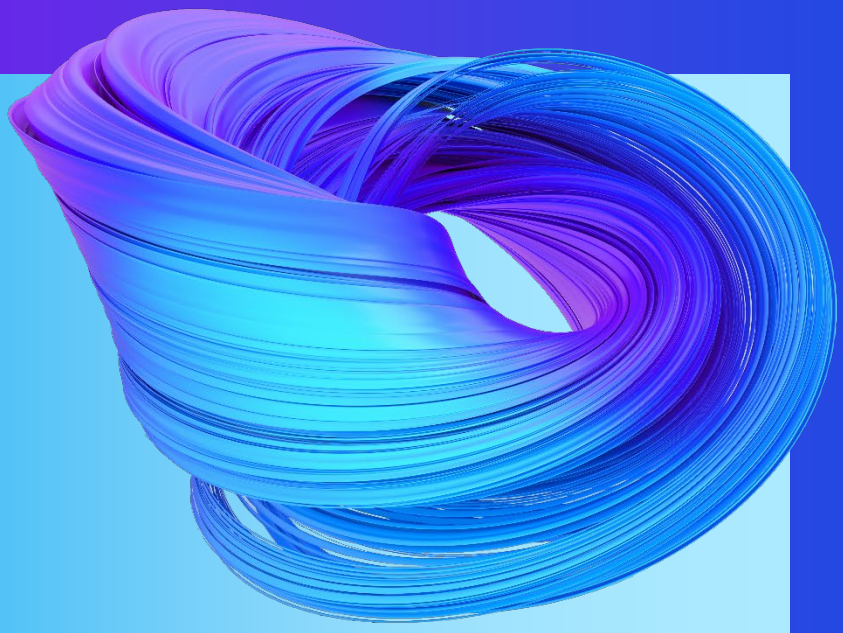
KPMG



Considerations in tackling FinOps Cloud Cost Optimisation

April 2024

Overview

As businesses continue to accelerate cloud adoption at an exponential rate, there is overwhelming increase in cloud spend with businesses incurring high cloud bills.
As per market estimates, **~$600 billion** was the estimated cloud expenditure in 2023 at a steep growth of 21.7% from 2022[1].
On the other hand, it is also estimated that on an average **35%** of **cloud expenditure was wasted** in 2023[2].

Some common factors contributing to this unforeseen rise in public cloud spend (apart from digital spikes due to recent pandemic) and cloud spend wastage are as follows:

- Lack of in-depth knowledge (both architectural and operational) of public/private/hybrid cloud set-up.
- Inability to adapt to introduction of new services and updates in various cloud offerings due to lack of awareness and training.
- Ambition of faster time to market enabling engineers to create infrastructure on demand, avoiding the traditional process of lengthy discussions and multiple approvals from finance and leadership to procure hardware.
- Over-provisioning of resources in cloud and lack of cloud cost management.
- Business divisions working in silos – finance, product, procurement, engineering etc.

Thus, optimising cloud cost has evolved as an important area of cloud financial manage2ment or **FinOps**.

The Cost Optimisation Lens

FinOps, which is a portmanteau of two domains '**Finance**' and '**DevOps**', is a collaborative effort between **finance**, **engineering**, and **operations**, where cloud spending is moderated and optimised but not curtailed unnecessarily. It involves aligning business, finance, and technology teams to maximise the value of cloud investment.

Cloud cost optimisation is a cardinal pillar of FinOps and comprises of five distinct lens that act as its anchors and strengthens it over time. It starts with collaboration amongst FinOps and other cloud teams followed by brainstorming and drafting meaningful goals and KPIs. Having robust cost reporting dashboards are imperative as one can optimise only when one knows what to optimise. Lastly, automating the optimisation techniques in due course of time and percolating the learnings within the team adds to the strength of cost optimisation.

Figure: Five Lens of Cloud Cost Optimisation

Cloud Cost Optimisation classification

While cloud cost optimisation is not a one-time activity but a continuous process, it is worth classifying cost optimisation initiatives for better planning, implementation and tracking purpose, as follows:

