High Tech Practice

Transforming infrastructure operations for a hybrid-cloud world

Despite the rapid growth of public-cloud adoption, internal-infrastructure teams are well positioned to add unique value. But they'll need to change their approach and evolve their service.

by Aykut Atali, Chandra Gnanasambandam, and Bhargs Srivathsan

October 2019

Public-cloud adoption is gathering momentum. By 2021, about 35 percent of all enterprise workloads will be on the public cloud, and 40 percent of companies will use two or more infrastructure-as-a-service (IaaS) and software-as-a-service (SaaS) providers, according to McKinsey's 2018 IT as a Service (ITaaS) Survey. Despite the growing interest, however, a good number of workloads will not be moving to the public cloud anytime soon.

In total, about 65 percent of workloads will continue to be hosted in private data centers and managed by internal-infrastructure teams over the next several years. There are a variety of reasons for this, including better total cost of ownership (TCO) in certain cases, the need to safeguard sensitive intellectual property, the absence of viable public-cloud providers in some countries, that skill-set enterprises have been built up around managing legacy systems, and the perceived need for control over security and regulatory needs.

These factors, combined with the growing use of edge computing, mean that a hybrid, multicloud infrastructure will become the de facto way of operating. The problem is that too few companies have adequately prepared for that reality. Many IT-infrastructure organizations lack a comprehensive strategy. Partly as a result, they have struggled to evolve their service. As public-cloud innovators offer attractive features, such as pay per use, high resiliency, and the ability to scale use with demand, the gaps are becoming all the more glaring. Instead of receiving a seamless hybrid-cloud experience, internal and external customers often face a discordant one. Moreover, companies across the industry face a strategic imperative to build faster and more effective delivery platforms to jump-start growth, speed time to market, and foster innovation—and technology is the keystone in enabling that capability.[1]

To address these issues, infrastructure teams must significantly alter their approach. By embracing world-class demand planning, capacity delivery, service operations, and strategic sourcing, organizations can achieve transformative returns. In addition to obtaining double-digit savings in labor, infrastructure, and capital expenditures, leaders that have applied these practices have accelerated end-to-end capacity delivery fourfold, gained a 20 to 30 percent improvement in infrastructure utilization, and improved customer satisfaction twofold.

A yawning performance gap

There's no doubt that "hyperscalers" such as Amazon Web Services, Google Cloud Platform, and Microsoft Azure have raised the bar for internal-infrastructure teams. Product owners, engineers, and business customers appreciate the high availability, flexible capacity arrangements, transparent cost structures, and rapid incident response that large public-cloud providers offer. Moreover, customers know they can count on hyperscalers to provide the latest technology.

These attributes have put the relative weaknesses of internal-infrastructure operations on sharp display. Product teams and external customers aren't able to see the true cost of what they're paying. Capacity is fixed, and the ordering cycle is so long that product teams often have to predict needs for one or more quarters in the future, increasing the risk of errors.

That said, internal-infrastructure teams have several natural advantages. Not only do they have deeper knowledge of the company and its customers but they can also deliver superior TCO in a number of use cases. Many have paid down the cost of their assets, in full or in part, and they own proprietary customer data. They also have the ability to tailor software and hardware solutions to customer needs.

By playing to these strengths and making relevant changes to skill sets, processes, and policies, internal-infrastructure teams can deliver significant value—and even increase the share of workloads managed—becoming indispensable partners to their organizations.

---

[1] See Oliver Bossert and Driek Desmet, "The platform play: How to operate like a tech company," February 2019, McKinsey.com.

Creating a world-class infrastructure organization

McKinsey research shows that CIOs increasingly recognize that organizations cannot capture agility benefits solely by moving applications to the cloud. Instead, they need to reassess the infrastructure stack and the way it works.[2] Improving cost, speed, flexibility, and service requires an integrated, end-to-end approach to transformation, given the interplay among functions, processes, and partners. Our data show that taking a holistic approach can unlock a twofold improvement in customer service, a 50 percent improvement in reliability and availability, and fourfold faster capacity deployment in addition to substantial productivity gains. Moreover, better collaboration within the broader company and partner ecosystem can help IT play a more pivotal role in supporting business innovation and growth (exhibit).

To achieve these gains, internal-infrastructure teams should follow four best practices.

