel over the wires. To successfully transmit data, the signals need to be amplified with additional power on the other end. Photons don't have those characteristics. Sending photons over fiber optic cable promises greater energy efficiency and faster processing than traditional copper wire transmission.

### Consolidation strategy

To identify and address IT efficiency opportunities, IT leaders should begin with a thorough review of their estate, according to the age, power, performance, and utilization ratios of assets. This exercise will help to understand which assets host what, and to what sources environmental and financial costs are associated. This discovery phase can also support an infrastructure modernization or data migration project when it's time to do so, which means it will not only deliver efficiency and sustainability benefits but also contribute to more holistic business objectives.

[22] Efficiency benefits and challenges of information technology lifecycle management, HPE, 2020
[23] HPE ProLiant shatters 37 world records, HPE Newsroom, 2019
[24] Performance and secure banking desktops with HPE Moonshot, HPE, 2018
[25] A Comparative Study of HDD and SSD RAIDs' Impact on Server Energy Consumption, Tomes, 2017
[26] How Photonics Can Help Prevent a Digital Energy Crisis, HPE, 2017

Apart from reducing carbon, water, and overall environmental footprint, here are some of the business benefits a hardware consolidation could bring to data center operators and digital transformation leaders:

- Reducing the number of servers in the IT environment, which saves on per server maintenance costs
- Bringing down data center footprint, or space, which extends the life of the existing data center, helping avoid the cost of building a new one, as well as driving up the ratio of compute power per square foot
- Minimizing TCO for the IT environment, which increases the number of servers, a technician can easily manage
- Reducing operating costs and avoidance of new costs such as electrical or heating, ventilation, and air conditioning (HVAC)
- Simplifying the IT environment, which facilitates remote server management and streamlines overall asset management

As a second step, if the server estate audit confirms a significant share of the estate is underutilized or not performing at satisfying power performance, a hardware consolidation project should be considered. The ROI of such a project shouldn't be hard to justify when recalling that a comatose server can cost up to $3,000 per year.[27] This also presents an opportunity to consider implementing a multicloud and hybrid cloud strategy that allocates workloads where they operate best—on-premises, colocation, private cloud, or public cloud.

There are five main types of server consolidation initiatives:

- **Centralization:** Co-locating servers and/or storage into fewer locations or one central location
- **Physical:** Consolidating servers or storage systems with the same application types or platforms onto fewer or larger systems with the same application type or platform
- **Data integration:** Combining data with different formats onto a similar format or platform
- **Application:** Consolidating servers or storage systems supporting different types of workloads onto fewer or larger systems
- **Storage:** Consolidating storage onto fewer or larger storage systems independent of server type, OS, or application

An example of an organization that implemented this strategy is Barclays. In 2013, Barclays undertook a server estate audit, or Zombie Hunt, and was able to retire more than 9,000 servers. This exercise saved $4.5 million annually in power, $1.3 million in maintenance costs, and cleared up close to 600 racks of data center space, 2.5 MW of cooling load, approximately 20,000 network ports, and 3,000 SAN ports.

Beyond the cost savings and streamlined support, Barclays also found that by decommissioning older servers, they were able to move away from technologies that were holding the business back, and replace them with IT services that were more future-friendly.[28] The Green Grid's Solving the Costly Zombie Server Problem white paper and keynote from Barclays' former head of IT operations, Paul Nally,[29] are helpful resources to better understand what it takes to implement such strategies.

Third-party advisors can also be considered to undertake such an audit when internal staffing is limited. For instance, Advisory and Professional Services provide consulting, financial, educational, and operational services to support an organization's digital transformation, leveraging intellectual property tools.

### Decoupling workloads from hardware layers: virtualization and containerization

In this section, we will explore how virtualization and containerization can drive equipment utilization up by decoupling workloads from the hardware layer. While this technique has been around for a long time already, it remains one of the most important levers of action to do more with less.

Virtual machines (VMs) and containers are two virtualization methods. Both aim at emulating computer systems that will allow the user to run multiple applications on the same hardware. For instance, applications functioning on a different operating system (OS) than the hardware OS (host OS) will be able to run. The hypervisor, in VMs' case, or runtime engine, in containers' case, are the software architectures enabling this decoupling of hardware and applications, which leads to dematerialization of infrastructure.

Virtualization provides part of the solution when tackling the shortfall of insufficient processor utilization. The ability to run multiple workloads on the same hardware leads to increased equipment efficiency and lowers operating cost, reduces additional space, and decreases energy usage. In both cases, virtualization drives equipment efficiency. On average, virtualized environments yield 2.5x more utilization in servers when compared to non-virtualized environments.[30]

[27] New data supports finding that 30 percent of servers are "Comatose," Koomey & Taylor, 2015
[28] Solving the Costly Zombie Server Problem, Green Grid, 2017
[29] Server Roundup Winner Barclays: Discipline in Decommissioning, Uptime Institute, 2013
[30] Beyond PUE: Tackling IT's wasted terawatts, the Uptime Institute, 2020

From a resource perspective, the main difference between VMs and containers lies in the lighter consumption of the latter. VMs require several guest OS layers to run different application types. These guest OS can take up a significant amount of the available resources. Conversely, containers package together everything needed to run a piece of software without the need for a guest OS to host it. As a result, there is more room to accommodate additional resources. This partially explains why containers are well-known to run better on bare metal versus already virtualized environments, with some software providers claiming 90% processor utilization versus 15% CPU utilization.[31] In general, leaving out the hypervisor needed in virtualized environment saves on software licensing cost as well.

When looking for hardware platforms that can drive better equipment efficiency in a virtual machine environment, one should look at disaggregated hyperconverged infrastructure (dHCI). Users often leverage traditional HCI solutions to get a single management pane for the environment, but the amount of compute and storage resources typically scale linearly and application environments do not scale that way.