01 Quick wins

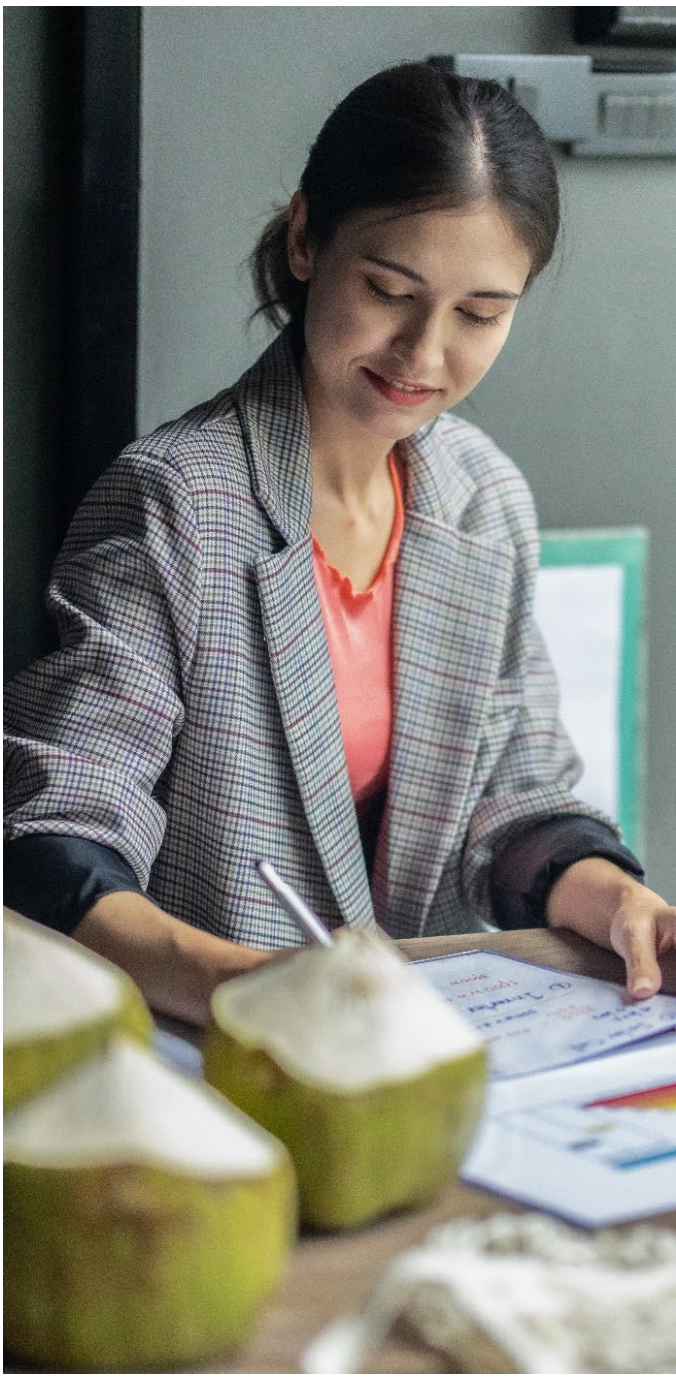
These are waste elimination strategies that can be achieved without a matured FinOps practice in place and with minimal risk to cloud operations. Some of these strategies include deletion of abandoned snapshots in non-production environments, clean-up of unattached IPs, termination of idle load balancers, removal of instances and storage volumes from non-production environment that have not been used for long duration etc.

02 Short-term cost optimisation

These strategies can be applicable to a reasonably mature FinOps practice aided by a capable Cost Reporting (show-back/ chargeback) model. This may involve cost optimisation strategies around production environment rightsizing, automated clean-up of abandoned resources, configuring auto start/stop of resources, committed usage discounts and reservations etc.

03 Long-term cost optimisation

These are complex optimisation strategies that require assistance from other FinOps arms such as automation, collaboration among multiple cloud account teams, and education of high-order initiatives - such as cloud architecture and policy/process re-design. A few examples include Kubernetes cluster optimisation, automated rightsizing, and container optimisation.



Our Approach to achieve FinOps Cost Optimisation

While there are numerous technical and non-technical levers to achieve cloud cost optimisation using the principles of FinOps conceptually, in real-world engagements there are certain critical considerations (depicted below) that must be carefully evaluated both before and during the engagement, that lay the foundation for realising cost optimisation effectively.

