Technical white paper

Check if the document is available in the language of your choice.

DELIVERING DIGITAL TRANSFORMATION EFFICIENTLY AND SUSTAINABLY

Part 2: Analysis of the environmental impact of ITaaS adoption

CONTENTS

INTRODUCTION

Digital transformation is a pervasive trend touching all industries and with it, the approach to IT delivery is evolving. According to a survey conducted by 451 research, 57% of enterprises are adopting a hybrid multicloud IT delivery strategy.[1] Yet, cloud customers waste up to 45% of their spend on provisioning resources they never end up using.[2] IT as a service (ITaaS) can address the key causes of inefficiencies in on-premises or collocated data center IT equipment, while establishing the right business processes to execute an efficient hybrid multicloud delivery model.

This yields not only financial and operational benefits for an IT organization, but can also address environmental concerns, which are increasingly central to the decision-making process of investors, consumers, and corporates alike. Green bonds or sustainability performance-linked financial products give organizations renewed incentives to assess material environmental, social, and governance (ESG) factors, set measurable goals and report progress. By removing information silos and giving access to core business data that informs materiality and goal settings, IT departments have a foundational role to play in an organization's sustainability journey.

The first part of this white paper defined ITaaS and highlighted how this approach helps organizations to effectively deliver on their hybrid multicloud IT strategy. It briefly touched upon the sustainability benefits, such as the associated reduction in use of resources and electricity.

This second part examines, in more detail, the environmental impacts of shifting to ITaaS. It emphasizes how ITaaS represents a leap toward circular IT and a shift toward supplier responsibility, whereby wasted infrastructure is eliminated as the manufacturer maintains custody over equipment and manages assets at end of use.

ALIGNING WITH CIRCULAR ECONOMY CONSUMPTION PRINCIPLES

A circular economy is an industrial system that decouples economic growth from the consumption of natural resources. It is restorative by intention and design; shifts toward the use of renewable energy; eliminates the use of toxic chemicals, which impair reuse; and aims for the elimination of waste through the superior design of materials, products, systems, and business models.[3]

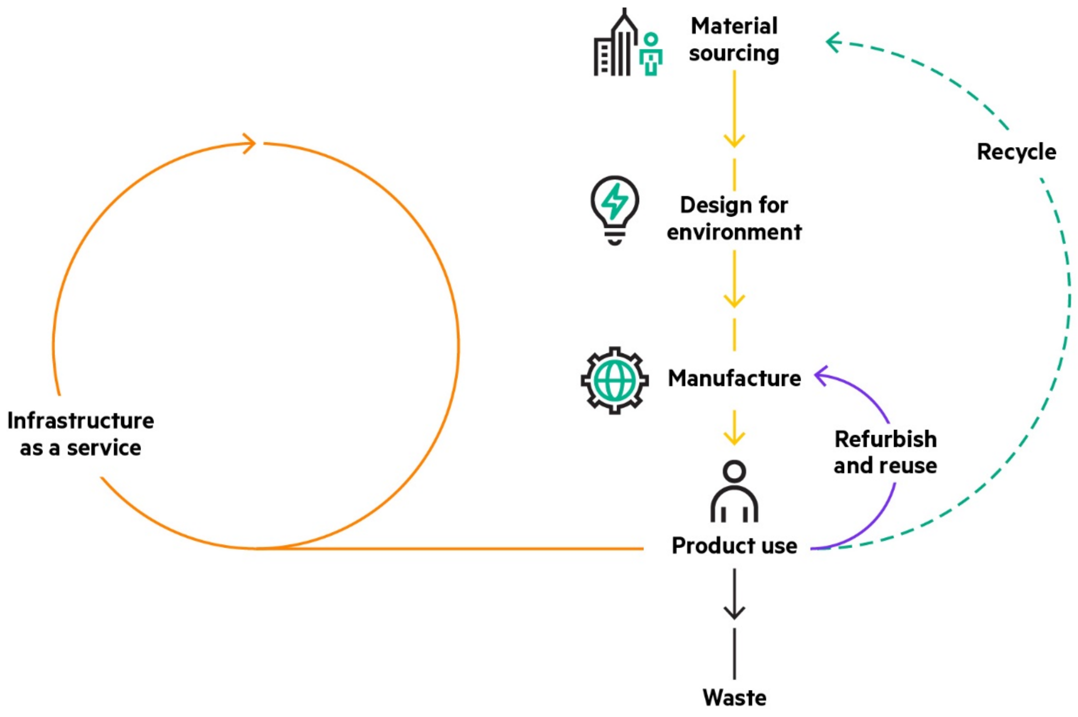

FIGURE 1. HPE circular economy approach[4]