World-class demand and capacity planning

Demand planning cannot be done on a month-to-month basis. It should be managed more like a budgeting and capital-planning process. Knowing if, where, and when to build a data center, rent space from a co-location facility, or purchase compute, storage, and networking capacity from a hyperscaler requires infrastructure teams to reliably estimate demand over a multiyear horizon. Preparing the land and constructing a new data center, for instance, can easily take up to two years. Likewise, renting space or buying reserved instances (RIs)[3]—or committed-use discounts—on the cloud can bind businesses to terms of anywhere from one to three years. Poor decisions can add millions of dollars in cost—and can leave businesses scrambling to make up for capacity shortfalls in some regions, while needing to manage excess capacity in other regions.

Optimizing data-center footprints and utilization rates takes strong sales- and operations-planning (S&OP) processes. Internal-infrastructure teams need to be able to translate the business's growth projections into a detailed resource forecast for each region. By capturing demand signals, such as revenue estimates for new and existing markets, and employing predictive machine-learning-enabled engines, infrastructure teams can adjust workload requirements with greater precision and speed.

Like with budgeting and capital planning, organizations need both a long-term plan that looks out three or more years as well as a short-term plan that is updated quarterly. For long-term planning, revenue targets should be translated first into usage growth (for example, core hours) and then into resource requirements (for instance, compute, storage, and network) based on product and platform road maps. For short-term planning, demand forecasts can be created from resource use at the cluster level by applying advanced analytics and overlaying sales and product intelligence. By improving planning in these ways, IT-infrastructure teams have the potential to lower data-center costs by 20 to 30 percent and achieve data-center fill rates of as much as 70 to 80 percent, among other benefits.
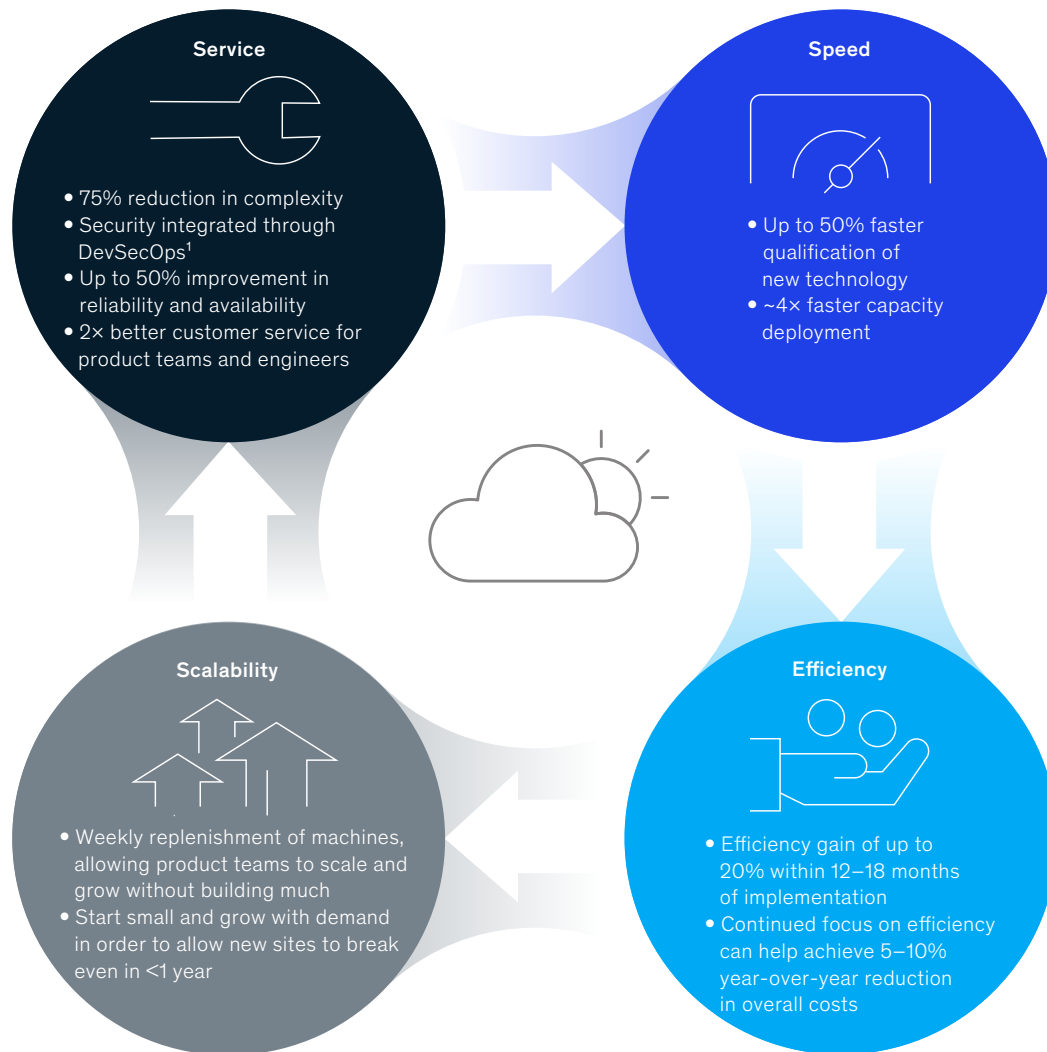
Rapid capacity delivery

