Cloud best practices

A guide for IT executives seeking to
transform their business through cloud
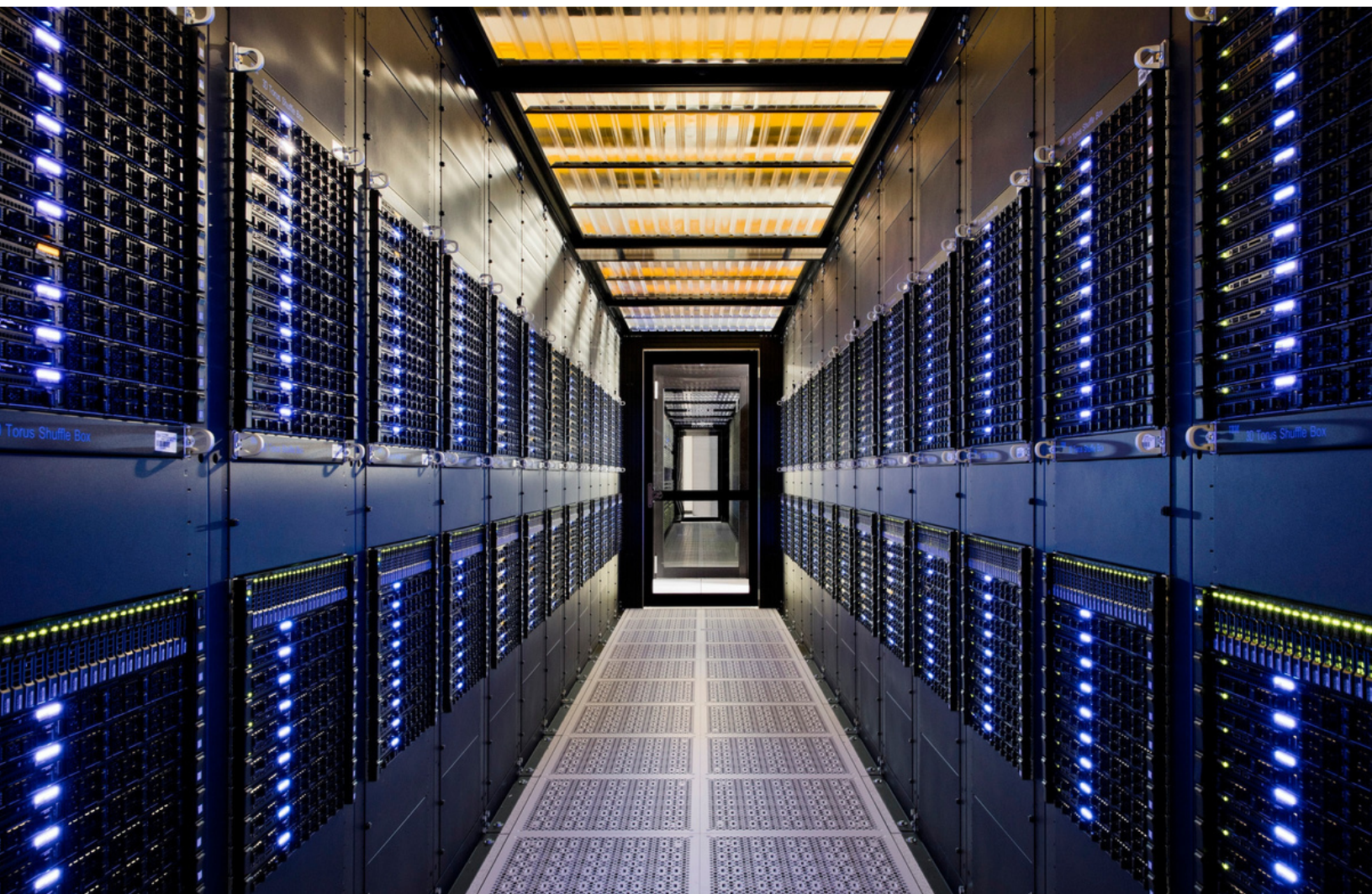
Table of contents

Introduction

It's the seventh of the month and your CFO just received a large cloud bill.

Again.

Like many times before, your cloud costs have reached an all-time high much sooner than predicted or budgeted for. Sound familiar? You're not alone.