[1] Having your cake and eating it too, 451 Research, 2019
[2] cloud.withgoogle.com/build/infrastructure/public-cloud-pricing-explained
[3] Ellen MacArthur Foundation
[4] h20195.www2.hpe.com/v2/Getdocument.aspx?docname=a00099273enw&skiphtml=1

As global demand for connectivity and computing power grows exponentially, information and communications technology (ICT) infrastructure is consuming energy and materials at unprecedented rates. ITaaS is a business model that gives an incentive to both parties in the service agreement to avoid overconsumption, assumes a comprehensive level of repair options, and builds-in reuse or recycles at the end of term. To that extent, an ITaaS agreement aligns with principles from the circular economy.

- **Material waste**—ITaaS avoids waste by avoiding overprovisioning or consumption of unnecessary assets (as discussed in part one of this white paper). When avoiding unnecessary assets, organizations also avoid the embodied carbon coming from extractions, logistics, and manufacturing of the equipment.
- **Repair and upgrade options**—Through its personalized maintenance service, ITaaS ensures equipment will be repaired or upgraded with spare parts to extend its lifecycle.
- **Refurbishment or recycling at end of use**—ITaaS assumes IT assets are refurbished or, when necessary, recycled at end of use. The operational processes leverage established supply chain processes. The consistency of the end-of-use governance processes with the design and manufacturing processes ensures a high level of accountability and traceability on the fate of the assets. This in turn ensures compliance with environmental and human rights regulations and protects the reputation of the organization consuming the ITaaS solution.
- **Data transparency**—ITaaS reports data transparently and specifically, giving access to financial, environmental, and asset management information that can demonstrate a company's IT efficiency and sustainability leadership.

An ITaaS approach delivers a personalized maintenance and support experience. This has allowed some organizations to gain confidence in deploying certified pre-owned equipment when their platform of choice was no longer available, or during workload or data center migrations, thereby participating in the re-use economy within the larger circular economy model.

FROM ENVIRONMENTAL IMPACT TO ENERGY IMPACT

Enterprise IT equipment has a significant environmental impact resulting from the materials, manufacturing, and logistics required to deliver the equipment to a customer. This is known as the embodied impact. However, the largest environmental impact of this equipment results from the customers' use of the asset, better known as the use-phase impact—which accounts for over 80% of the compute carbon footprint.[5, 6]

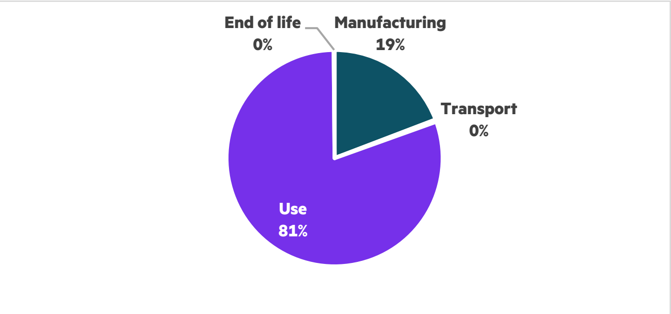

FIGURE 2. Average rack-mount server carbon footprint throughout its lifecycle

Once deployed at a customer site, the environmental impact of data center IT equipment is primarily due to electricity usage to power the IT and cooling equipment, as well as water consumption for evaporative cooling. The more equipment that is deployed, the more power and space it will consume, and consequently, the more water and energy will be required to cool it down. In this section, we focus on the primary environmental impact of transitioning to ITaaS: energy. This serves as a proxy for the broader environmental footprint given that energy consumption has a direct impact on cooling water and other resources consumption.