Many internal-infrastructure teams maintain an array of server configurations and deploy capacity on a make-to-order basis, with little inventory held in stock and no room for rapid deployment. During the typical two- to six-month window from order to delivery, needs can change, leading to new configurations and additional work. Rather than customizing machines individually—with hundreds of different machine types, as is often the case now—infrastructure teams should standardize their configurations and segment their supply chains.

Based on our experience, a handful of common configurations can cover more than 90 percent of demand in most organizations. Standardizing configurations around that core number enables infrastructure teams to move to a make-to-forecast model, rationalize demand around a select number of suppliers, negotiate more favorable inventory

---

[2] Arul Elumalai and Roger Roberts, *Unlocking business acceleration in a hybrid cloud world*, August 2019, McKinsey.com.

[3] A reserved instance provides a discount in return for committing to a certain amount of resources or capacity for a multiyear term, usually between one and three years.

Leaders with world-class cloud operations will achieve performance improvement across all critical aspects.

Service

- 75% reduction in complexity
- Security integrated through DevSecOps[1]
- Up to 50% improvement in reliability and availability
- 2× better customer service for product teams and engineers

Speed

- Up to 50% faster qualification of new technology
- ~4× faster capacity deployment

Scalability

- Weekly replenishment of machines, allowing product teams to scale and grow without building much
- Start small and grow with demand in order to allow new sites to break even in <1 year

Efficiency

- Efficiency gain of up to 20% within 12–18 months of implementation
- Continued focus on efficiency can help achieve 5–10% year-over-year reduction in overall costs

[1]Building security into infrastructure and applications from the start of the development cycle.

terms, and shorten lead times with a high degree of confidence. These changes have allowed internal-infrastructure teams to shrink order-to-delivery time to four weeks or less.

Smart service operations
Internal-infrastructure teams face high labor costs, both from their in-house base of employees as well as from their outsourcing contracts. Those expenses often amount to 20 to 30 percent of total

operating costs. To gain greater efficiency and cost performance, infrastructure teams need to do three things: inventory their core activities and determine which activities to retain in house and which to outsource, take advantage of lean process redesign and automation, and establish smart contracts and managed-service agreements with their outsourcing vendors to drive ongoing efficiency gains and greater third-party accountability.

Activity-based resource planning can help determine both the physical head count needed to maintain on-premise services and the processes to outsource and automate to gain the necessary scale and cost advantages. To extract further value, infrastructure teams should adopt metrics-driven performance management with their third-party providers. For example, infrastructure teams seldom invoke the penalty clauses in their third-party contracts, even when service-level agreements (SLAs) are unmet. Conducting regular service-delivery reviews could help catch these oversights. In addition, examining time and material spend across different providers and geographies can allow teams to consolidate workloads in lower-cost hubs and potentially convert contracts to smart contracts or managed-service agreements. For example, contracts to manage a set of network or storage devices can easily be converted to a managed-service contract and potentially consolidated under a single third-party provider. These steps have the potential to free up 15 to 20 percent in costs spent on both employees and third parties.

Strategic sourcing

In many infrastructure organizations, capital expenditures are often split across a large number of suppliers, leading to transactional relationships that reduce economies of scale and limit scope for co-innovation. To address this, internal-infrastructure teams should consolidate the supply base, eliminating or minimizing use of the vendors with which they do only small volumes of business. Greater purchasing power and commitment allows infrastructure teams to create strategic longer-term partnerships. With stronger negotiating leverage, they can lock in superior pricing, hedge against market fluctuations, and capture future savings as compute and storage costs fall over time (following Moore's law principles). A more strategic approach to sourcing can also open up opportunities for codesign in ways that improve TCO and performance. Through this more strategic approach to sourcing, internal-infrastructure teams can shorten capacity-deployment lead times by roughly 50 percent, eliminate redundant spend, and lower capital expenditures by 10 to 15 percent.