Many organizations tell us this same story. Top talents, the technical movers and shakers of the company, are pulled away from key tasks to fix cloud-related issues. In many cases, despite proactive and even heroic efforts, cloud bills are still running amok and any changes you make may put application performance at risk. The glorified promises of cloud benefits and savings aren't being seen—instead, customers become trapped in an endless break-fix loop as they strive for transformative cloud optimization.

Since 2016, organizations using cloud-based computing put optimizing existing cloud use—cost savings—as their top priority. In 2020, over 73% of organizations rated cloud costs as their highest IT priority.1 The more advanced and mature customers are in their journey, the higher they rate this challenge—concluding that managing cloud costs isn't a skills issue, but rather the nature of the cloud.

Nonetheless, cloud computing is a worthwhile investment despite the new challenges that come with it.

It's an investment in increased agility, elasticity and better cost intelligence—all of which are essential strengths for a modern organization to harness.

Agility is ultimately about the speed of innovation. This speed of innovation is achieved when businesses can regularly deliver new functionality to users, continually improve their products and methodically increase the value that they deliver to their users. Agility rids businesses of the need to spend months on product requirements—as traditional "waterfall" methodologies advocate—before executing and building. Rather, it helps enable businesses to deliver small improvements that are iterated upon based on user feedback and expanded regularly, increasing the pace of innovation through trial and error. Being able to consistently deliver improvements to customers substantially increases your competitive edge in the market. The cloud can significantly advance an organization's ability to be agile by ensuring two main goals:

– Application teams always have the required resources they need to build, test and release applications. Within minutes, any type of resource can be deployed and available for use. On premises, many of our clients have reported that the average wait time for an application owner to get new resources provisioned is 6 – 8 weeks. This wait time translates to perpetual delays in development, user acceptance testing (UAT) and even production. The cloud mitigates these delays in that it allows for new resources to be deployed on demand, within minutes—under the condition that you put in your credit card details, of course.

– Cloud providers deliver application components as managed services, known as platform as a service (PaaS), that implement best-in-class capabilities. These managed services grant application developers the capacity to direct their attention elsewhere as their expertise is no longer needed to build these components from scratch or exhaust their limited time waiting for new virtual machines (VMs). Focus on delivering capabilities and innovative solutions that prove valuable to your customers as opposed to "keeping the lights on" managing your IT environment. This process applies to certain scaling abilities, the need to update and patch different application components and so on.

Elasticity is inarguably the core element of the cloud. As a matter of fact, Amazon Elastic Compute Cloud (Amazon EC2), Amazon's infrastructure as a service (IaaS), is by far its most adopted service. Elasticity is the ability to expand and contract the environment based on changes in user demand, a critical element in unlocking the ability to pay for only what you need. On premises, even the most sophisticated organization must purchase hardware to meet peak demand, yet the majority of that hardware will sit idle through nonpeak times.

Cost intelligence requires organizations to move beyond simply thinking about cost savings. Cost intelligence is the mechanism that enables organizations to successfully find the balance between cost management and net new investment in innovation and growth—with the ultimate goal of improving customer satisfaction.

While agility, elasticity and cost intelligence are all powerful drivers to cloud adoption, most organizations aren't immediately realizing these benefits. The barriers to these benefits can't be broken down overnight; the more advanced customers are in the adoption of cloud and employing cloud-native methodologies services, the more challenging optimization becomes.

Cost optimization

Positioning the optimization of existing cloud spend as the top initiative for cloud users or staring at rising cloud bills won't reduce your spend, nor will sending reports back and forth. Not to minimize the importance of cost visibility and reporting—cost visibility is a critical foundation of any organization's cloud cost optimization strategy and required to establish the needed accountability—but it's simply one component of a cohesive strategy. To optimize a cloud environment, you must implement actionable change—but again, it's easier said than done.

Now, for the million-dollar question: Why is cost a continuous problem for organizations?

Because cost isn't the only challenge: it's merely the result or a symptom of a different challenge, how to safely assure application performance.

Cost optimization is a perpetual pain point for organizations because one can't truly separate performance from cost; the two considerations are tightly coupled.

For example, an organization could easily optimize for cost alone by moving all applications to the smallest instance type, but no one is willing to take the performance hit.

