Deloitte.

Enabling the data center of the future

The case for on-premise XaaS data centers

Introduction

The accelerated and sustained growth of cloud services adoption has transformed the world of information technology (IT) over the past decade, repositioning enterprise IT—previously considered a pure support function—as a primary player in operations, new product development, and in-house innovation hubs. Indeed, the benefits that the cloud has helped catalyze across end users and small and medium enterprises are numerous, substantial, and undeniable.

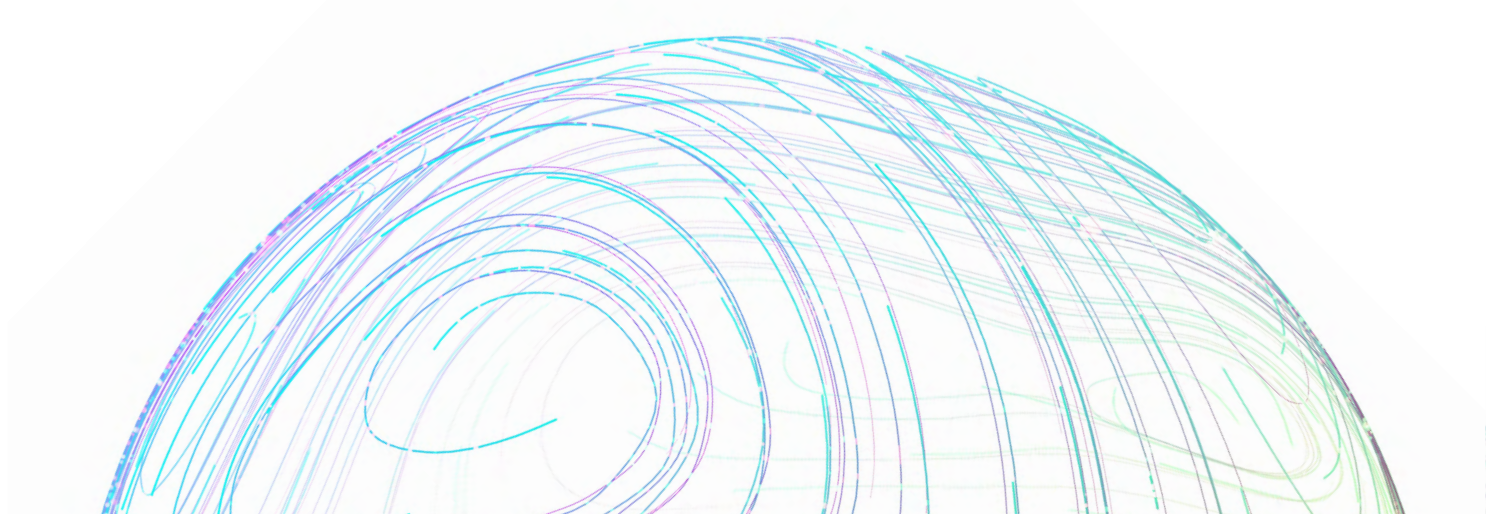
IT, once a major upfront investment and a barrier to entry for small companies, is now available at any scale and is fully self-service. As a result, a small player can leverage cloud services to spend more time on mission-critical activities dedicated to growing the company without having to invest in additional talent and one-time capital expenditures (CapEx) for back-end operation assets. The cloud's flexibility and scalability provide an ideal platform to quickly ideate, build, validate, and grow a product. In light of the benefits associated with the cloud, one could argue that starting in the cloud should be a baseline strategy for most companies.

Still, the cloud's flexibility and scalability may come at a cost premium, and a company operating at scale may find their costs of operating within the cloud to exceed that of maintaining an on-premise (on-prem) operation: Equivalent workloads running in the cloud under public prices are estimated to be 10–12 times more expensive than the cost of running on-premise data centers—even accounting for talent, hardware, and real estate costs.[1] This observation should not come as a surprise since publicly listed cloud prices often incorporate negotiated discounting, especially in business-to-business (B2B) environments. As a result, a company's cloud cost can *still* be more expensive than running data center operations on-premise, in spite of discount pricing from committed capacity agreements.

The rationale for hosting workloads on-premise may extend beyond finances. In this paper, we will explore additional circumstances where it makes sense to keep or move workloads on-premise. For these cases, we will discuss how to operationalize on-premise hosting to realize a competitive advantage, answering the question: Should all our workloads live in the cloud or should we repatriate any workloads to our on-premise data center?