[5] Internal HPE data following the Product Attributes to Impact Algorithm (PAIA) evaluation methodology, 2019
[6] msl.mit.edu/projects/paia/main.html

ENERGY IMPLICATIONS OF ITAAS

Eliminating comatose equipment

IT equipment that is powered-on in the data center but not performing useful work, sometimes known as zombie equipment, is a known challenge for IT operations managers. Although IT professionals have been aware of the magnitude of zombie IT equipment in data centers for over a decade, studies continue to find that 20–30% of servers in data centers are unproductive.[7]

Significant power savings can result from the removal and recycling of this obsolete IT equipment. For instance, winners of an annual Uptime Institute Server Roundup contest seeking to reduce data center energy use achieved impressive results, including reducing power consumption by 2.5 megawatts through the removal of 9000 servers.[8]

To help organizations take the necessary steps to reduce zombie equipment, The Green Grid developed a white paper entitled Solving the Costly Zombie Server Problem, Detecting and Preventing Resource Inefficiency and Waste in the IT Estate.[9]

Increasing server, storage, and network utilization

Servers

In its paper, Beyond PUE, Tackling IT's wasted Terawatts, the Uptime Institute identifies increasing server utilization as one of the few major energy saving opportunities in IT, alongside optimizing the server refresh lifecycle, right-sizing redundancy requirements, and consolidating the infrastructure.[10]

When operating at low load, data center servers are inefficient, as highlighted by Figure 3. The SPEC Power performance benchmark[11] was developed by an industry organization known as the Standard Performance Evaluation Corporation Server, to measure performance per watt, known as energy effectiveness. The benchmark measures energy consumption at various load levels. Figure 3 displays the results published by all participating vendors from the SPECpower_ssj® 2008 benchmark since its inception in 2007 to January 2020, comparing power consumption at idle, 10% and 50% of the target load versus the power consumption at 100% of the target load.

The results highlight:

- A significant improvement in power effectiveness between 2007 and 2020
- A dramatic range of effectiveness profiles between the different systems tested
- While the most recent and energy effective systems can draw as low as 9% of the target load power when idle, the average power draw at idle for systems tested 2018–2020 is still high at 20%
- At 10% load performance, which fits the profile of comatose servers, the power draw ranges from 20% to 50%, with a recent average of 35%

This data indicates that although power effectiveness has improved, idle servers remain a key inefficiency, drawing a significant portion of their full load power while doing no useful work.

[7] anthesisgroup.com/report-zombie-and-comatose-servers-redux-jon-taylor-and-jonathan-koomey/
[8] datacenterknowledge.com/archives/2014/05/12/barclays-fires-9000-idle-servers-data-centers
[9] thegreengrid.org/en/resources/library-and-tools/456-WP
[10] uptimeinstitute.com/beyond-pue-tackling-it's-wasted-terawatts
[11] spec.org/cgi-bin/osgresults