Cost may be a symptom, but it's not the root cause organizations should strive to resolve.

Let's look deeper into the root cause of the issue. The promise of the cloud is that you pay for only what you need, but the reality is that you pay for the resources that you provision and allocate to applications. Approximately 80% of companies report receiving bills 2 – 3 times larger than what they budgeted for; the gap between consumption and allocation is the cause of these large and unexpected bills. These cloud bills continue to creep up because organizations have yet to understand the actual resourcing needs of their cloud applications and the significance of continuously optimizing resource allocations across the entirety of the estate. As a result, application resource allocation is often a padded guess, and rarely—if ever—is it a decision subjected to comprehensive reevaluation.

The result? The vast majority of cloud estates are static and highly overprovisioned; so it comes as no surprise when we end up complaining about the cost.

Our mission

As organizations rush to the cloud—especially due to the impact of COVID-19—while continuously seeking to optimize both cost and performance, enterprise IT teams are placed under tremendous pressure. The cloud addresses an old pain point: infrastructure supply is static while application demand is dynamic, but matching demand with supply in real time across multiple metrics and dimensions requires making far more decisions than initially predicted.

Organizations shouldn't be expected to compromise on innovation for the years of experience required to gather a comprehensive understanding of the cloud. Cloud is a continuously moving target, requiring organizations to adapt and relearn as the technology develops month over month. There are limited resources that provide a nitty-gritty and meaty breakdown of best cloud practices, and even more limited resources that provide information tailored to a specific organizational role. We hope that this white paper will fill this gap.

The information in this white paper will answer questions we receive from our clients on a day-to-day basis. Our goal is to help industry leaders avoid making the initial, common and justified mistakes when migrating to the cloud, moving beyond the learning curve and preventing errors that you weren't made aware of at the start of your cloud journey.

The best cloud practices don't have an option for "one size fits all". It's necessary to understand how each role in an organization can best navigate responsibilities and goals in the cloud while having the ability to enable other team members in their cloud-based roles.

We'll provide our suggestions, recommendations and guidelines for cloud best practices so you can maximize value and performance for both your cloud team and environment. Our goal is to provide you with a comprehensive guide as to what you can be doing with your company's and cloud provider's existing resources, helping enable you to do more with less.

Cloud best practices for CIOs

CIOs have three main responsibilities when transforming their companies into cloud powerhouses:

–Set the strategy and the goals.
–Build the right organizational structure.
–Drive accountability and incentives.

This white paper discusses the importance of all these tasks, as well as our recommendations and strategies on selecting the right course of action for your company.

We begin with strategy development.

Strategy

The foundational element of a successful cloud journey is the organization's strategy. The right strategy needs to be realistic, achievable, quantifiable and, most importantly, aligned with the broader company strategy. The cloud isn't the end goal, but rather a medium to aid in reaching the overarching goals of the company. The role of IT should support—and even accelerate—business goals by enabling the organization to better innovate and delight its customers.

A case study in cloud strategy

A leading North American bank made the decision to shift from a financial-services-focused company and transform into a software-focused company. While it continued to provide financial services to its customers, the bank realized that rather than increasing its labor force, it needed to use software to deliver a higher-quality and more competitive service to its customers. As a result, its CIO announced a cloud-first strategy with the goal of moving entirely to the cloud within just a few years. In this case, its CIO was confident that the agility and elasticity the cloud provides would enable the bank to innovate faster, focus on user value and scale with lightning speed according to user demand, as opposed to reactively building static infrastructure.

We have worked with many CIOs that initially made strategic statements, such as "80% of our applications will be in the cloud in two years," but they had no concrete implementation of these strategic goals in their long-term planning. These initiatives not only fail often but may even cause organizations to lose focus of their original mission. Many organizations set overly ambitious goals that prove to be unrealistic, only to double down on them even when they're past the point of providing any significant value.

One organization we worked with set an initial goal of having all its applications in the cloud by the end of 2020. As the organization moved along in its journey, it encountered many challenges and was forced to develop very complex workarounds to try and accomplish the original goal.

Once again, an attempt to increase bandwidth and innovation regretfully transpired into troubleshooting toward an unattainable goal. After taking a step back, the organization realized that the cloud wasn't the problem—its goal was. The cloud was a great fit for some of its applications, but a poor fit for others.