Rather, there will either be a requirement for more compute resources or for more storage resources. The result of traditional HCI architectures is, therefore, often stranded and idling resources. By decoupling compute from storage, paired with an Intelligent Data Platform technology, dHCI provides 30 to 50% better data reduction than traditional HCI platforms and 20 to 30% better CPU utilization for workloads per server.[32] This, in turn, allows for a reduction in the number of compute servers and drives down associated power and heat costs. In summary, dHCI is an efficient IT solution that maximizes IT processing power and storage capabilities while minimizing cost and resource demand—doing the most amount of work with the least amount of equipment and environment footprint. dHCI brings the simplicity and virtualization benefits of HCI without having to compromise on equipment efficiency.

### Sustainable consumption models: servitization and as-a-service models

Part one of this paper, Accelerate digital transformation with sustainability, IT efficiency, offers numerous benefits of developing a sustainable IT strategy; however, these efforts can require significant time, efforts, and resources that may be better spent on other activities. That is understandable given IT leaders are being asked to complete more digital transformation projects with less staffing and resources. With the emergence and maturity of infrastructure-as-a-service (IaaS) models, which provide access to technology versus ownership of hardware, digital transformation leaders can now implement a sustainable IT strategy in their data center without having to worry about the continuous operational aspects. This section outlines many ways that aaS models deliver on joint IT and sustainability objectives.

Traditional IT hardware vendors have become solution providers. HPE, for instance, has pledged to offer its entire portfolio through an as-a-service, pay-as-you-go consumption model by 2022. This will allow digital transformation leaders to benefit from a cloud experience, on-premises and at the edge, with HPE GreenLake cloud services offering, which includes an option to provision on-premises data centers with new compute or storage assets in less than 14 days. This means there is no need to overprovision to cope with needs that will come in six months or a year.

Consumption-based models typically monitor current hardware requirements for running workloads and predict future hardware needs and timelines based on actual workload information. This allows customers to avoid significant overprovisioning of IT equipment, increase equipment utilization rates, optimize refresh cycles to get the latest and most efficient generations, and reduce the need for (and energy consumption of) associated resources such as cooling, UPS, and power conversion equipment. On top of that, users only pay for what they use.

Because metering and charging are based on actual usage, these metrics are available to users in near real-time. These analytics include IT efficiency performance indicators such as product and data center-level utilization, in a single pane of glass, including visibility of workloads that are on-premises, in colocations, private clouds, and public clouds.

As-a-service (aaS) adoption also enables career growth and opportunity for IT professionals. Despite the changes ahead and apprehension from younger IT decision-makers, about nine in ten (87%) of respondents in an HPE survey agree that aaS adoption will advance their career and envision a shift in their roles from day-to-day IT support toward business strategy. Globally, seven in ten IT decision-makers say aaS adoption will make their role within their organization more important (72%) and give them more control over data (70%), and budget (71%).[33]

Finally, the as-a-service consumption model brings one additional benefit to the sustainability equation, in this case, the vendor, who is an IT solution provider, owns the asset from the beginning to the end. Sustainability professionals refer to this as extended manufacturer responsibility, a concept whereby manufacturers of products bear a significant degree of responsibility for the environmental impacts of their products throughout the product lifecycle, including downstream impacts from the disposal of the products. Producers accept their responsibility to minimize lifecycle environmental impacts, both legally and financially, and are incentivized to repair, refurbish, and reuse assets to extend their useful life. It's not difficult to understand the importance of keeping IT vendors and solution providers accountable when considering that more than 50 million tons of e-waste are generated every year.[34] E-waste is the fastest growing waste stream in the world and finite critical raw materials like gold or cobalt are buried in landfills when they could be recovered.

* May be subject to minimums or reserve capacity may apply
[31] Five reasons you should run containers on bare metal, not VMs, Diamanti, July 2018
[32] HPE InfoSight data from HPE Nimble Storage dHCI, HPE, 2020
[33] HPE 2019 As-a-Service: Driving Change Report
[34] Time to seize opportunity, tackle challenge of e-waste, United Nations, 2019

The results are tangible. HPE modeling found that implementing our as-a-service delivery models versus a traditional capital expenditure model can result in 33% energy cost savings over five years[35] and a 30% reduction in TCO.[36]

**Efficient disaster recovery**

Lastly, this section explores the importance of balancing resilience and efficiency in the implementation of a sustainable IT strategy.

In 2020, an increasing proportion of data center outages cost more than $1 million (now nearly one in six vs. one in 10 in 2019) and a greater percentage cost between $100,000 and $1 million (40% vs. 28%).[37] The findings also show that on-site power problems remain the single biggest cause of outages, followed by software/systems and networks.

Resilience and disaster recovery are essential dimensions of any IT strategy. A single component or point of failure can cause significant damage to the entire infrastructure and associated applications and services. Data centers are often considered as a critical infrastructure for a society that requires as much care as a power station or a hospital—and rightfully so when considering these data centers host and process data to support applications used by the healthcare, public sector, banking, and other critical sectors.

Traditional resilience is often achieved with infrastructure redundancy and standalone systems used only in case of emergency. These can be either designed with N+1, 2N, or 2N+1 capacities, with "N" representing the amount of infrastructure needed to operate the facility at a full IT load. As an example, a 2N system is fully redundant, with an independent, mirrored system that can take over if the first system fails.