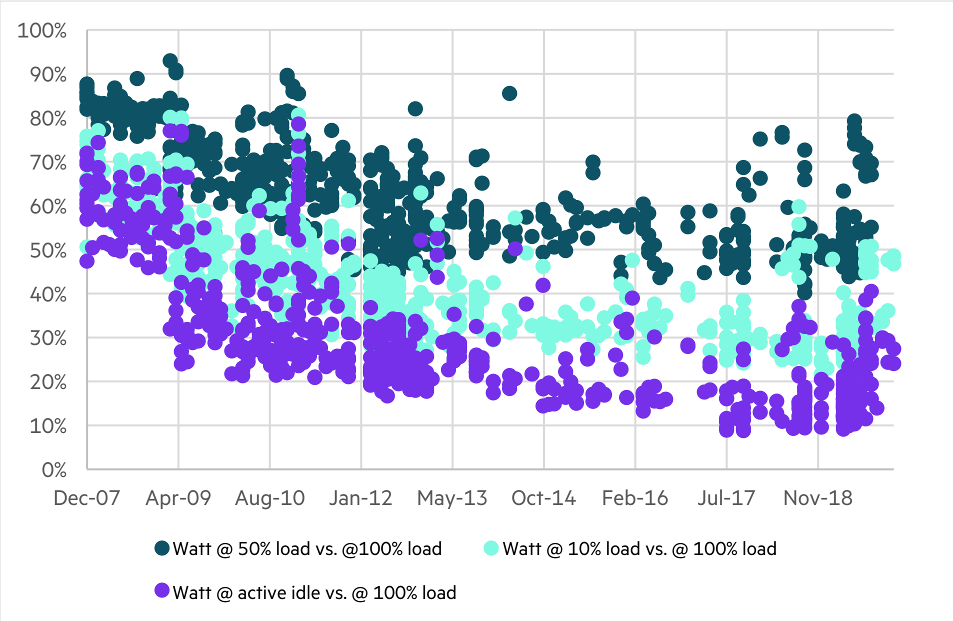

FIGURE 3. SPECpower_ssj 2008 public results over time

Storage

Data center storage and cooling for storage is estimated to consume over 25% of the power in an average data center, yet there is no industry standard benchmark that provides power efficiency insights for storage arrays.[12] The ENERGY STAR® label, which relies on a standard test suite developed by the Storage Networking Industry Association (SNIA), partially addresses that gap but does not disclose detailed benchmark results. This void requires IT departments to make decisions based on vendors' dimensioning tools or the advice of storage consultants.

The absolute power consumption of storage arrays increases as the quantity of media increases. Capacity per watt or IOPS per watt are metrics that can be used to assess power effectiveness of storage arrays and are usually optimized when the arrays are fully populated, as the storage array controller consumption overhead is distributed over its full capacity. Figure 4 illustrates a more than 6X power effectiveness improvement when an array is at full capacity.

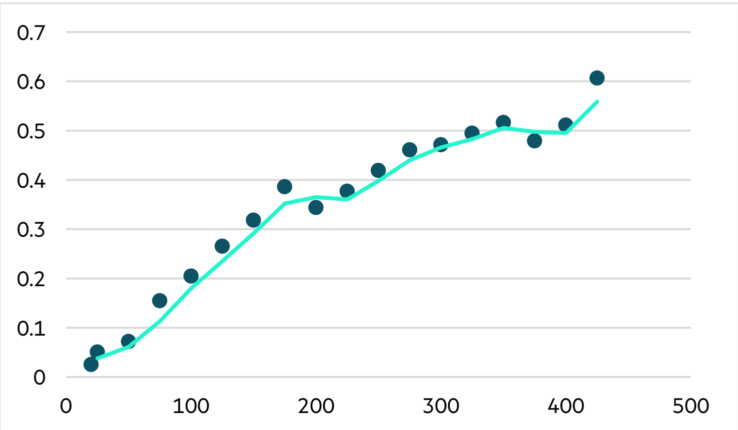

FIGURE 4. Example of effective TB per watt in a storage array as effective capacity grows

[12] Kim and Rotem, "Energy Proportionality for Disk Storage Using Replication"

Network

Enterprise networks consume about 8% of the overall ICT power consumption.[13] More than power effectiveness, the need to cope with growth in data and types of devices, while delivering an expected performance level and avoiding congestion, are the factors that drive network deployments, particularly at the edge. A context of growth that is difficult to forecast presents the risk of overprovisioning and sub-optimal use of available resources.

A network-as-a-service approach enables organizations to:

- Leverage best practices in network deployment, as well as management and software tools to deliver the best connectivity experience by user tier while avoiding overprovisioning
- Track the actual usage patterns to inform forecasting and planning decisions
- Ensure energy saving features are effectively configured and implemented

Conclusion

The data presented here highlights that for server and storage technologies, power proportionality, or power consumption effectiveness improve when equipment is used at above 50% of its full load capacity. By eliminating overprovisioning, ITaaS targets hardware utilization levels of 80%. ITaaS, through a metered infrastructure approach and backed by detailed reporting tools, therefore, addresses the energy efficiency challenges of overprovisioned and under-utilized hardware. While not essential, having clarity on the start and end-state estimated required capacity is important to ensure optimal efficiency of the systems deployed.

When metering and charging are based on a workload level, organizations may lose track of the actual underlying hardware utilization. However, through reporting and software management tools such as HPE OneView, HPE InfoSight, and Aruba NetInsight, as well as a closed loop approach to reviewing dimensioning assumptions, organizations can leverage the ITaaS governance framework to measure and monitor actual infrastructure utilization. This approach can significantly reduce and even eliminate, overprovisioning.

Optimizing the refresh cycle

ITaaS agreements include an optional technology refresh at the term of the contract. The timing is flexible to adapt to the profile of the specific technology and workload.