Some organizations we talk to even consider repatriation, for example, moving their applications back on prem as a result. But it wouldn't change the reasons why they wanted to move to the cloud or allow them to continue and advance their businesses as they originally planned. Instead, some customers took a different approach; they evaluated the individual application profiles and identified the applications that would see the most benefit from running in the cloud. For them, it was the applications that were actively being developed—some already in production but being expanded and some yet to be released. From there they built a plan not only to move those applications to the cloud but to make sure they had a way to achieve elasticity and take advantage of the wide variety of options that were available to them in the cloud. Once they found a reliable way to do it, they expanded to additional applications.

The main way to avoid these kinds of setbacks is to ensure the strategy is realistic and achievable. A data-driven approach ensures the strategy falls into those two categories.

The first step to cloud strategy success

It's important to determine if a core competency of the company is to have full control of its applications and IT resources. If so, the cloud is likely not a good fit because companies lose the ability to have full control over their resources. Such examples are applications where sub-millisecond transactions and responses are critical, such as high-frequency trading where dedicated high throughput and low-latency network paths are built to serve these sensitive transactions. If it's unnecessary for a company to have full control, the cloud is a good fit because it relieves the stress of resource allocation and application management.

Although some experimentation is necessary for moving applications to the cloud, IT teams should do their best to provide a comprehensive list of each application they're currently managing. The IT team should also provide accompanying data around the time consumed managing each application and exactly what's being managed during that time.

Once all the applications have been assessed, the team will need to decide which applications are a good fit for the cloud and best suited to reap its benefits. Despite popular belief, it's rarely the best option to move all apps to the cloud. Some apps will require the same amount of labor to manage in the cloud as they would on premises. Each app should be evaluated for its ability to function with increased efficiency and agility in the cloud.

Consider compliance

Before continuing, let's discuss compliance restrictions. Your team needs to ensure that there aren't any compliance restrictions that would prevent any application from moving fluidly to the cloud.

Following the compliance restrictions assessment, you need to establish how dynamic the environment is. In a dynamic cloud environment, applications can automatically adjust to changes in demand or workloads. The cloud is a good fit for dynamic applications, whereas static applications with consistent demand are likely not a good fit for the cloud.

The key questions to ask your team are:

– How much manual effort is required when addressing fluctuating demand?
– How dynamic are your individual applications?
– Are the application candidates dynamic, a good fit for the cloud and able to grow and contract with demand?

After you've evaluated your core competencies and dynamic environment, it's an appropriate time to evaluate your cloud strategy.