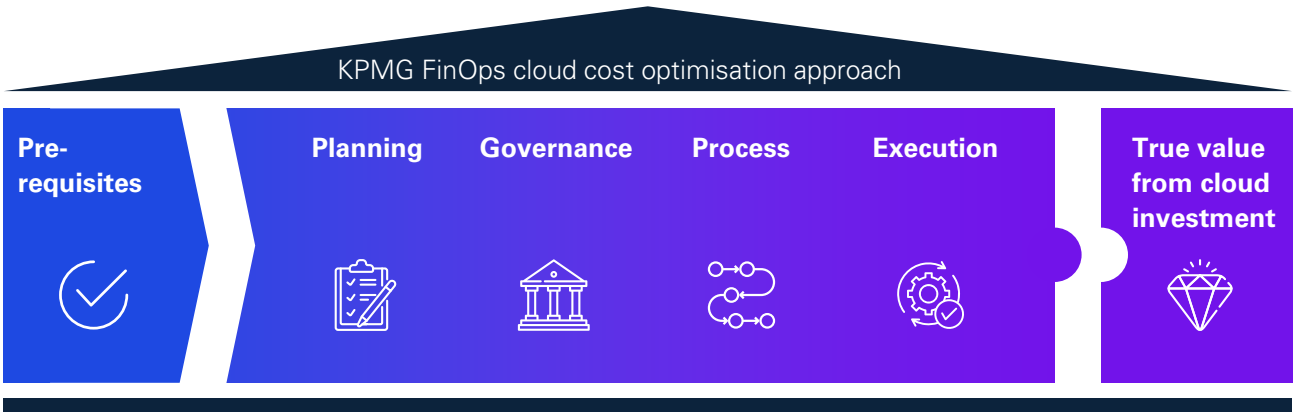


Figure: KPMG FinOps Cloud Cost Optimisation Approach

Cloud Cost Optimisation – Pre-requisites

Some of the key aspects that one needs to be cognizant to ensure readiness before embarking on a FinOps cost-optimisation engagement are as follows:

- Carefully analyse the structure and terms of the contract being signed (if consulting a third party for cost savings assessment and implementation) and ensure sufficient risk coverage is in place for commitment of savings, as assessment is always based on past data, and it does not provide a view on the current data or future business needs.
- Have initial round of workshops with the stakeholders to discuss and get a buy-in on the possible cost optimisation plan and initiatives.
- Get 'master level view' access of the target environment across cloud vendors at an organisation level. This would help in having a unified view on the historical cloud spend across the IT footprint.
- Obtain read-only access to all available native/third party FinOps tool(s) to aid assessment of cloud usage reports.
- Get access to client's current architecture blueprints from their repository.
- Request for provision of test environment in each cloud platform to ensure execution team can validate the code to automate as well as implement identified cost savings levers.

Cloud Cost Optimisation – During the Engagement

Post confirmation of pre-requisites, cost optimisation needs to be analysed from various vectors to ensure we get a practical, feasible and realistic view of the actual cloud cost optimisation. The key considerations of FinOps Cost Optimisation are as follows:

Planning

- Fix the goal-post in terms of baselining the client's cloud environment with respect to the timeline and duration for which assessment will be performed. This will ensure that all cost savings from this timeline are recorded during execution, and any environment changes in the interim (which are natural in cloud estate) do not affect the value of savings realisation.
- Categorise savings initiatives into various buckets based on parameters such as achievability and accessibility during execution.
- Based on discussion with various stakeholders, define a clear roadmap for execution into appropriate sprints based on environment complexity, account size and criticality, categories of cost savings initiatives, client's IT freeze periods, business peak seasons etc.

Governance

- Have an executive buy-in on the engagement for both assessment and implementation, and a clearly defined process/workflow for execution of cost savings initiatives including execution paths for all possible scenarios.
- Have a communication and escalation matrix for oversight, tracking, and alignment of program as per the defined roadmap.
- Create a FinOps Governance board that will comprise of key stakeholders who will govern, approve, and oversee the execution of cost savings initiatives with defined roles and responsibilities.
- Define the periodicity of Governance board cadence with clearly documented agenda, actions, and outcomes.

Process tool recommendations

- FinOps Tool insights with respect to cost optimisation are often based on past data with no business context and insights into future business plans/roadmap.
- The tool recommendations of projected cost savings are often hyperbolic and can have multiple redundancies, and no factor of achievement probability.
  - In reality, all recommendations need to be thoroughly validated first, redundancies need to be weeded out next, and probability of achievement to be factored based on discussion with business and technical stakeholders before coming up with the potential savings realisation pot.
  - Example: In most FinOps tools, cost savings due to Reserved Instance Plan and Rightsizing are showcased separately on the same instance. But in reality, rightsizing cannot be applied to instances which are part of existing Reserved Instance subscription or proposed Reserved Instance plan.

- Many a time, tools also provide recommendations based on the current footprint for which realistic implementation may not be feasible (example: constraints due to business workload requirements, operational requirements etc.).

- The tool(s) can certainly act as key enabler to provide direction on key avenues, validation of observations, and save lot of human time and effort, but over-reliance on the tool to achieve cost optimisation can result in misplaced expectations.

Execution

- Adopt a gradual approach targeting low hanging fruits (such as waste elimination) first, which helps in gaining confidence of management/stakeholders and use their active support to execute higher order initiatives.

- Do not always focus on highest cost savings scenario, as this may limit flexibility in business requirements. Also explore alternate cost savings options which may offer moderate cost savings but provide higher level of flexibility to meet business requirements, thus being more relevant and effective.

- Drive execution of cost savings across various teams through identification and recognition of Value champions, gamification of progress made among various business accounts which are under the purview of implementation.

- Do not blindly rely on recommendations from tool(s), and do not immediately act on all of them. Use these recommendations as a starting point to analyse and discuss with relevant teams on the extent of their feasibility with minimal to zero business disruption.