However, resilience comes often with a cost, both financially and environmentally. The problem is that resilience has traditionally meant redundancy, and redundancy means additional resources that, at some point, consume natural resources and have an environmental footprint. For instance, a generator as a redundant backup to a grid power supply will usually run on diesel, a high-carbon intensity fuel and will also generate heat while running. Unutilized equipment in data centers can cost up to $3,000 per comatose server (as explained previously), yet the average total cost per minute of an unplanned outage increased to $5,617 in 2010 (it was $7,908 in 2013 and $8,851 in 2016).[38] Given this, transformation leaders must reconcile the need for both data center efficiency and resilience.

It's also important to consider alternative solutions to resilience that complement and help reduce the need for redundancy. Intelligence is one. In June 2018, IDC completed a worldwide survey asking end-user customers about their experiences with and aspirations for cloud-based predictive analytics. Customers see the use of artificial intelligence (AI)/machine learning (ML) algorithms to drive improvements in their infrastructure and processes in a positive light. The use of AI/ML to drive increased customer value was desired by 85.7% of respondents. There was a high interest in increasingly autonomous operations (73.8%) to drive improved productivity and 71% of respondents wanted to see increased use of AI/ML to help drive autonomous operations.[39] AI has proven how useful it can be for anticipating, predicting, and taking action on potential failures and outages.

By leveraging the 1,250 trillion data points and 1.58 million saved hours[40] of predictive analytics coming from its customer installed data base, HPE developed HPE InfoSight, a tool that uses AI to help IT leaders predict and prevent infrastructure problems before they happen. The benefits are clear: HPE InfoSight automatically resolves up to 86% of IT environment issues, meaning businesses save 86% of system management time.[41] In 2018, it saved one customer more than $1 million in storage spend.[42]

A concrete example of how this intelligence brings resilience without adding redundancy is HPE Primera, the tier 1 mission-critical storage platform that leverages AI from HPE InfoSight, allows 100% guaranteed availability.[43] IT leaders get high availability and resiliency to achieve the agility and efficiency a modern data center demands.

[35] HPE internal calculations, HPE GreenLake Environmental Savings, 2021
[36] The Total Economic Impact™ of HPE GreenLake Flex Capacity, Forrester, 2020
[37] Global Survey of IT and Data Center Managers, Uptime Institute, 2020
[38] Cost of Data Center Outages, Ponemon Institute & Emerson Network Power, 2016
[39] Cloud-Based Predictive Analytics Survey, IDC, 2018
[40] Hewlett Packard Enterprise redefines mission-critical storage with new platform designed for the intelligence era, HPE, 2019
[41] Based on HPE Nimble Storage's installed base, 2021
[42] How the world's best artificial intelligence engine is redefining HPE 3PAR, 2019
[43] 100% Availability Guarantee, HPE, 2019–2021

## Brainstorm box

IT leaders should audit the annual electricity consumption of their existing infrastructure at the IT infrastructure and data center infrastructure level. Assess the energy efficiency of your current IT infrastructure* by completing the following tables.

| **Server infrastructure** | | | |
|---|---|---|---|
| Compute workloads | | | |

| **Consolidated (virtualized, containers, and more)** | **%** | **Non-consolidated** | **%** |
|---|---|---|---|
| Processing unit utilization (avg.) | % | Processing unit utilization (avg.) | % |
| Memory utilization (avg.) | % | Memory utilization (avg.) | % |
| Storage utilization (avg.) | % | Storage utilization (avg.) | |

| **Direct-Attached Storage** | **%** |
|---|---|
| Flask/Disk ratio | % |
| Compaction enabled storage | % |
| Average compaction factor | |
| Power Supply Ratings | |
| 80 Plus Bronze and above | % |

| **Storage Infrastructure** | **%** |
|---|---|
| Storage optimization | |
| Average used capacities | % |
| Flask/Disk ratio | |
| Compaction enabled ratio | % |
| Average compaction factor | |
| Power Supply Ratings | |
| 80 Plus Bronze and above | % |

| **Networking Infrastructure** | **%** |
|---|---|
| Network optimization | |
| Average network traffic load | % |
| Fiber/Copper ratio | % |
| Energy-efficiency Ethernet ratio | % |
| Average compaction factor | |
| Active/Powered port ratio | % |
| Power Supply Ratings | |
| 80 Plus Bronze and above | % |

| **Rack Infrastructure** | **%** |
|---|---|
| Power distribution | |
| Intelligent PDUs ratio | % |

* Another key element to consider for each infrastructure category is the operating temperature classification. This will be explored in the next section.

## b. Adjusting your data center infrastructure

Once IT leaders have achieved infrastructure efficiencies, the second step is to focus on the ancillary equipment, or the mechanical and electrical infrastructure.

It's no secret that IT hardware creates a large amount of heat while running. That means significant power, water, and cost goes into cooling it. Cooling systems can account for nearly half of total electricity usage in data centers and, to make this even more complex, studies[44] show that an increase in rack density means an increase in cooling requirements as well.

Cooling efficiency varies drastically between organizations. As previously explained, PUE improvements are stalling; however, there are still opportunities, especially in on-premises and colocation data centers, to lower PUE by minimizing the need for cooling.

Key opportunities to reduce the impact of cooling include:

- Reducing the cooling needs of the IT equipment by reducing heat generation, or
- Better utilizing the cooling by more efficiently separating hot and cool air, or
- Substituting traditional mechanical cooling with naturally generated cooling

In addition, digital transformation leaders and data center operators should consider water scarcity as an operational risk. Today, 1.42 billion people live in areas of high or extremely high water vulnerability.[45] The ICT industry is estimated to account for 0.2% of global water consumption, equivalent to more than 242 billion packs of mineral water (9 liters) per year.[46] Yet, only 50% of data center operators measure their water consumption, including that which is used by cooling infrastructure. Positioning data centers in non-water-stressed regions, as well as ensuring that technologies in place are using as little fresh water as possible, are both considerations that should be included in any comprehensive sustainable IT strategy.