Choosing a cloud provider

The first decision to make is the number of cloud providers you want to use. According to Turbonomic's 2021 State of Multicloud Report, a growing number of organizations are choosing to use multiple cloud providers, with 30% of organizations using three or more clouds. The major reason these customers are choosing multiple cloud providers is that each provider has different strengths and, based on the needs of each application, a different provider may be the best fit. That being said, almost every enterprise standardizes on a main cloud provider and additional cloud providers are supported and allowed but only based on the necessity of the applications. Sometimes these exceptions need to be approved on a case-by-case basis. The reason behind this method is simply a centralization and consolidation benefit to enable these enterprises to negotiate better deals and discounts with the provider of choice.
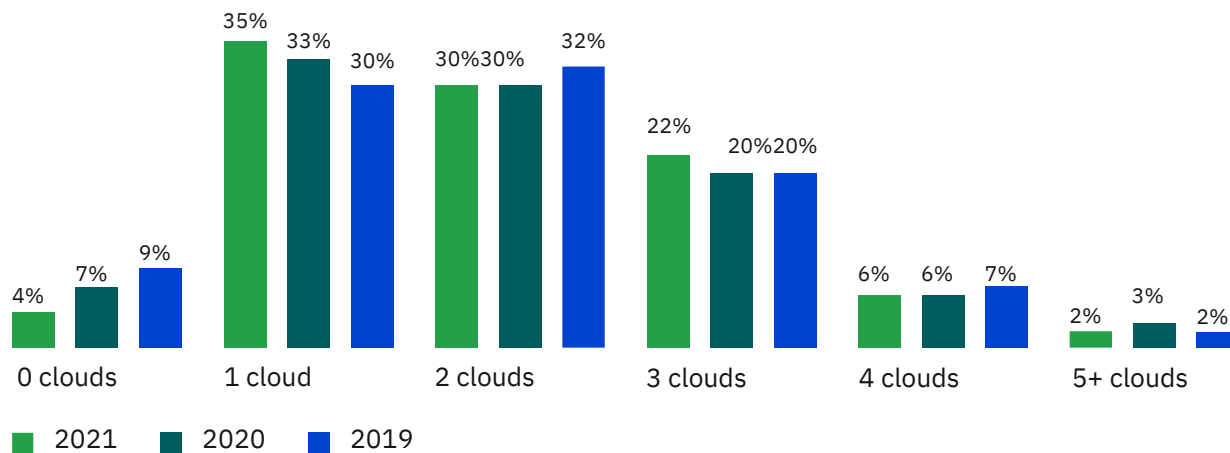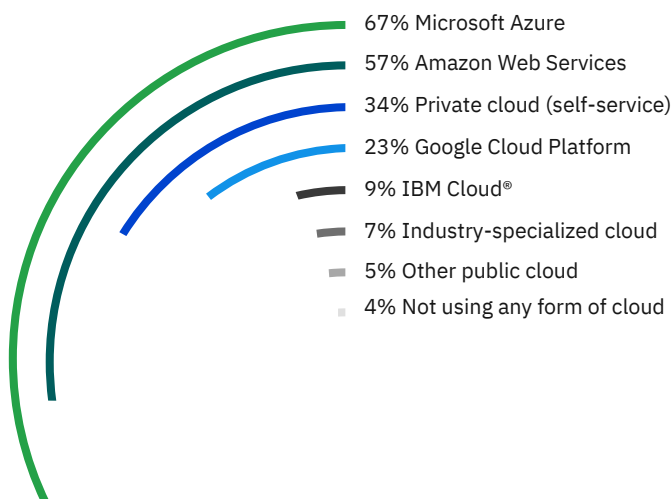
(Number of people that took the survey each year: 2021 = 819; 2020 = 938; 2019 = 736)

Figure 1. The use of multiple cloud providers is increasing.

Source: 2021 State of Multicloud Report

Each provider offers varying advantages and disadvantages. From our 2021 State of Multicloud Report, we're seeing a few interesting trends with current and future cloud provider adoptions:
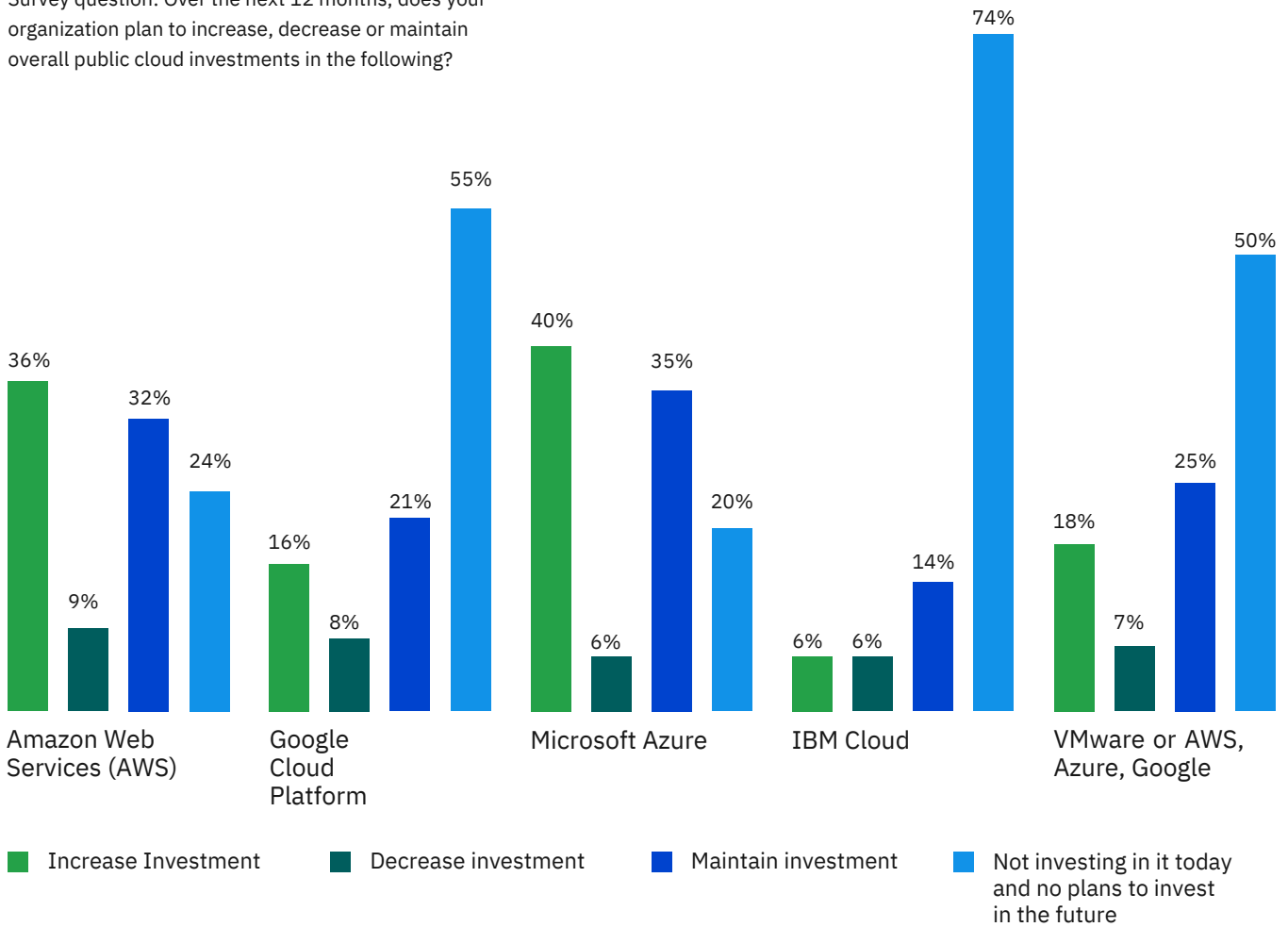