- Create policies and automate the identified levers using IaC (Infra-as-Code) templates to achieve faster and seamless execution and avoid potential configuration drift in the long term.

- Understand the context by talking to relevant technical and business stakeholders to see if the best-practices or recommendations apply to the given environment.

  - Example: Don't apply a blanket shutdown on non-production compute instances outside local business hours. Rather, understand the business requirements from engineering teams. There can be offshoring/nearshoring support involved with extended team working in different time-zone which need to be factored while implementing such cost savings initiatives.

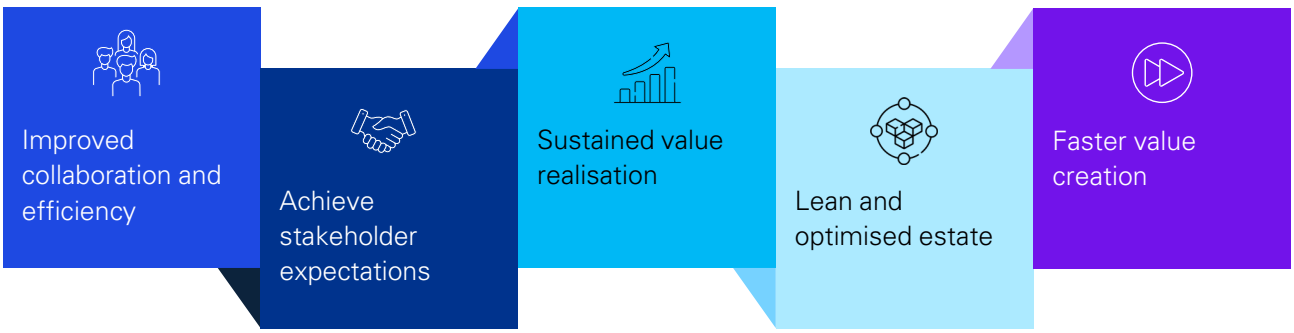
- FinOps can be driven either by a completely centralised team, or in a hybrid model involving a central hub team and a logical spoke team extending to each business unit. The size, maturity and complexity of the organisation needs to be assessed before zeroing on the most suitable FinOps structure. However, in none of the cases should FinOps be set-up in a siloed or decentralised manner since it results in sub-optimal value realisation in the long run.

- An effective FinOps team structure ensures the following:
  - Optimal and effective implementation of FinOps across entire cloud estate, attributed to learnings and experiences gained over time by the FinOps central team.
  - Centralised automation that reduces manual effort and probability of human error.
  - Effective cost through centrally governed and controlled enterprise discount plans, reservations, volume discounts and usage credits.
  - Engineers and operations team focus on their core competency and not worry about pricing or rate negotiations.
  - Tracking of unit cost economics represents true savings and is an effective unit of measure relative to cloud spend per account or aggregate cloud spend, since the latter can spike up/down based on business demands and seasonal variations.

Summing up

It is evident that organisations must adopt FinOps to manage their cloud footprint, especially to optimise escalating cloud cost, derive maximum value out of their cloud investments and reduce wastage. As new age businesses seek more avenues to optimise their cloud spend, a structured approach and adhering to practical considerations in dealing with cost optimisation need to be considered and worked upon more closely to make the engagement successful.

Some key outcomes that are delivered using our stated approach are depicted below.



An optimised cloud ecosystem will reduce operational expenses and contribute to the overall bottom-line of the organisation leading to increase in the value of an organisation's cloud investment. To know more on how we handle FinOps Cloud Cost Optimisation in action, contact us!

References:

[1] Gartner Worldwide Public Cloud Spending

[2] Gartner Survey on Cloud spend wastage

Contacts

Suharsh Paldewar

Manager, Cloud Tx – CIO Advisory

suharshpaldewar@kpmg.com

https://www.linkedin.com/in/suharshp/

Suharsh is part of KPMG's Cloud Transformation team. He is passionate about solving clients' problems by using a blend of technology, business appreciation and commercial acumen to drive transformations in Enterprise Cloud, FinOps, SRE adoption, IT service management, and provide compelling solutions for Cloud and IT Infrastructure services.

Arunava Mukherjee

Assistant Manager, Cloud Tx – CIO Advisory

arunavam@kpmg.com

https://www.linkedin.com/in/arunava11/

Arunava is part of KPMG's Cloud Transformation team specialising in FinOps, cloud migration and cloud management. He enables clients to not only manage their cloud migration journeys but also derive more value out of their investments in cloud by ensuring they have an optimised and efficient cloud ecosystem.

Some or all of the services described herein may not be permissible for KPMG audited entities and their affiliates or related entities.

kpmg.com/uk

Document Classification: KPMG Public

Create: CRT143124D | March 2024