This section will explore steps that IT leaders can take to optimize their cooling infrastructure by implementing efficient solutions and alternative cooling sources that will lower the environmental impact of the data center's ancillary equipment.

### Hardware cooling needs

It's important to keep in mind that IT equipment generates heat and will fail at critical temperatures. Therefore, the objective should be to allow the IT equipment to operate at the highest levels of performance with the least amount of mechanical cooling. To do so, there are two main questions IT leaders should ask themselves:

- Is the current operating temperature of the data center appropriate for the installed IT equipment?
- Does the data center contain excessive cooling?

For both questions, the American Society of Heating, Refrigeration, and Air-Conditioning Engineers (ASHRAE) guidelines can help. ASHRAE is the body that governs the standard for the accepted air temperature and humidity ranges for data center environments.

Reducing the need for and quantity of mechanical cooling is often a first step. Making sure that new equipment purchased will be able to operate in higher temperature environments is also important. For instance, HPE servers are designed to run in ASHRAE A3 and A4 environments (wider temperature and humidity spectrums, as stated by ASHRAE), which allows for operation in ambient air cooling (free cooling) environments, significantly reducing data center cooling power demand. In many cases, increasing data center temperature will raise the energy requirements of the IT system only marginally while saving the data center significant cooling costs.

### Making cooling more efficient with containment

Another way for IT leaders to drive their PUE down is to manage the airflow inside the data center efficiently. Establishing an airflow management strategy can cut energy costs, support the installation of additional compute nodes with the same amount of installed cooling capacity, and reduce the environmental footprint, including greenhouse gas (GHG) emissions and water consumption associated with cooling.

The basics of data center airflow management require separating the cooling air from the hot air returning form the IT equipment. If both airstreams mix, the efficiency of the system diminishes. Hot and cold air containment is an effective and well-known method to meet this objective.[47]

[44, 47] Global Survey of IT and Data Center Managers, the Uptime Institute, 2020
[45] Water Security for All, UNICEF, 2021
[46] The environmental footprint of the Digital World, GreenIT.fr, 2020

An additional consideration is the reuse of waste heat. This can enhance the efficiency and sustainability credentials of a data center even further while benefiting another organization that could use it for district heating or industrial purposes. There is also an opportunity to increase return on investment (ROI) by monetizing the heat reuse as some municipal and district heating schemes allow for financial compensation against heat injection in their grid.

**Alternative cooling sources**

Finally, when possible and available, sustainable IT strategies should include the use of alternative cooling sources, which are naturally generated and, therefore, reduce the cost of cooling the data center. Data centers have traditionally been cooled by compressor-based cooling systems, which are expensive and, as discussed previously, sometimes inefficient. Colder countries such as Sweden or Iceland have been using ambient air to cool their data centers when the external temperature is cool enough to create a sufficient temperature difference from the operating temperature of the IT equipment. This concept is called free cooling and can achieve very low PUEs. The two most used free-cooling system architectures are:

- **Direct free cooling (DFC):** The outside ambient air or natural water source is, after being filtered, directly used to cool the IT equipment through ventilation.
- **Indirect free cooling (IFC):** The outside ambient air or natural water source is used to cool the data center operating air through heat exchangers.

DFC has the theoretical advantage of maximizing the cooling potential of outside air or natural water sources like rivers and lakes, without the use of any additional cooling equipment. However, it does inconveniently require the purification of that air or water, including dehumidification and filtering to remove moisture, dust, and other polluting agents that could significantly increase operating costs and shorten hardware lifespan. A fresh air handling unit that combines all devices for this purification process is needed to make sure the DFC process is performant, but it can be expensive and reduce the cost benefits of free cooling if not properly sized and designed.

IFC finds a compromise by benefiting from the outside ambient air or natural water source cooling without endangering the data center inside air quality. This is done by expelling the hot air inside the data center to the outside, thereby, lowering the temperature of the air inside without mixing both air streams. Various heat exchanger designs follow this concept, one example being the rotating wheel exchanger.

The inconvenience of such a technique is the obvious loss in cooling efficiency when compared to direct free cooling. As a result, this technique is less suitable when outside temperatures are warmer and it often needs to be coupled with electrical chilling units to cover for higher temperature periods. Another inconvenience of such a system is the space required. A rotating wheel exchanger takes significant space because of the relatively low air-to-air heat transfer efficiency.

As in any project, the environmental benefits and impacts of the entire lifecycle need to be measured to achieve the intended result of lowering the environmental impacts of IT infrastructures. Many projects like these have been implemented without first assessing the potential environmental savings, resulting in sophisticated but unutilized cooling technologies.

**Liquid cooling**

Liquid cooling is increasingly considered as an alternative solution to ambient and free cooling, particularly for intensive workloads that necessitate the use of efficient heat dissipation technologies. Water and other specialty liquids used for liquid cooling are more efficient for heat and therefore are often used in rack densities above 50 kW. Many HPE HPC products are available with immersion or warm-water direct liquid cooling as a standard feature. In addition, HPE offers Adaptive Rack Cooling, an in-rack cooling solution for localized heat loads from high-density equipment using closed-loop liquid cooling.

There are, many other ways to improve the efficiency of data center auxiliary equipment, especially cooling, however, they are beyond the scope of this paper. Building innovation, insulation, automation, and immersed oil cooling are all technologies and techniques that could make a data center more resource efficient.

