

OPTIMISING CLOUD SPEND BY RIGHT-SIZING AND RIGHT-BUYING

Why 35% of cloud spend is being wasted and how to save it

For those in the IT and business world, the decade just gone will go down in technological history as the 'Decade of the Cloud'. For businesses, it represented quite literally a whole new world; a world in which scalability, flexibility and reduced IT spend were all possible. It was endorsed by many cloud-native start-ups that have since become the world's leading tech unicorns – the likes of Facebook, Amazon and Netflix. As a result, the Cloud went from being a back-office consideration to a key business tenet that goes as high as the C-suite.

While the last ten years have seen the Cloud become a business 'must-have', it is the next ten years that will really count as businesses expand the depth and breadth of their usage. The coming decade will define who is making the best possible use of the Cloud and who is losing out – and it is not the back-office IT team who will be the deciding factor. It is up to the C-suite to not just recognise cloud cost optimisation as a top priority, but to ensure their team has the right tools and understanding to better align cloud spend with cloud usage.

This whitepaper will provide a comprehensive overview for business leaders of how and why they are currently wasting, on average, 35% of cloud spend¹. It will then look into what tools they should be considering to help them with right-sizing and right-buying their cloud usage; that is, optimising the Cloud by prescriptively aligning both the size, type and usage of the Cloud instance with their exact needs.

Cloud central

How central is the Cloud to your business? The closer you get to a firm's IT department, the more accurate the answer is likely to be. The truth is, it is more central than most business leaders would think. If you were to ask any CEO where their business was based, they might tell you its geographical location: say, New York or London. In reality, more and more businesses are shifting off-premise, to the point at which the more truthful answer for many might be 'the Cloud'.

Today 94% of enterprises use the Cloud². The Public Cloud has experienced a particular boom with a growth rate three times that of the Private Cloud, making it a more than \$230 billion industry. Consequently, it will be the main focus of this whitepaper.

In tandem with the Cloud's rise in popularity, cloud estates have become much larger and more complex. Eighty-four percent of enterprises have a multi-cloud strategy, running applications in, or experimenting with, almost five clouds on average according to 2019 figures³.

The lost 35%

Imagine if 35% of your employees turned up for work every day but did no work. You would be paying almost a third of your workforce an annual salary simply to exist.

¹<https://resources.flexera.com/web/media/documents/rightscale-2019-state-of-the-cloud-report-from-flexera.pdf>

²See reference above

³See reference above

It's hard to imagine as it simply wouldn't happen – you wouldn't have hired them in the first place or they would be put to work somewhere they are more needed.

Yet this is happening every hour within most businesses' cloud estates. On average, 35% of cloud spend is being wasted. That equates to a minimum of \$360,000 down the drain every year for 50% of firms (those spending more than \$1.2 million on cloud per annum), and at least \$3.6 million for 13% (those spending more than \$12 million).

On top of this, because the Public Cloud charges per time used (be it by the hour, minute or second), whether a firm is using its full capacity or not, 35% of cloud capacity going unused is not a one-time loss; it's happening on an ongoing basis. While having enough headroom is vital for a firm to avoid outages in the event of peak user traffic, even factoring in huge bursts in demand this 35% is not getting touched. Unused headroom therefore represents an obvious area of savings to be made if firms were to put the right tools in place.

It is therefore no surprise that cloud spend has been ranked consistently by enterprises as their top challenge year-on-year, regardless of their cloud maturity. What may come as a surprise, however, is that despite its priority ranking, the majority of firms are still struggling to rein in spending. This can be pinpointed to two key factors: a 'like-for-like' approach to cloud migration and maintenance, and a lack of accurate, trustworthy tools.

The problem with 'like-for-like'

Most firms are still taking a 'like-for-like' approach to cloud migration and maintenance. The 'like-for-like' approach – sometimes termed 'lift-and-shift' – involves taking a simple inventory of all their virtual machines, mapping them to the closest fit instance type in their chosen provider and then firing them up. While enterprises should in theory be able to run their applications and workloads in this environment, and in fact many view it as the 'easy option', this results in any inefficiencies that exist in the on-premise world being

replicated in the Cloud. 'Like-for-like' is how businesses end up paying far too much from day one.

Broken down, the 'like-for-like' approach is problematic for a number of reasons. First and foremost, it fails to take into account the time-based payment model of the Cloud. Both in-cloud and on-premise, enterprises tend to lean on the side of caution by paying for more capacity than they need due to the fallout of being under-capacity; that is, outages and potentially devastating financial and reputation loss. The difference is, in a data centre environment, you've already bought and therefore own the servers you are using to run workloads. Meaning from a cost perspective, it doesn't matter if they are running unnecessarily. However, if the same round-the-clock timeframe for a server to be running is replicated 'like-for-like' in the cloud, unlike on-premise, you'll be paying for every hour, minute or second you're not using it.

Another factor a 'like-for-like' approach fails to account for is that, while the on-premise estate is fixed, the Cloud estate is innately flexible and scalable. Therefore, while it may make sense for the on-premise estate to build in extra capacity to allow for potential business growth and minimise the need for costly updates and additions to existing hardware, the same cannot be said for the cloud. Unlike on-premise, businesses do not need to be paying for 'just-in-case' capacity.