67% Microsoft Azure

57% Amazon Web Services

34% Private cloud (self-service)

23% Google Cloud Platform

9% IBM Cloud®

7% Industry-specialized cloud

5% Other public cloud

4% Not using any form of cloud

Last year, for the first time, Microsoft Azure overtook Amazon Web Services (AWS) as the leading cloud provider. It's continuing to maintain that lead. Also, notably, the percentage of respondents that aren't using any form of cloud computing today dropped by half in 2021, from 8% to 4%.

(Number of people that took the survey = 819)

Figure 2. In 2021, 67% of organizations are using Microsoft Azure, up from 61% in 2020.

Source: 2021 State of Multicloud Report

Survey question: Over the next 12 months, does your organization plan to increase, decrease or maintain overall public cloud investments in the following?

Amazon Web Services (AWS): 36%, 9%, 32%, 24%

Google Cloud Platform: 16%, 8%, 21%, 55%

Microsoft Azure: 40%, 6%, 35%, 20%

IBM Cloud: 6%, 6%, 14%, 74%

VMware or AWS, Azure, Google: 18%, 7%, 25%, 50%

Legend:
- ■ Increase Investment
- ■ Decrease investment
- ■ Maintain investment
- ■ Not investing in it today and no plans to invest in the future

(Number of people that took the survey = 819)

Figure 3: Percentage of investment in public cloud services in a 12-month period

Source: 2021 State of Multicloud Report

A thorough research investigation should be done to see which provider or providers would best fit your application needs and strategy. Gartner has extensive research and evaluations on the largest cloud providers, including:

• Solution Scorecard for Amazon Web Services IaaS+PaaS
• Solution Scorecard for Microsoft Azure IaaS+PaaS
• Solution Scorecard for Google Cloud Platform IaaS+PaaS

How to approach new application development

Now that you've begun your cloud journey, where and how do you develop new applications? There are two potential approaches, as it's doubtful that you'll remain static with no new application development. These approaches are addressed in greenfield versus brownfield. A greenfield approach in software development is developing an application from scratch without any dependencies on existing software or code, whereas brownfield development is dependent on or takes into account existing software. See Wade, 2018. It's much more difficult to develop applications from scratch in the cloud, so your organization needs to evaluate its future plans for application development. Do you intend to build your applications in the cloud going forward or maintain an on-prem strategy for future developments?