## 3. Looking beyond the data center: a lifecycle and science-based approach

At this point, IT leaders should have their own operations in order, or at least have a plan to do so. As discussed previously, for data center operators, the environmental impact produced by their operations is likely to comprise the largest portion of their footprint. That is why starting with IT sustainability and data center sustainability is the right approach. However, this doesn't mean impacts from other parts of the value chain shouldn't be addressed. For instance, at HPE, while it is widely recognized that targeting energy reductions in the product use phase will be most impactful, a set of other initiatives and targets have been defined across the entire value chain, including the supply chain and operations.

Although focusing time, effort, and money on the largest impacts first will be most strategic, it should be noted that impacts beyond the operational or financial control are increasingly scrutinized by customers and investors. The benefits are numerous:

- Mitigation of reputational risks associated with bad practices in supply chain
- Increased resilience of the value chain by identifying and remediating potential environmental, social, and corporate governance (ESG) risks
- Lower TCO by reducing waste in the value chain

---

**Brainstorm box**

Here are some useful questions to understand why your sustainable IT strategy should extend beyond the data center walls:

- What external stakeholders commercially depend on your organization and what level of understanding do you have of their activities, operations, and processes?
- If something unethical should occur in the supply chain, is there a way to identify this and react swiftly?
- Are the products, solutions, and services provided to the IT organization going to contribute or detract from the objectives of the sustainable IT strategy? If yes, how?

Are there supplier engagement processes in place that encourage continuous engagement on sustainability-related issues between all parties? Do all organizations have an opportunity to provide their expertise and knowledge with transparency?

---

This section aims to provide IT and digital transformation leaders with additional guidance to widen the scope of their sustainable IT strategies. The following section describes the importance of working toward a resilient and sustainable supply chain, while the second part focuses on the benefits of responsible asset disposal.

### a. Sustainable procurement

To achieve success in IT sustainability efforts, organizations need to build consensus both internally and externally. When asked if having an IT sustainability policy or program influences their vendor/partner selection process, over 70% of respondents prefer to work with a partner that has either demonstrated or communicated a commitment to sustainability.[48] This clearly shows how important choosing trustworthy partners is becoming when doing business and thinking sustainably. Having a responsible supplier base used to be a nice to have or a mere differentiator. It is now becoming a must-have with demonstrated business benefits.

Recognizing the increasing complexity and criticality of sustainable IT procurement, HPE developed a white paper that recommends general principles for the development of sustainable IT purchasing guidelines.[49] This white paper can aid organizations in integrating environmental, social and governance factors into their decision-making, alongside existing financial criteria. As explained in the paper, the criteria suggested are based upon the following principles:

- **Fairness and equitability:** Procurement criteria should be fair, equitable, and consider sustainability aspects throughout the entire lifecycle of a solution. Sustainable procurement should ensure fair and standard treatment for all.
- **Harmonization and recognition of international standards:** Because environmental standards vary across the global marketplace, it is important to recognize and support the harmonization of these standards and procurement schemes, particularly in the development of evaluation criteria, tools, and methodologies.

[48] IT's Role in Driving Sustainability Progress, IDC, July 2020
[49] Sustainable IT Purchasing Guidelines, HPE, 2021

- **Prioritization:** As explained in the first section of this white paper, the criteria chosen should reflect and align with the social and environmental priorities of the organization. Decision-making criteria must align with the organization's overall sustainability goals and prioritize procurement specifications that are based on reputable science.
- **Measurability and verifiability:** These criteria should be measurable and refer to an existing standard, comparable with competing products and verifiable by the purchaser.

The HPE Sustainable IT Purchasing Guidelines outline two attribute categories that should be considered to improve sustainable outcomes—product attributes and generic attributes.

### Product attributes

Product attributes cover a variety of topics such as efficiency, acoustics, materials, power consumption, and temperature monitoring. While all are important, eco-design is increasingly seen at the top of sustainable procurement officers' agendas. Eco-design is an approach that considers the environmental impact of the entire lifecycle of the product. It helps minimize the environmental footprint of a product by reducing energy and resource waste at the different stages of the product lifecycle—manufacturing, use, and disposal. A European Commission report[50] suggests that 80% of all product-related environmental impacts are determined during the design phase of a product, reinforcing this trend of designing with sustainability in mind.

HPE's eco-design program, Design for Environment (DfE), was created in 1992 and is still ongoing and evolving today. It is a holistic approach to product design that results in lower product power consumption and reduced waste, reducing TCO for users, and contributing to a more circular economy. The following principles guide HPE engineers when developing new solutions.[51]

- **Energy efficiency:** Innovating solutions that can accomplish more compute, storage, and networking with less power and with more dense architectures. This includes increasing energy performance, real-time thermal and power monitoring at the system level, dynamic power capping, and power proportionality, as well as power supply improvements to reach at least 94% efficiency. (See the IT infrastructure management and reporting tools section of part one of this paper for more).
- **Materials innovation:** Using less and better materials reduces product footprints. Restricting materials of concern, such as some flame retardants and plastic additives, allows HPE products to meet global regulatory requirements. This helps ensure that customers can purchase and use HPE products in any region of the world. This proactive removal of problematic substances helps maximize the ability for products to have residual value at the end of the customer's use, increasing value-return to users.
- **Longevity and recyclability:** HPE makes it easier for users to extend the useful life of IT products—designing for longer lifecycles and providing guidance on self-repair and upgrades. The Recyclability Assessment Tool (RAT) from HPE enables the calculation and improvement of the end-of-life recyclability of products. Based on these assessments, 90% of the material in HPE products is recyclable, on average.[52] Whereas many consumer goods are designed for a lifespan of only a few of years, HPE designs products for lifecycles exceeding five years.

### Generic attributes

Generic attributes cover product packaging, end-of-use services (covered in the asset end-of-use management section), leased and managed services (covered in the following section), supply chain responsibility, and organizational performance. While all topics are important, digital transformation leaders should continue to focus their efforts based on a data-driven and science-based approach based on areas of most significant impact. According to lifecycle analysis, after the use phase of data center products, the supply chain is where the largest environmental impact lies. Ensuring the supply chain behind the assets you purchase or use is aligned with both social and environmental best practices is of the utmost importance.