A lack of tools

Monitoring and updating modern IT estates is no longer a task that can be done manually. Without comprehensive oversight of their estates, firms are unable to identify where inefficiencies lie and what tools need to be deployed in what servers to address them.

In order to improve cloud-cost efficiency, understanding the relationship between the number of users hitting a platform and the number of transactions going through a platform is key. This data allows firms to 'right-size' and 'right-buy' their cloud estates based on real-time and forecasted usage.

Unfortunately, a lack of trust in the tools available on the marketplace has caused firms to avoid outsourcing their cloud cost optimisation.

While cloud providers offer some data and tools to help manage costs, it is understandably difficult for firms to trust they are giving the best possible money-saving advice. Given cloud providers are paid based on their customers' usage, it's equivalent to the fox guarding the henhouse; a conflict of interest. Without access to the underlying data supporting these recommendations, firms are right to be wary.

As a result, most firms are still relying on manual processing. This has left them unable to collect critical data in a coherent way, let alone analyse it to understand the impact on existing and future capacity.

As a result of both the all-too-common 'like-for-like' approach and a lack of trusted tools, only a minority of companies have been able to effectively address the issue of overspending. With massive financial gains on the line, there are two key processes that firms must be implementing, both at the point of migration and on an ongoing basis once in the Cloud, in order to optimise spend: right-sizing and right-buying.

Getting it right: right-sizing and right-buying

Moving to the Cloud requires much greater pre-emption and change management than many firms anticipate. Instead of taking a 'like-for-like' cloud migration approach, firms must begin their cloud journey by right-sizing their estate; that is, doing a comprehensive stocktake of the demand profile of all business workloads. Getting this right is essential as each increment in instance size is twice as expensive as the previous one, so even being one size too big means an enterprise is paying double what it should.

In order to accurately right-size, firms must first identify the sizes of the instances needed for each application workload.

They can do this by analysing which machines can be downsized from core count, memory and allocated storage, and which can be switched off altogether. This requires statistical analysis of highly granular resource utilization data. Using simple averages will significantly underplay resource requirements, while using peak without an understanding of how often those peaks occur and for how long can significantly overplay resource requirements. They must also understand the capacity of the on-premise estate if their Hybrid-IT allows them to bring work off-cloud if necessary. Applications or services in the Cloud can also be right-sized based on their actual use of resources.

The second central tenet is right-buying. This involves accurately analysing and optimising the type of instance (container, VM or 'bare metal'), the duration the instance exists for, the location and region the instance is created in, the type of purchase (on demand or reserved capacity), and the best way to buy it.

On top of this is an ongoing analysis of the billing engines of the cloud providers to identify optimal usage and policies. By undertaking a comparison of multiple cloud providers and the on-premise estate, firms will be able to select the best purchase plan for optimum price based on what their use profile actually looks like.

A combination of right-sizing and right-buying is the only way to ensure workflow demands are exactly catered for, taking peak demand into account with no costly excess headroom. These are not just one-off processes to be implemented at the cloud migration stage, but ongoing undertakings as an enterprise gets settled and expands within the world of Cloud. The process of right-sizing and right-buying one's estate is a continuous one that should last a business' entire cloud lifespan.

The right tool for the present

Cloud Cost Optimisation (CCO), a new feature available within ITRS Capacity Planner, is centred around these two processes. Based on the right-sizing



and right-buying processes, the CCO solution is able to optimise cloud spend by recommending improvements to a production estate and identifying wasted capacity, and therefore money.

Most importantly, ITRS Capacity Planner's new CCO solution backs up its recommendations by presenting the data in an easy-to-visualise way. In this way, CCO is able to prove why it makes the recommendations it does, giving firms the peace of mind that they are making the best choices for their cloud estates.

For example, in the cloud migration process, CCO prices up the 'like-for-like' migration and presents a clear comparison with the right-sized estate to demonstrate potential savings to be made.

The right tool for the future

Once a firm understands and has optimised its existing environment, it then needs to be able to predict its trajectory and what demands will need to be met in the near and long-term future.

ITRS Capacity Planner's CCO solution facilitates this via application demand modelling, which determines the relationship between application volumes and resource utilisation. In this way, it can identify capacity bottlenecks and predict future platform requirements.

In tandem, 'Forward Thinking' scenario modelling allows firms to model the impact of this demand across services. From there, they can address resource constraints and determine the end-to-end cost of servicing growth.

In this way, CCO allows businesses to identify the optimum configuration of their various applications to support not just existing but projected demand and the cost of servicing it.

To enable ongoing adaptation and optimisation as the enterprise changes or scales up, CCO is also able to integrate with a variety of service management platforms that control or automate workflows. For example, ServiceNow; if Capacity Planner suggests a change to an environment, that would go into ServiceNow as a recommendation and raise a change ticket before entering the workflow. By feeding recommendations into service management and policy-based orchestration platforms, CCO ensures that application configurations are always optimised.

Conclusion

In the world of Cloud, information is power. Having the right tools to not only gather information from the entire estate, but analyse it and comprehensively present clear, data-backed recommendations is key. By building a cloud network on the principles of right-sizing and right-buying, enterprises will be able to optimise their environments for the right workload configuration and accurately plan their monthly cloud spend leading to more accurate cloud usage. More importantly, they'll be able to return that 35% of wasted cloud spend to the wider business to make vital improvements and deliver better returns.