The answer is twofold: 1) Strategically balancing a multi-cloud strategy that intelligently identifies workloads, operations, and teams that are operating in the cloud, in a hybrid cloud/on-premise arrangement, or in on-premise data centers; and 2) using everything-as-a-service (XaaS) to enable the on-premise data center of the future.

Gartner projects that by 2025, 85% of infrastructure strategies will integrate on-premise, colocation, cloud and edge delivery options, compared to 20% in 2020.[2] The issue is not cloud versus an on-premise data center, but rather balancing flexibility and realizing optimization at scale.



Amid cloud's growth, is the on-premise data center going away?

Trends show that cloud providers may continue to gain market share in generalized storage and compute in coming years. In fact, both Statista and Synergy Research Group cited 2019 as the first year that cloud infrastructure services (CIS) spend surpassed traditional data center hardware and software spend (figures 1 and 2).[3] However, CIS growth has not come at the expense of data center hardware and software spend, which has roughly remained constant at a compound annual growth rate (CAGR) of roughly 2.2% over the past 10 years.

Data supports the idea that enterprises are spending more on IT infrastructure services today than before, but will this trend continue? How will enterprises manage their on-premise data center operations in the years to come, and will there eventually be a decline in future data center spend?

We see hybrid cloud/on-premise and distributed cloud becoming the future of IT, and many organizations in the technology industry echo this hypothesis.

According to Dell Technologies, "The inevitable result of this on-prem vs. cloud debate is that organizations frequently adopt multiple public cloud solutions and while also investing in private cloud infrastructure, creating a multi-cloud environment that is intended to meet all needs."[4] Similarly, Amazon Web Services (AWS) extended its reach in 2019 by offering products for infrastructure on-premise as part of its AWS Outposts product. AWS Outposts' list of benefits cites enabling a consistent hybrid experience by running a host of AWS storage and compute services on-premise, recognizing use cases and industries that will remain on-premise in the future.[5]

Table 1 presents six key factors that make workloads strong candidates for staying (or repatriating to) on-premise operations to achieve an optimal operational mix.

Importantly, IT managers should apply rigor beyond analyzing if an application, workload, or dataset *can* be migrated to the cloud to understand if it *should* be migrated from a business value point of view.