In addition, infrastructure teams should consider whether it makes sense for them to move to an original-design-manufacturing (ODM) model, in which the team designs the machines and procures parts in house and then outsources the build. That shift requires scale and solid in-house design, as well as supply-chain and maintenance capabilities, but it can give teams more control over their sourcing and inventory. In addition, it can allow them to cut out the high-end features and components they don't need, unlocking potential savings of 20 to 40 percent of total equipment costs. Finally, infrastructure teams should take a similarly strategic approach to their sourcing arrangements with hyperscalers (see sidebar, "'Hyperscaler' optimization").

Putting the right building blocks in place

Instead of infrastructure teams seeing themselves as traditional IT organizations, they need to function like strategic partners to their companies. They should adapt their governance, talent, and performance-management systems to support greater customer centricity, accountability, and collaboration.

Stronger engagement with product owners can help infrastructure leaders jointly shape solutions—ensuring, for instance, that when a new platform is introduced, the infrastructure-architecture team, sourcing team, and product-operations team create configurations with the right cost and performance trade-offs. Creating a dedicated sales and operations council can help infrastructure leaders embed forecasting and budgeting discipline into their planning efforts—improving resource utilization and business value.

From a governance perspective, each data-center site should be managed as a separate profit-and-loss center with internal charge-back mechanisms introduced to drive visibility and accountability. Creating business-unit-specific bills can be a powerful way to educate application teams about cloud-based costs and, as a result, rightsize demand. Performance management needs to evolve

'Hyperscaler' optimization

'Hyperscaler' spend is a large and growing part of total infrastructure spend. To get more value from that spend, teams need to do the following:

— Restructure contracts for the right mix of resources to find the optimal balance of regions, reserved instances, and on-demand services[1]—

and optimize configuration and service-level agreements based on improved demand planning.

— Establish transparency across IT, business, and finance to understand the drivers of demand and their cost implications—and to monitor spend and usage at a granular level.

— Automate resource planning for further optimization. Use artificial intelligence and machine learning to detect workload requirements and adjust resources for self-healing resource planning—for example, terminating or resizing unused instances, deleting unattached storage, and flagging aged snapshots and disassociated IP addresses.

[1] A reserved instance provides a discount in return for committing to a certain amount of resources or capacity for a multiyear term, usually between one and three years. On-demand service allows a cloud user to pay per use, an option that provides the most flexibility, although at a higher price point.

similarly. Digital systems and dashboards can instill consistent measurement across the organization, and a regular cadence of reviews with product teams can ensure their needs and those of end customers are well understood and consistently met.

Attracting the same talent that hyperscalers are hiring will force IT-infrastructure teams to think and plan differently. Working in a multicloud environment, for instance, will require infrastructure teams to have the necessary design and engineering talent in house. To enable world-class cloud operations, IT-infrastructure teams will need architects to help the business units redesign workloads for the public-cloud capabilities. They'll also need data scientists and operations-research practitioners to help with capacity planning. In addition, teams must have sufficient depth in operational areas such as supply-market analysis, cleansheet-based negotiations, activity-based resource modeling, demand planning, and contract

management. Roles and decision-making rights across the internal and partner ecosystem must be defined as well. While some talent may be upskilled from within the existing IT organization, teams will also need to pull from nontraditional pools and complement with smart outsourcing services. Some may even choose to invest in research arms or partnerships in tech hubs like Silicon Valley that can provide the culture and peer cohort needed to recruit and retain in-demand talent.

———————

Hyperscalers are getting a lot of attention these days, and for good reason. They can deliver above-average speed, scale, service, and efficiency. But no matter how good they are, public-cloud providers won't be able to displace the essential role that internal-infrastructure teams play. From achieving superior TCO in some instances to satisfying very specific use-case needs, internal-

infrastructure teams are in a position to add unique value and maintain their critical role in a multicloud environment. To do so, however, they'll need to improve demand and capacity planning, rationalize configurations, embrace digital service operations, and take a more strategic approach to sourcing.

If internal-infrastructure teams can manage that, then they will be able to deliver the world-class cost, service, and performance so many of their customers inside and outside their organizations increasingly expect.

Aykut Atali is an associate partner in McKinsey's Silicon Valley office, where **Chandra Gnanasambandam** is a senior partner and **Bhargs Srivathsan** is a consultant.