The key principles and initiatives that underpin our own program are described in HPE Supply Chain Responsibility—Our Approach. These principles and initiatives are easily transferable to other industries and organizations with key objectives to:

- Protect and elevate workers in the supply chain
- Reduce global and community environmental impact
- Benefit the company and its ecosystem

Moreover, transparency, collaboration, and broader business alignment should underpin any responsible supply chain program.

[50] Sustainable Product Policy, European Commission, 2018
[51] Design for the Circular Economy from HPE
[52] 2019 HPE Living Progress Report, HPE, 2020

**Transparency:** Openly reporting supply chain risks and social and environmental performance data results in a more robust and resilient supply chain. Transparency is critical to effective stakeholder engagement, informed decision-making, accountability, and progress toward cross-sectoral issues. For instance, HPE publicly reports key information about its suppliers and their social and environmental responsibility (SER) performance year-over-year, which helps hold them and their suppliers accountable.[53] Transparency can also inspire wider change by encouraging others within and outside the IT sector to adopt a similar approach.

**Collaboration:** Collaborating, by engaging with a broad range of stakeholders to research and better understand issues of concern regarding SER in our supply chain, is critical as well. These stakeholders include workers (through interviews, surveys, capability building programs, and ethics concerns reporting system), industry bodies, governments, socially responsible investors, and NGOs.

Stakeholder engagement is a critical step toward a coordinated and effective response to important social and environmental challenges. This includes working with industry peers and consortia to influence industry alignment and direction, as well as capability-building programs that address the most significant supply chain SER issues as identified by audit trends, external stakeholder input, and other intelligence. At HPE, supplier assessments are often paired with capability-building opportunities to facilitate improvement.

**Business alignment:** Finally, a responsible supply chain program is unlikely to succeed if it's not directly linked to the procurement process and connected to the business process. If the sustainability performance of suppliers does not have a direct influence on purchasing decisions and the award of a contract, suppliers won't be incentivized to perform better or to continue performing well.

ESG and/or SER criteria are increasingly being integrated into procurement scorecards. These criteria have increased from just 5% of RFP scores to more than 15 or 20% in recent years, with the remainder being left for quality and financial parameters.[54] In addition to integrating social and environmental criteria into our supplier scorecard, HPE has committed that 80% of our manufacturing suppliers, by spend, will have science-based emissions reduction targets for their operations by 2025.[55]

---

## Brainstorm box

To put this into practice, work with your procurement organization to create a questionnaire or checklist to evaluate potential IT equipment purchases.

Product attribute considerations may include:

- Does the supplier adhere to GreenScreen for Safer Chemicals method for chemical hazard assessment of their products?
- Are parts that must be recycled separately, easily separable?
- Are power and thermal monitoring capabilities embedded in the product?
- Generic attribute considerations may include:
- Does the vendor offer product repair, refurbishment, or recycling as part of the initial contract to guarantee proper handling at decommissioning?
- Are product packaging material types reasonable in weight and separable, and are they composed of multiple materials?
- Does the supplier have a third-party validation of sub-supplier audit results?

Further guidance and a sample questionnaire can be found in the HPE Sustainable IT Purchasing Guidelines.

---

### Sustainable cloud procurement

With cloud services taking an increasing share of IT budgets, it is important to understand how to assess their sustainability performance to make the right decision. There is a general assumption that outsourcing IT to a public cloud is a net positive environmental move. But how can digital transformation leaders be certain and show evidence of such reduction? How can an IT leader confidently report on the environmental footprint associated with their outsourced IT services?

Cloud providers have been reluctant to disclose the environmental performance of their services. In the fall of 2017, the Green Electronics Council (GEC) sponsored an Arizona State University graduate capstone project that found that cloud service providers' publicly available sustainability data was limited and that the sustainability metrics and terminology used by those providers were inconsistent and confusing.[56] Third-party studies[57] show that on average, public cloud providers do perform better environmentally, however, they are not far ahead from the average enterprise data centers and sometimes perform worse than best-in-class enterprise data center performers.

[53] Hewlett Packard Enterprise Suppliers, 2020
[54] Based on HPE internal data, 2021
[55] 2020 Living Progress Report, HPE, 2021
[56] Purchasers Guide for Sustainability and Cloud Service Procurements, GEC, 2019
[57] Beyond PUE: Tackling IT's wasted terawatts, Uptime Institute, 2020

Cloud providers are usually less transparent when it comes to providing environmental information associated with services provided to their customers. They usually report their corporate environmental footprint, rather than a customer workload footprint. Breaking this down per functional unit (customer, application, server, virtual machine, m2, and such) is virtually impossible to determine using the data provided by the public cloud vendor. The reality is the benefits of outsourcing usually come with the inconvenience of losing oversight of the environmental data that organizations are accustomed to accessing. For instance, whereas it is relatively simple to access your data center's power and water consumption via utility bills, determining the power and water consumption associated with the services from a public cloud data center that hosts hundreds or even thousands of different customers is much more complex.

To make it even more complex, 85% of organizations are now using multiple clouds in their business,[58] using different cloud procurement models depending on their applications, and procurement strategies (IaaS, platform as a service, and software as a service). This makes it increasingly challenging for digital transformation leaders to assess the environmental footprint of such services.