Figure 1. Global cloud and data center spending 2020

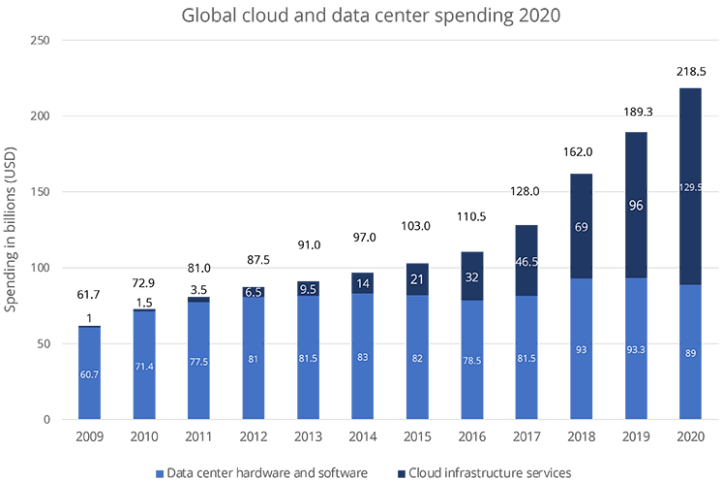


Figure 2. Enterprise spending on cloud data centers

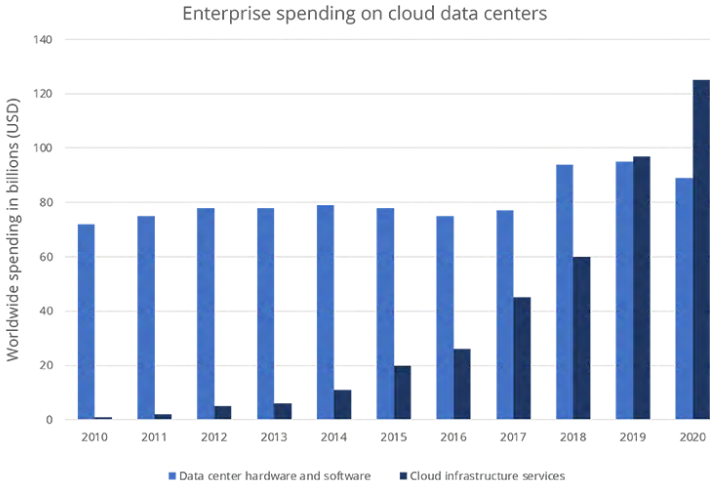


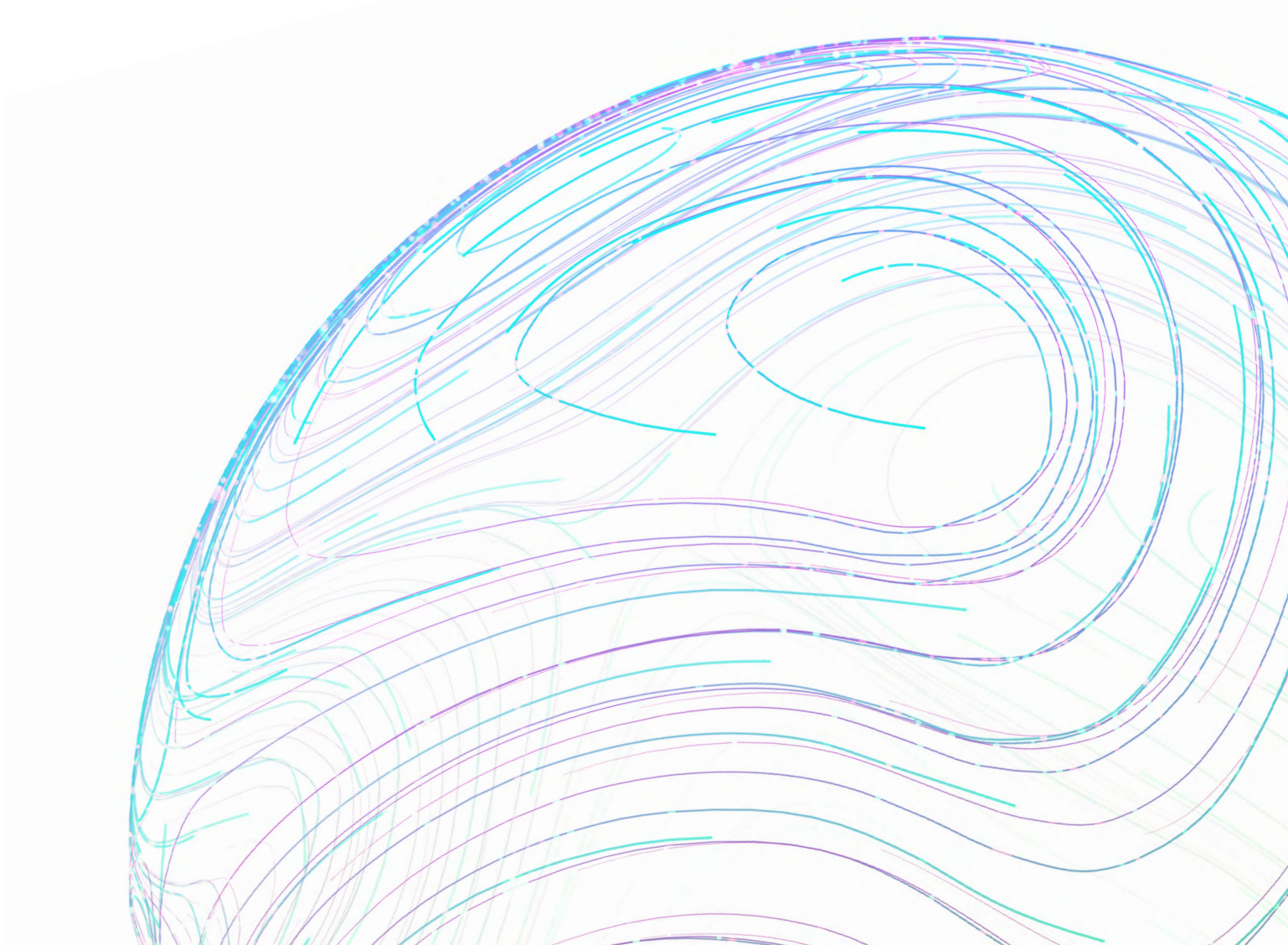
Table 1. Key factors and illustrations for candidates of on-premise workloads