Research by The Uptime Institute found servers older than three years old make up 40% of the installed base, yet they perform less than 7% of the work while consuming 66% of the power.[14] Being intentional in the lifecycle management of their IT estate enables IT departments to take advantage of performance improvements and reduced footprint from newer technologies. This in turn can yield material energy effectiveness improvements. The white paper Efficiency Benefits and Challenges of IT Lifecycle Management provides a comprehensive assessment of the benefits of optimizing technology refresh lifecycles.

Through its contractual structure and the technology refresh option, ITaaS addresses energy consumption challenges associated with sub-optimal asset lifecycle management.

[13] mdpi.com/2071-1050/10/9/3027
[14] intercompbusiness.com/wp-content/uploads/2019/06/Why-Faster-Refresh-Cycles-And-Modern-Infrastructure-Management-Are-Critical-To-Business-Success.pdf
uptimeinstitute.com/beyond-pue-tackling-it%E2%80%99s-wasted-terawatts

Overprovisioning due to redundancy requirements

Redundancy requirements can be driven by application availability or disaster recovery (DR) requirements. Because of those requirements, more equipment gets deployed than the workload traffic requires to address full or partial failure of a system or subsystem.

High availability

Hardware-level high availability (HA) is usually achieved by provisioning redundant components, such as power supplies, fans, or CPUs.[15] Hardware HA, therefore, inevitably consumes more resources.

At the software level, HA architecture approaches can be different between legacy and modern cloud-native applications but will often leverage technologies that distribute the workloads over several nodes, such as clusters or load balancers. This approach helps minimize the redundancy overhead to achieve the targeted availability percentage. Some applications can achieve 99.999% availability based on an n+1 hardware node deployment.

ITaaS generally presents the same efficiency opportunities for highly available applications as it does for applications that do not have a HA framework, as it will bring up the utilization level of the baseline infrastructure. However, in some cases, ITaaS will allow organizations to achieve 100% availability while minimizing the number of redundant systems deployed. Technology such as HPE InfoSight that leverage artificial intelligence (AI) to enable autonomous operations, combined with services, allow organizations to achieve 100% availability guarantee for a single storage array.[16] This approach combining services and AI-driven IT operations drives a very efficient use of deployed resources.

Disaster recovery

Efficient DR starts with establishing recovery point objective (RPO) and recovery time objective (RTO) for various types of data and applications. This drives the decisions regarding DR architecture and tools. For applications requiring the most aggressive RPO and RTO metrics, adopting active-active or active-active-active DR architectures can be more energy-efficient. In such configurations, it is possible to keep utilization levels above 50% through careful DR requirements segmentation and prioritization.[17] This approach drives an optimization of the redundancy needs and, therefore, minimizes the environmental impact.

Where a longer RTO is acceptable, an active-cold-standby option can also be a way to minimize the energy impact from DR requirements.

In these scenarios, an ITaaS delivery model again ensures the baseline utilization level is maximized and that DR requirements are not themselves plagued by overprovisioning. In this case, ITaaS offers an additional benefit of addressing the requirements for robust governance processes, which underpin DR planning and execution.

While DR requirements certainly justify a certain level of overprovisioning, the energy consumption impact can often be optimized through careful segmentation of the data and an aaS approach.

ITaaS and data center facilities considerations

The energy effectiveness of data center facilities, often measured as power usage effectiveness (PUE) at the data center level, has a significant influence on the environmental impact of IT. Data center PUE improvement opportunities should, therefore, be considered when aiming to improve overall IT energy effectiveness and minimize its environmental impact. However, many companies focus PUE improvement efforts on the facilities infrastructure and ignore the efficiency of the IT equipment. Energy effectiveness improvement at the facilities-level needs to go hand-in-hand with efforts to improve the efficiency of the IT load.

ITaaS can be delivered in an organization's owned on-premises data center or in a co-located data center. In the latter scenario, ITaaS can sometimes include in its scope the data center facilities, management processes, and costs. From an energy impact and business process perspective, this is a good way to ensure operational focus on both sides of the PUE metric: data center energy effectiveness and IT load utilization levels.

In a scenario where ITaaS encompasses the data center facilities, while not necessarily part of the immediate services scope of work, a road map can be established to address energy effectiveness opportunities in the cooling, power distribution, and through the adoption of monitoring and automation tools.