Nevertheless, this shouldn't stop digital transformation leaders from requesting that their chosen cloud provider report the information needed to assess their environmental footprint. In fact, in some instances, this can be an essential requirement of the purchasing process. Digital transformation leaders must remain accountable for the environmental impact of their IT organization even when outsourcing. The environmental impacts of these services remain, and the out-of-sight, out-of-mind mentality should not stand. These impacts are simply relocated somewhere else, in another data center, and managed by different people. In GHG accounting terms, outsourcing simply changes the category for which the impacts will be associated, usually from scope 1 and 2 (direct operations) to scope 3 (upstream and downstream). IT organizations should treat cloud providers the same as any other service provider and hold them accountable for their ESG performance.

techUK's Lost in Migration: Attributing Carbon to the Cloud when outsourcing to data centers and cloud paper,[59] helps identify ways in which digital transformation leaders can understand the impacts of outsourcing IT to third-party vendors, or at least estimate them robustly enough to inform their decision-making. The paper details four different options, depending on the maturity of your cloud service provider and of your organization, to come up with a roughly accurate idea of the GHG emissions associated with third-party IT services:

1. Asking the supplier
2. Calculating the emissions
3. Estimating the emissions
4. Comparing it to existing literature

techUK's paper then details the different tools and guides that are available to digital transformation leaders should they chose any one of these four options. The first option will typically yield the most accurate assessment and the fourth the least accurate, dependent on the quality of the data.

Although the paper does not account for environmental impacts other than GHG emissions, the methodology and the concept suggested should also work for other factors such as water. If digital transformation leaders are willing to rely on their supplier to provide this data (option 1), it is preferable to link this requirement to the contract, at renewal or as part of the tender, to leverage purchasing power as a tool for negotiation. The GEC published a Sustainable Cloud Services Purchaser Guide[60] in March 2019, to which HPE contributed as well, that outlines the right questions to ask suppliers when it comes to tendering these third-party IT services. The questions are divided into three main categories:

- Sustainability attributes of the public cloud providers, at a corporate level
- Sustainability attributes of the cloud services provided, at a data center and infrastructure level
- Sustainability attributes of the cloud services provided, from an energy source perspective

In a world where more than 90% of businesses have multiple cloud environments in place,[61] it is paramount to integrate a sustainability dimension to ensure it is indeed an environmental net positive gain and not a way to displace the issue elsewhere.

[58] 85% of companies now operating in a multi-cloud environment, IBM, 2019
[59] Lost in Migration: Attributing Carbon to the Cloud, techUK, 2019
[60] Purchasers Guide for Sustainability and Cloud-Service Procurements, GEC, 2019
[61] Going hybrid: Demand for cloud and managed services across Asia Pacific, 451 Research, 2019

## b. Asset end-of-use management

A critical part of any lifecycle management strategy is the appropriate disposition of end-of-use assets being replaced. Circular economy principles provide a helpful framework to inform end-of-use decisions to minimize the environmental impact. Technology refresh strategies need to consider three primary areas of risk:

- **Security and compliance:** Is the asset disposal following cybersecurity best practices and applicable government privacy regulations?
- **Environmental:** Is the equipment processed following labor, chemical substances, e-waste, and environmental regulations?
- **Economic:** Is the economic model transparent, scalable, and replicable?

At end of use, assets can either be recycled or upcycled, in whole or in parts. Without examining the specific benefits of recycling versus upcycling, upcycling is generally less energy intensive than recycling. Upcycling opportunities for the system, or its parts, are correlated to the asset age, configuration, and associated secondary market demand. When IT assets cannot be refurbished and reused due to age or disrepair, they should be recycled under applicable regulations and best practices. The appropriate processing of e-waste is a growing challenge that has been acknowledged by the United Nations and many governments worldwide.

Recycling or refurbishing assets through an organization that lacks adequate certifications, asset tracking, and assurance programs throughout the commodity lifecycle pose significant reputational risks for organizations.

---

**Brainstorm box**

IT leaders should ensure that a responsible and secure asset decommissioning policy is in place across the organization. This will require a partnership with a credible vendor that will reduce risks related to asset disposition. Consider the following when selecting a vendor:

- Select a service provider that offers global reuse and recycling services and has specialized skills that encompass secondary market knowledge and data privacy concerns
- Select specialized vendors, such as HPE Financial Services, who can offer detailed reporting regarding the fate of your assets. For instance, are you able to report the following data:
  - Total reuse of decommissioned assets (metric tons, approximate)
  - Total recycling of decommissioned assets (metric tons, approximate)
  - Total metric tons of waste diverted from landfill
  - Total emissions reductions from reuse versus recycling

---

## c. Decarbonization

According to the World Economic Forum's Global Risk Report, climate change and the failure to act are considered to be the most critical threat to humanity and the planet in the field of environmental sustainability, and the second top global risk in any category.[62] The effects are already being felt around the world and, according to the U.N., climate change is already the cause of over 150,000 deaths annually.[63] Given these threats to business and humanity alike, decarbonization should be one of the key environmental priorities of a sustainable IT strategy. These strategies are usually divided into two workstreams:

- Drive efficiency across the value chain to reduce the need for resources, for example, reduce the PUE ratio of the data center by investing in a more efficient cooling technology.
- Use low-carbon resources that help minimize the carbon intensity of activities, for example, use low-carbon energy to power the data center.

This paper thus far has primarily covered the first workstream, demonstrating the need to perform more work with less hardware (equipment efficiency), less power (energy efficiency), and less ancillary equipment (resource efficiency). Unfortunately, digital transformation leaders can only drive efficiency to a certain viable limit, and they will inevitably need to consume resources and create environmental impacts to run their operations. For instance, even if the PUE ratio reaches 1, energy must still be consumed to power the hardware stack. That's where the second low-carbon workstream kicks in to further reduce the environmental footprint.