Key factors	Description
Assets are subject to ***external regulatory*** and/or ***compliance requirements***	Application, workload, or database has jurisdictional privacy, security, data residency, or other regulatory and compliance requirements that will pose a significant challenge, create risk, or require on-premise residency. *Example: Since the enforcement of the General Data Protection Regulation (GDPR) in May 2018, "Software-as-a-service (SaaS) providers that collect data on EU residents need to have much stronger controls around the collection, tracking, sharing and protection of that user data." As the collection of that user data grows, some assert that "the cost of a hosted, multi-tenant delivery model [may increase] relative to a single-tenant, 'on-premises' delivery model that limits data collection."*[6]
Application, workload, or database requires ***specialty hardware*** and/or special ***configuration***	Application, workload, or database requires a specific hardware asset or configuration that is difficult to find in the public cloud market. The combination of technical specs requires customization best suited for in-house and on-premise execution. *Example: Machine learning for video-graphic content and engineering/architecture simulation modeling are recommended to run with at least 16GB of RAM (up to 72GB in extreme cases) with very few virtual CPUs. Such configurations are best created in-house given that a hyperscaler will simply offer its biggest instance type available, which is not a cost-effective option.*
Application, workload, or database requires exceptionally low **latency**	Application and workloads require an exceptionally low latency to deliver on mission. The company's latency needs require maintaining a private network with customized configuration of hardware assets. *Example: Online gaming and e-sports experiences require the lowest latency possible, while rendering detailed graphics, security, multiplayer live-play, and streaming to external audience platforms. While cloud providers vouch to reduce latency to as much as 20% of standard acceptable lag, on-premise solutions are still preferred to support the highest quality gameplay.*[7]
Time, effort, and complexity to refactor and ***migrate an existing application and copious number of resources***	Workload requires a predictable and numerous amount of resources, and its availability is mission-critical without a forecastable need for legacy functionality changes. *Example: Health insurance companies adjudicated up to 19.4 million in-network claims in 2022 according to cms.gov.*[8] *Business systems are often built on legacy mainframe architecture, and the code reflects decades of business rules in adjudication engines on-premise. Mainframe continues to provide secure, fast, and reliable execution for legacy back-office systems unlikely to change.*
Capacity for application or workload needs to easily be adjusted by ***adding or removing servers on-premise***	Application and workloads are containerized and run on horizontally scalable server clusters that can be adjusted for capacity by adding or removing servers. *Example: Platform9 offers a managed service for maintaining on-premise deployments of Kubernetes clusters, giving in-house developers a cloudlike experience and the ability to develop highly scalable applications in a private cloud based on the capacity of the compute cluster available. By adding hardware to a data center, the capacity of these applications is increased proportionately.*
Operating an Internet of Things (IoT) solution at scale generates high ***egress costs***	IoT products incur high and growing egress costs, as they transfer terabytes of data through their digital solutions each day. *Example: Considering that certain cloud providers charge up to $0.12/GB for egress fees, extensive IoT solutions, which involve transferring terabytes of data daily, are a growing cost to take into consideration.*

The role of XaaS in the data center of the future

As organizations continue to move into the hybrid and distributed cloud world of IT, on-premise data centers will need the ability to respond quickly to changing needs to overcome the general view that on-premise operations tend to sacrifice flexibility for control. Companies should consider shifting to as-a-service (aaS) business models to create more flexibility within the on-premise data center and address pain points experienced with traditional one-time hardware asset purchases. According to Deloitte's 2021 edition of its *Everything-as-a-Service (XaaS) Study*, more than 50% of companies classified as leading adopters of as-a-service models reported improved decision-making, better collaboration, accelerated experimentation, and an ability to reinvent business processes as organizational benefits of adopting XaaS offers in their business operations.[9]

In an XaaS model for product-centric businesses, a customer consumes hardware assets and maintenance as a service (aaS), giving IT control and greater flexibility over what they consume and pay for. XaaS offerings for product-centric businesses typically bundle hardware with software and support services to give users optimal management of their assets and the ability to rightsize their on-premise hardware asset inventory according to their needs at any point in the growth journey. This typically removes the need for bulk investments in hardware assets, instead enabling ongoing increment and asset mix adjustments according to business needs. In this way, the hardware asset provider builds a relationship driven by the customer's success and growth.