In the mid- to long-term, bundling IT and data center infrastructure in one aaS contract, brings opportunities to align the data center design strategy with the hybrid and digital transformation strategy. This approach enables organizations to stay ahead of power density and cooling requirements, foster innovation, and leverage energy effectiveness best practices.

[15] hpe.com/us/en/servers/nonstop.html
[16] h20195.www2.hpe.com/V2/GetDocument.aspx?docname=a00074521ENW
[17] redbooks.ibm.com/redpapers/pdfs/redp5109.pdf

HYBRID AND MULTICLOUD ENERGY EFFICIENCY

As previously mentioned, the majority of IT architectures today assume a hybrid delivery mode that comprises on-premises resources for certain workloads and shared or dedicated cloud-based infrastructure for others, as well as platform and software as a service.

Shifting some workloads to a third-party service provider does not bring large energy savings opportunities other than operating in more efficient data centers. However, cloud customers waste up to 45% of their spend on resources they never end up using.[18] This is because cloud business models do not consider a KPI that focuses on maximizing infrastructure utilization and right sizing the underlying hardware equipment. If not managed with the right governance processes, cloud deployments can cause IT sprawl and increased energy consumption.

Given that oversubscribed idle instances are a significant business opportunity, cloud service providers should ideally have actual infrastructure utilization levels that are higher than in a traditional delivery model for the provider to have a successful and efficient business model. Access to the actual utilization level information, however, is difficult. Research indicates that average utilization of hyperscale data centers infrastructure, while aiming for 40–70%, actually fluctuates between 10 and 50%. As mentioned in the book The Datacenter as a Computer, "this activity profile turns out to be a perfect mismatch with the energy efficiency profile of modern servers in that they spend most of their time in the load region where they are most inefficient."[19]

Hybrid deployment models, on the other hand, do present efficiency improvement opportunities when governance and control challenges are addressed as part of the original design phase. To mitigate the inefficiency risk of cloud resources sprawl, organizations need to have business processes in place that focus on continuous compliance and cost controls. Continuous compliance controls ensure that hybrid deployments are meeting security and data governance requirements; these are critical to hybrid cloud deployment success but are beyond the scope of this paper.[20] Continuous cost controls can help inform policies, assign budgets based on business forecast, and monitor expenses against the allotted budget. Having the cost control processes in place will identify any deviation from plan, which is a first step toward tracking resources sprawl and inefficiencies.

While these measures do not address the overall cloud service provider efficiency, they do tackle what is in direct control of the cloud user.

SUMMARY AND CONCLUSIONS

In conclusion, an ITaaS delivery model addresses the top three areas of concerns responsible for inefficient on-premises or collocated data center IT energy consumption: zombie equipment, over provisioning, and lifecycle management.

Additionally, an ITaaS approach can establish the right business processes to execute an efficient hybrid multicloud delivery model, eliminating sprawl in the multicloud environment through a strong governance framework, including budget allocation and compliance metrics.

Digital transformation initiatives allow organizations to identify and address efficiency opportunities and drive innovation. Efficiency improvements serve both business and sustainability objectives by making better use of available resources and gradually engineering waste (whether of energy, time, or material) out of the business processes.

Through digital transformation consumed aaS, IT departments are, therefore, in a unique position to drive both infrastructure efficiency, as well as business efficiency and corporate sustainability.

To take advantage of the full potential for operational, resources, and energy efficiency offered by an ITaaS model, enterprises need to partner with vendors that offer them transparent reporting tools that will allow them to have a clear line of sight to actual resources deployment and utilization, across their hybrid multicloud estate. Under those circumstances, ITaaS delivery models, by focusing vendors incentive on right-sizing solution deployments and revisiting lifecycle decisions at specific checkpoints, present an effective path forward to a more efficient and sustainable IT.

[18] cloud.withgoogle.com/build/infrastructure/public-cloud-pricing-explained
[19] nrdc.org/sites/default/files/data-center-efficiency-assessment-IP.pdf
nrdc.org/sites/default/files/NRDC_WSP_Cloud_Computing_White_Paper.pdf
web.eecs.umich.edu/~mosharaf/Readings/DC-Computer.pdf
[20] hpe.com/sg/en/services/cloud-services-partners.html

LEARN MORE AT

hpe.com/us/en/living-progress.html

a50002144ENW, July 2020