This section will focus on low-carbon energy, however, it is important to acknowledge that other resources consumed by data centers, such as water, also have environmentally friendly alternatives. As an example, Apple and Google have both used treated sewage effluent (TSE)

[62] Global Risks Report, World Economic Forum, 2021
[63] Climate change, World Health Organization, 2021

water to in of their data center's evaporative cooling system,[64] as has the U.S. National Security Agency.[65] In Apple's case, this innovation was estimated to save nearly 5 million gallons of fresh water per year.[66]

While this method is not as well-known as free cooling, it demonstrates that alternatives exist. If recycled water is not available or a viable option, some water streams may be less carbon intensive than others for a variety of reasons beyond the scope of this paper. In any case, IT leaders should partner with their suppliers to determine what less carbon-intensive water supply options exist.

The same is true for power. Renewable and low-carbon energy supplies are becoming increasingly available to a larger number of users and geographies. First, it is important to know that data center operators will typically need to use the grid as a provider of renewable energy as they are unlikely to be able to produce all of the energy needed on-site (for example, via solar panels). This is due to the small energy density per surface unit of renewables.

Second, the energy supply will usually be sourced only in part from renewable or low-carbon sources. This is often the case for the simple reason that renewable sources are difficult to store and, therefore, have intermittency challenges. When that's the case, less sustainable energy sources might be used.

The Intelligent Low-Carbon Power Sourcing for Data Centers white paper by BSR guides in understanding the complexity of sourcing this type of power for data centers. It first describes the different risks and opportunities of low-carbon sources associated with different types of data centers (cloud, enterprise, managed, and rented). Second, it suggests a strategy for data center operators and digital transformation leaders to source low-carbon energy, which includes aligning with a corporate will, asking the right questions, assessing the options available and prioritizing opportunities.

In addition, the best practices guidelines of the EU Code of Conduct for Energy Efficiency in data centers (part 3.2.8), details the different EU standards that can guide decision-making. For instance, the EN 50600-4-3 standard mentions the renewable energy factor (REF) as the ratio to use when disclosing the renewable energy share used in a data center and how to calculate it.

## d. Edge computing and decentralized architectures

Edge computing is increasingly seen as the architecture of the future. According to Gartner, by 2022, more than 50% of enterprise data will be created at the edge.[67] After a race toward centralization with public cloud adoption, organizations and experts are finding that computing data and gaining insights where the data is created, at the edge, offers several benefits. In short, edge computing is a distributed computing paradigm that brings computation and data storage closer to the location where it is needed. The benefits include:

- **Low latency:** For applications that are time sensitive and require sub-millisecond latency, moving the compute node where the action is, at the edge, is a critical need. With the elimination of the data round-trip to the centralized architecture, edge architectures usually outperform traditional cloud models.
- **Connectivity:** Edge architectures are usually preferred in remote locations where there is weak to no connection to a centralized system.
- **Security:** Sensitive information security protocols may prohibit data from being hosted in the cloud. By staying at the edge, they remain in a physical perimeter that data owners fully control.
- **Resilience:** By not aggregating large amounts of data in the same place, there is an element of geographical redundancy with edge architectures that can help mitigate outages that a centralized system would be unable to overcome.

There are few papers addressing the environmental benefits of decentralized versus centralized architectures. The Institute of Electrical and Electronics Engineers published a white paper in 2019 titled Estimating Energy Consumption of Cloud, Fog, and Edge Computing Infrastructures.[68] The paper compares the energy consumption of different architectures, more or less centralized, to guide policymakers and organizations in their decisions.

The results are striking. Because distributed systems don't require large cooling systems and eliminate a large portion of the energy consumption associated with network and intra-data centers data roundtrips, fully decentralized architectures could save between 14 and 25% power when compared to fully centralized architectures. On the other hand, some worry distributed architectures and the rise of the Internet of Things (IoT) and edge devices could increase the amount of electronic waste and make it even harder to trace.

[64] Google details its use of treated wastewater at Georgia data center, The Verge, 2012
[65] NSA Will Cool its Secret Servers With Waste Water, Data Center Knowledge, 2014
[66] Apple's newest innovation: Wastewater treatment to cool Prineville data centers, Oregon Live, 2016
[67] Top 10 trends impacting infrastructure and operations, Gartner, 2020
[68] Estimating Energy Consumption of Cloud, Fog, and Edge Computing Infrastructures, IEEE, 2019

## 4. Conclusion

The emergence of a hybrid cloud, digital world requires IT leaders to develop a holistic strategy that drives efficiencies across their entire IT estate, whether on-premises, in the public cloud, or a colocation. As we have outlined, enterprises focused on their digital transformation strategies are realizing that as they increase efficiencies and optimize their infrastructures, they are also minimizing their environmental impacts—in other words, doing more with less. Although each organization is unique, significant efficiency opportunities likely remain untapped and all organizations can benefit from embedding sustainability and efficiency principles in their IT strategies, whether to accelerate their digital transformation, reduce risk, or unlock innovation.

This paper recommends that IT leaders begin by assessing inefficiencies within their owned IT estates, looking beyond PUE to the efficiency of IT hardware, which has the biggest potential for environmental and financial savings from reduced energy use. Then, a wider set of initiatives should be implemented across the value chain to ensure ethical and sustainable practices in the supply chain and at end-of-use. In many cases, the transition to an as-a-service IT delivery model can offer a low-touch solution that takes care of operational efficiencies and lifecycle management while offering flexibility and control. With the right technologies, processes, and business models, any IT organization can plan an active role in implementing a sustainable digital transformation that increases the competitiveness of their enterprise.

## Learn more at

HPE.com/us/en/living-progress/sustainable-it.html

Chat now (sales)

Hewlett Packard Enterprise



a50004223ENW, Rev. 2