Once you have established your application development plan, it's time to define what moves and exactly where it moves. It would be a waste of valuable time to go through the effort of implementing a strategy and then realizing the uncertainty as to exactly where you're moving each application.

Troubleshooting is the final step in strategy development. Backup alternatives need to be defined and established in case applications aren't a good fit for the cloud or function differently than predicted.

A strong strategy for a successful shift to the cloud addresses all these points.

Figure 4: Shifting to the cloud

Organizational structure

In IT and cloud operations (CloudOps), departments and teams can have a centralized, decentralized or hybrid organizational structure. In a centralized organizational structure, all software, users and applications are connected to a central server that stores the information and data as opposed to a decentralized network, wherein separate user groups store data and information relevant to that specific group. See Kumar, 2020.

An exclusively centralized structure isn't possible in a cloud-based environment, because a number of applications and software will be hosted elsewhere. With that said, a completely decentralized structure isn't necessarily ideal either in that team members lack control of their operations.

Centralized structure

What are the benefits of a centralized structure?

This structure is a good fit for existing applications currently being centrally managed. The team can maintain the current management systems and operations in maintenance mode—where no features are being added and no improvements are being made. This structure is a good fit for back-office applications and third-party vendors, as they're not being developed and can be easily centrally managed. Teams can have control of additional applications and resources not developed within the company, rather than having them managed offsite.

Most significantly, a centralized structure is the easiest when seeking to control what is done, when it's done and where it's done, both in your team and the organization as a whole. All data, applications and software are stored in the same location, making for quick access times, efficient resource allocation and full visibility. When the team is in control of all its resources, it enables easier forecasting and budgeting for resources in the company.

Decentralized structure

A decentralized structure creates an environment better fit for self-developed applications. Decentralization invokes higher agility and scalability as companies aren't limited to the resources owned by the company. Additionally, this structure enables easier accountability and management.

The transition from a fully centralized IT organization to a decentralized one isn't trivial—there's a significant cultural shift that needs to be managed. While every organization has some central function, most organizations find themselves somewhere along the spectrum, neither fully centralized nor decentralized.

Hybrid structure

The two most common hybrid organizational structures are split by either environment type or by function.

–By environment type: The most common split is by environment type, where nonproduction is decentralized and production, along with back-office apps, is centralized. With the development and test workloads already decentralized, the main benefit of this approach is the increased agility benefits. In this approach, developers and quality assurance (QA) engineers are empowered to deploy any resources needed to deliver quick, reliable and high-quality applications. Developers and QA engineers are no longer burdened with the production environment, along with its particular set of challenges. These production environment challenges are managed by central IT—the team with decades of experience in running and managing such estates—empowered to run the applications under their management however they see fit.

–By function: Other organizations split the management and ownership of the environment by function. In most cases, the decisions relevant to the application itself are owned by the lines of business (LOB) and the application owners, whereas the functions that are broader than an individual application are managed centrally. In many cases, central IT incorporates a new function or team, a cloud center of excellence or a financial operations (FinOps) team. This new team is customarily responsible for the cloud estate as a whole and the following broad functions:

•Drive and define best practices, including the hierarchical and organizational structure of the accounts and subscriptions.
•Create policies on security and tagging hygiene, manage suspension schedules, and drive standard tooling for management across the organization.
•Purchase reserved instances (RIs) or saving plans, managing these functions over time.
•Drive accountability across the decentralized functions, in some cases negotiating both enterprise agreements and discounts as well as licensing agreements for third-party software.

Accountability and incentives

Given that most organizations are at least partially decentralized, there needs to be a certain level of governance and accountability to maintain control of the environment. If no guardrails are being enforced, the decentralized approach can turn into the "Wild West."

We equate it to giving teenagers their first credit card, but with no limits or guidelines, just a quick and easy way to get anything and everything they desire. As you can imagine, it can become disastrous very quickly.

This issue has happened to many of the clients we've worked with; one unfortunate clients was left with no choice but to call an emergency board meeting, requesting an 8-digit increase to their budget due to inflated and out-of-control cloud bills. It's a CIO's nightmare—but one that's preventable with the right accountability and incentive system.