In table 2, Deloitte research shows that XaaS can solve the following pain points that may arise when product-centric businesses run on-premise data centers under a traditional one-time purchase for hardware assets.

Table 2. XaaS addresses one-time purchase data center pain points

Pain points	One-time purchase model	XaaS purchase model
Reduced time to value due to complex, time-consuming CapEx purchase cycle	Hardware assets purchased in one-time transactions are typically classified as a CapEx. CapEx contracts typically involve high dollar amounts and negotiations can be complex and time-consuming, requiring months of internal approval and putting the onus on IT to demonstrate the business value of the asset purchase.	In an XaaS model, customer does not purchase the asset; instead, they incur much lower monthly fees that are typically classified as recurring operating expenses (OpEx). As a result, it is usually easier to obtain corporate approval, making XaaS better suited for incremental purchases and subscription changes in response to evolving customer needs.
Hardware asset obsolescence due to lengthy intervals between hardware asset renewals	Customer must trigger a new CapEx purchase cycle after a typical hardware life cycle of two to five years.[10]	Since the customer does not own the asset, they do not have to worry about asset obsolescence, as the provider is tasked with managing the asset's useful life.
Significant upfront investment required	High upfront investment limits the customer's capital allocation opportunities. It also deters customers with cash budget limitations.	A monthly recurring fee model allows customers to consume hardware assets without a significant upfront investment. (Most companies do require a minimum commitment for a specified term period.)
Limited ability to support wide variety of workloads due to fixed hardware asset mix	Customer must trigger a new CapEx purchase cycle to change data center hardware asset mix. Also, they cannot return or exchange hardware assets with a one-time purchase model.	Customer can adjust their hardware asset mix through upgrades, add-ons, or renewals to accommodate a variety of workloads.
Internal teams may be constrained by fixed capacity and run-time to accommodate respective workloads	Customer must trigger a new CapEx purchase cycle to increase a single data center's capacity.	Customer may adjust total hardware asset count at any time under the same contract: • Add/remove hardware to scale within the terms of the contract considering minimum commitments • Ability to rightsize hardware assets
Physical asset ownership is an ongoing responsibility to maintain	Customer must trigger warranty for replacement of faulty equipment. Warranty time span is limited.	Ongoing service for replacement of faulty equipment is the responsibility of the provider and typically included in a service offering, for example: • Rack fans • Faulty drives • Cables

By solving these and other customer pain points, XaaS has the potential to provide the following benefits:

- **Asset flexibility**
  Customer may adjust on-premise hardware asset mix incrementally at any point in time.
- **Scalability on demand**
  Customer may adjust capacity and pay only for the assets on-premise at any point in time.
- **Purchasing experience**
  Customer pays for IT hardware assets and maintenance as a recurring OpEx expense.
- **Ongoing support**
  Customer selects the desired level of support to help optimally manage assets.

We acknowledge that despite the added flexibility that comes with consuming hardware assets as a service, not all customers have the physical space, energy capacity, or cooling systems to adjust capacity on-premise. Our research found that hardware providers are responding by offering modular data center solutions designed to address customers with limited space, edge computing needs, and even modular cryptocurrency mining operations.

These solutions revive concepts proved in the mid-2000s, allowing customers to purchase a "data center in a box," a fully integrated, plug-and-play data center inside a shipping container designed to expand onsite data center operations. Other promising models include partnerships between colocation data center providers and server management software providers. Such market solutions further support the idea that there will continue to be an appetite for on-premise or colocated data center operations, particularly those that allow companies to rightsize IT and that demonstrate a real commitment for providers to grow with their customers.

It is important that the data center of tomorrow position itself as a scalable and flexible resource aligned with the hybrid cloud architecture of the future. By consuming hardware assets under an XaaS model, customers can position their on-premise data centers as internal, operationally optimized resources that deliver a consistent cloudlike experience.

Beyond the data center of tomorrow

This paper does not intend to narrow the relevancy of XaaS business models to on-premise data center operations. The flexibility and scalability that XaaS models provide also apply to many other promising business operations technologies. For example, the number of connected devices is forecasted to be nearly double from 15.14 billion in 2023 to 29.4 billion by 2030.[11]

Companies will need to manage fleets of these devices that require ownership and consumption models that fit a particular business need. XaaS models are positioned well to support both device fleet management and edge computing capabilities that allow changing needs for scaling and hardware flexibility.