Stages of accountability

For most of our clients, accountability comes first and in stages to avoid creating significant conflict or discomfort. To create accountability, accurate visibility needs to be established as to what each department, business unit or application team is deploying, running, using and paying for. This process is commonly referred to as showback.

Once the teams have a comprehensive understanding of their spend and are aware that this information is visible and available to the entire organization, they naturally feel obliged to justify their actions. With this new visibility, the various teams will typically make an increased effort to ensure the environment is being managed and maintained efficiently. It's the first in a series of steps to create accountability.

The next step is chargeback. After showing the different LOBs their spend over a few months, most organizations move to chargeback the business unit for that spend. That business unit then needs to justify the spend against their own budget. Often, this justification alone drives the desired outcomes, but in some cases isn't sufficient.

Most organizations define the LOBs budgets for application hosting services based on past spend, not actual needs. As a result, past overprovisioning and overspending creates unnecessarily inflated budgets. We've realized that with some clients, these budgets are defined arbitrarily. For business units spending above their budget, asking for a budget increase is easier than figuring out how to reduce their spend without negatively impacting the performance of the applications they're responsible for, the velocity of the team or the reliability of the service they're delivering.

The bottom line is that the spend determines the budget, and the budget determines the spend. Therefore, the showback, and even the chargeback, are usually not enough to drive the desired behavior. With a few of our clients, overprovisioning has reached an absurd extreme—one where we were told that the business units were uninterested in removing or deallocating unused resources because they were concerned that their budgets would be lowered as a result.

Breaking the loop

In teaming with industry leaders, we've concluded that the only way to break this loop is by implementing the right incentive system to drive a change in behavior across the organization as a whole. This method breeds an enormous amount of opportunity for creativity in developing the appropriate incentive system, as the right fit is dependent upon the culture of each organization.

In our experience, the accountability and incentives that deliver the best results are as follows:

– Departmental bonus: a certain fixed bonus to the entire business unit if it successfully reduces its spend by a certain amount or percentage. The head of the business unit has the discretion on how to spend that bonus. Most commonly, we have seen that it either splits evenly or proportionally among team members or is used as funds for an organizational event, gifts or even sprucing up the office space.

– Management by objectives (MBO): Organizations add a certain amount to employee MBO, making their quarterly or annual bonuses dependent upon a certain amount of reduction. This approach is much stricter, mostly used by organizations that have tried other approaches in the past and struggled to get the desired results.

– Competitions: For organizations with a strong competitive culture, competition usually drives the best outcomes. Competitions can be held internally within business units or across them. They can be between individuals, while others are between the business units themselves. These competitions can be honor-based or reward-based. The extent of the problem, the urgency in which it needs to be resolved and the company culture all need to be considered when determining the type of competition that will drive the best outcome.

With each of these accountability and incentive suggestions, a clear system for tracking outcomes and progress against the goal is necessary for these implementations to deliver the desired outcome.

Conclusion

The right incentives and accountability systems, well-defined strategic goals and a clear organizational structure make for a successful leader and CIO. CIOs can shift their teams and the entirety of the organization to a highly competitive position, taking full advantage of the agility and the elasticity the cloud provides. Establishing a realistic and measurable strategy, where each application is assessed for cloud performance, ensures that applications and teams are being managed efficiently. The right organizational structure for your team and applications, typically defined by a centralized and decentralized hybrid split by the environment or function, aids your team in effectively implementing the defined strategy. While accountability isn't always sufficient in accomplishing financial goals, incentive systems can be strong drivers to reduce budgets and increase productivity. Implementing these three components into your organization helps ensure increased cost intelligence, optimized cloud spend and incomparable performance.

About IBM Turbonomic

IBM Turbonomic provides application resource management (ARM) software used by customers to help assure application performance and governance by dynamically resourcing applications across hybrid and multicloud environments. Turbonomic network performance management (NPM) provides modern monitoring and analytics solutions to help assure continuous network performance at scale across multivendor networks for enterprises, carriers and managed services providers.

To learn more, contact your IBM Business Partner:

Bechtle Schweiz AG
+41793476882 | erich.geiger@bechtle.com

IBM

1 RightScale 2016 State of the Cloud Report, RightScale, 2016; 2020 Flexera
State of the Cloud Report, Flexera, 2020

IBM