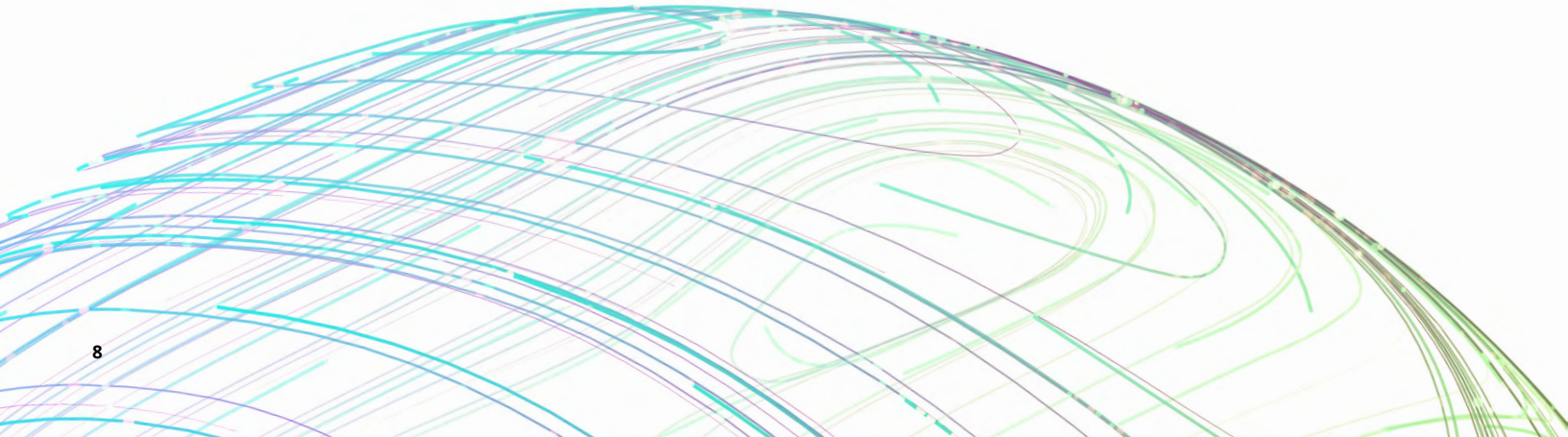
Contacts

Jordan Cole
Everything-as-a-Service Practice Lead
Deloitte Consulting LLP
jorcole@deloitte.com

Veronica Pulido
Everything-as-a-Service Leader
Deloitte Consulting LLP
vepulido@deloitte.com

Contributors

Jorge Reyes
Malavika Sharma
Rohan Talukdar
Julia Castanet



Endnotes

1. Sarah Wang and Martin Casado, "The cost of cloud, a trillion dollar paradox," Andreessen Horowitz, May 27, 2021.
2. David Cappuccio and Roger Williams, "The everywhere enterprise: A Gartner Q&A with David Cappuccio," Gartner, October 6, 2020.
3. Lionel Sujay Vailshery, "Global cloud and data center spending 2021," May 15, 2023; Synergy Research Group, "2020 – The year that cloud service revenues finally dwarfed enterprise spending on data centers," March 18, 2021.
4. Dell Technologies, "Hybrid cloud management," accessed May 22, 2023.
5. Amazon Web Services, "AWS Outposts Family overview," accessed May 22, 2023.
6. Taylor Wakefield, "GDPR – SaaS reconsiders on-premises software," Teleport, March 26, 2018.
7. Liz Alton, "An introductory guide to esports technology," Connected, September 27, 2019.
8. Centers for Medicare & Medicaid Services, "Health Insurance Exchange Public Use Files (Exchange PUFs)," accessed May 22, 2023.
9. Susanne Hupfer et al., *Gaining and sustaining a competitive edge with cloud and as-a-service IT*, Deloitte Insights, 2021.
10. Petroc Taylor, "Data center server replacement frequency 2020," March 2021.
11. Lionel Sujay Vailshery, "Number of IoT connected devices worldwide 2019–2021, with forecasts to 2030," November 22, 2022.

Deloitte.